II. Interdisciplinary Conference on Mechanics, Computers and Electrics

ICMECE 2022

06-07 October 2022

Barcelona / SPAIN

www.icmece.org

PROCEEDINGS

Edited by

Prof. Dr. Erol Kurt

ISBN: 978-605-70842-1-7

Organizing Institutions

Electrical and Computer Engineering Research Group - ecerg.com

Projenia R&D Consultancy Service Limited Company - projenia.net

Published by
Erol Kurt
on 15th March 2023
(Ankara/Turkey)

Book typesetted and designed by
Assoc. Prof. Dr.Yunus Uzun and Bekir Dursun

Cooperative Institutions

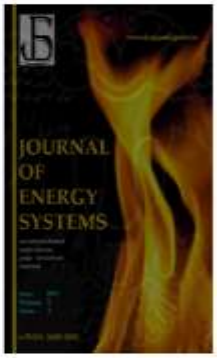
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola d'Enginyeria de Barcelona Est

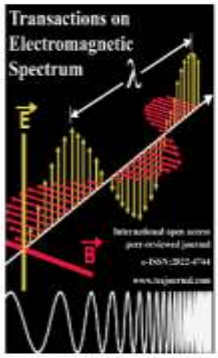
Universidad del Pais Vasco
Euskal Herriko Unibertsitatea

CROM
CENTER FOR RESEARCH ON MICROGRIDS

ALENAR
Gazi University Technology Faculty
Alternative Energy Researches Laboratory
Alternatif Enerji Araştırmaları Laboratuvarı

Supporting Institutions





Committees

Conference Founder & Chairman

Prof. Dr. Erol KURT, Gazi University, Turkey
Prof. Dr. Nicu Bizon, University of Pitesti, Romania
Prof. Dr. Jose Manuel Lopez Guede, University of Basque Country, Spain
Prof. Dr. Jose Matas Alcala, Barcelona East School of Eng.(EEBE), Politechnic Uni. Catalunia (UPC), Spain

International Organizing Committee

Prof. Dr. Adnan Sozen, Gazi University, Turkey
Prof. Dr. Carlos Rubio-Maya, Univ. Michoacana de San Nicolás de Hidalgo, Mexico
Prof. Dr. Eleonora Guseinoviene, Klaipeda University, Lithuania
Assoc. Prof. Dr. Fatih Mehmet ÖZKAL, Atatürk University, Turkey
Assoc. Prof. Dr. Fontina Petrakopoulou-Robinson, Universidad Carlos III de Madrid, Spain
Prof. Dr. Herminio Martinez, Barcelona East School of Engineering (EEBE), BarcelonaTech (UPC), Spain
Prof. Dr. Hadiyanto Hadiyanto, Indonesian Institute of Sciences, Indonesia
Prof. Dr. Ian Hunter, University of Leeds, UK
Prof. Dr. Jose M. Guerrero, Aalborg University, Denmark
Prof. Dr. Mehmet Önder Efe, Hacettepe University, Turkey
Prof. Dr. Mehmet Tekerek, Kahramanmaras Sutcu Imam University, Turkey
Dr. Mikel Larrañaga, Basque Country University (UPV/EHU), Spain
Prof. Dr. Nicu Bizon, Pitesti University, Romania
Prof. Dr. Peter Childs, Imperial College London, UK
Prof. Dr. Raoul Rashid Nigmatullin, Kazan National Research Tech. Univ., Tatarystan, Russia
Prof. Dr. Saad Mekhilef, University of Malaya, Malaysia
Prof. Dr. Saeed Badshah, Int. Islam. Uni. Islamabad, Pakistan
Prof. Dr. Sagdulla L. Lutpullaev, Uzbekistan Academy of Sciences, Uzbekistan
Prof. Dr. Serguei Martemianov, University of Poitiers, France
Assoc. Prof. Dr. Yunus Uzun, Aksaray University, Türkiye

International Scientific Committee

Assoc. Prof. Dr.Adem Tekerek, Gazi University, Turkey
Assoc. Prof. Dr. Ali Jazie, University of Al-Qadisiyah, Iraq
Dr. Ana Boyano, University of Basque Country, Spain
Dr. Angeliki Chatzidimitriou, Aristotle University of Thessaloniki, Greece
Prof. Dr. Anand Nayyar, Duy Tan University, Da Nang, Vietnam
Prof. Dr. Antonio Soria Verdugo, Universidad Carlos III de Madrid, Spain
Dr. Athar Waseem, International Islamic University, Pakistan
Prof. Dr. Aybaba Hançerlioğulları, Kastamonu University, Turkey
Prof. Dr. Ayman El-Hag, American University of Sharjah, UAE
Prof. Dr. Bernabé Marí Soucase, Polytechnical University of Valencia, Spain
Assist. Prof. Dr. Burak Akin, Yildiz Technical University, Turkey
Assoc. Prof. Dr.Bünyamin Ciylan, Gazi University, Turkey
Assoc. Prof. Dr. Bünyamin Tamyürek, Eskisehir Osmangazi University, Turkey
Prof. Dr. Can Çınar, Gazi University, Turkey
Prof. Dr. Carolina Marugan Cruz, Universidad Carlos III de Madrid, Spain
Assoc. Prof. Dr. Christiane Hennig, German Biomass Research Center, Germany
Prof. Dr. Corneliu Marinescu, Transilvania University of Brasov, Romania
Assoc. Prof. Coşku Kasnakoğlu, TOBB Ekonomy and Technology University, Turkey
Assoc. Prof. Dr. Çigdem Yangin Gömeç, Istanbul Technical University, Turkey

Dr. Danny Müeller, Technische Universität Wien, Austria
Assoc. Prof. Dr. Diana Zalostiba, Riga Technical University, Latvia
Dr. Eduar Eduardo Zarza, CIEMAT Solar Platform of Almeria, Spain
Dr. Ekaitz Zulueta, University of Basque Country, Spain
Prof. Dr. Eleonora Guseinoviene, Klaipeda University, Lithuania
Prof. Dr. Eyyup Tel, Osmaniye Korkut Ata University, Turkey
Prof. Dr. Farqad Al-Hadeethi, Royal Scientific Society of Jordan, Jordan
Prof. Dr. Ferda Hacıvelioğlu, Gebze Teknik University, Turkey
Assoc. Prof. Dr. Fontina Petrakopoulou-Robinson, Universidad Carlos III de Madrid, Spain
Prof. Dr. Francesco Calise, University of Naples Federico II, Italy
Assoc. Prof. Dr. Francesco Cottone, University of Perugia, Italy
Prof. Dr. Guang-Bin Huang, Nanyang Technological University, Singapore
Prof. Dr. Guido Van Oost, University of Gent, Belgium
Prof. Dr. Güngör Bal, Gazi University, Turkey
Prof. Dr. Hakan Ateş, Gazi University, Turkey
Prof. Dr. Hakan Çiftçi, Gazi University, Turkey
Prof. Dr. H. Mehmet Şahin, Karabük University, Turkey
Prof. Dr. Haitham Abu-Rub, Texas A&M University at Qatar, Qatar
Prof. Dr. Halil İbrahim Ünal, Gazi University, Turkey
Assoc. Prof Dr. Hasan Köten, Medeniyet University, Turkey
Prof. Dr. Herman Vermaak, Central University of Technology, Free State, South Africa
Prof. Dr. Hüseyin Çelikkan, Gazi University, Turkey
Prof. Dr. Ibrahim Dincer, University of Ontario, Canada
Prof. Dr. Ibrahim Sefa, Gazi University, Turkey
Prof. Dr. Igor Fernandez, Basque Country University, Spain(UPV/EHU)
Prof. Dr. Ilya Galkin, Riga Technical University, Latvia
Assoc. Prof. Dr. Ilona Sárvári Horváth, University of Borås, Sweden
Prof. Dr. Jongho Yoon, Hanbat National University, S. Korea
Prof. Dr. Jongsoon Song, Chosun University, S. Korea
Prof. Dr. Jorge R. Frade, University of Aveiro, Portugal
Prof. Dr. Jose A. Aguado, University of Malaga, Spain
Assoc. Prof. Dr. Jose A. Ramos Hernanz, Universidad del Pais Vasco, Spain
Assoc. Prof. Dr. Jose M. Lopez Guede, Universidad del Pais Vasco, Spain
Prof. Dr. Josep Guerrero, Aalborg University, Denmark
Assoc. Prof. Dr. M. Hanefi Calp, Karadeniz Teknik University, Turkey
Assoc. Prof. Dr. K.Premkumar, Rajalakshmi Engineering College, Chennai, India
Prof. Dr. Kozo Taguchi, Ritsumeikan University, Japan
Prof. Dr. Lütfi Öksüz, Isparta Suleyman Demirel University, Turkey
Prof. Dr. Leijun Xu, Jiangsu University, China
Prof. Dr. Jun Yang, Huazhong University of Science and Technology, China
Dr. Loreto V. Gutierrez, CIEMAT Solar Platform of Almeria, Spain
Prof. Dr. Mahmood Ghoranneviss, Islamic Azad University, Iran
Prof. Dr. Maria Venegas, Universidad Carlos III de Madrid, Spain
Assoc. Prof. Dr. Mario E. Magana, Oregon state University, USA
Prof. Dr. Maris Klavins, University of Latvia, Latvia
Prof. Dr. Mehmet Tekerek, Kahramanmaras Sutcu Imam University, Turkey
Assoc. Prof. Dr. Merih Palandöken, İzmir Katip Çelebi University, Turkey
Prof. Dr. Metin Gürü, Gazi University, Turkey
Prof. Dr. Milan Stork, University of West Bohemia, Czech Republic
Prof. Dr. Mohammad N. A. Hawlader, International Islamic University, Malaysia
Prof. Dr. Munir Nayfeh, University of Illinois at Urbana-Champaign, USA
Prof. Dr. Murat Yücel, Gazi University, Turkey
Prof. Dr. Muris Torlak, Sarajevo University, Bosnia and Herzegovina
Prof. Dr. Mustafa Ilbas, Gazi University, Turkey
Prof. Dr. Mustafa Aktaş, Gazi University, Turkey
Prof. Dr. Mykola Radchenko, Admiral Makarov National University of Shipbuilding, Ukraine
Prof. Dr. N. Nasimuddin, Institute for Infocomm Research, Singapore

Dr. Nam Choon Baek, Korea Institute of Energy Research, S. Korea
Prof. Dr. Namazov Subhan Nadiroglu, Azerbaijan Technical University, Azerbaijan
Prof. Dr. Narasimha G. Reddy, Lamar University, USA
Assoc. Prof Dr. Natalia Tintaru, Vilnius University, Lithuania
Prof. Dr. Nicolae Paraschiv, Petroleum - Gas University of Ploiesti, Romania
Prof. Dr. Nicu Bizon, Pitesti University, Romania
Prof. Dr. Nikolay Djagarov, Nikola Vaptsarov Naval Academy, Bulgaria
Dr. Nilufar R. Avezova, Uzbekistan Academy of Sciences, Uzbekistan
Prof. Dr. Nurettin Topaloğlu, Gazi University, Turkey
Dr. Pablo Eguia, Basque Country University (UPV/EHU), Spain
Prof. Dr. Pedro Juan Roig, Universidad Miguel Hernández, Spain
Prof. Dr. Peter Lund, Aaalto University, Finland
Prof. Dr. Poul Alberg Østergaard, Aalborg University, Denmark
Prof. Dr. Rachid Chenni, University Mentouri of Constantine, Algeria
Prof. Dr. Rafael K. Jardan, Budapest Univ. of Technology and Economics, Hungary
Prof. Dr. Rafaela Hillerbrand, RWTH Aachen University, Germany
Prof. Dr. Rakesh Kumar, NIT Tiruchirappalli, India
Prof. Dr. Raoul Rashid Nigmatullin, Kazan National Research Technical University,Russia
Dr. Rosaria Villari, Italian National Agency for New Technologies, Italy
Prof. Dr. Saffa Riffat, Nottingham University, UK
Dr. Seçil Karatay, Kastamonu Univeristy, Turkey
Assoc. Prof. Dr. Selami Balcı, KMU University, Turkey
Prof. Dr. H. Serdar Yücesu, Gazi University, Turkey
Prof. Dr. Shadia J. Ikhmayies, Al Isra University, Jordan
Prof. Dr. Sing Lee, Institute for Plasma Focus Studies, Australia
Prof. Dr. Sor Saw Heoh, Nilai University, Malaysia
Prof. Dr. Souad A.M. Albathi, Int. Islamic Uni. Malaysia, Malaysia
Prof. Dr. Sujit Barhate, Savitribai Phune Pune University, India
Prof. Dr. Tahir Güllüoğlu, Harran University, Turkey
Assoc. Prof. Dr. T.Thamizhselvan, Rajalakshmi Engineering College, Chennai, India
Prof. Dr. Tae Hee Lee, Hanyang University, S. Korea
Dr. Unai Fernandez-Gamiz, University of Basque Country, Spain
Prof. Dr. V. Jagannathan, Bhabha Atomic Research Center, India
Prof. Dr. Wail N. Al-Rifaie, Philadelphia University, Jordan
Prof. Dr. Wang Ru-Zhu, Shanghai Jiao Tong University, China
Assoc. Prof. Dr. Yong Song, Institute of Nuclear Energy Safety Technology, China
Prof. Dr. Zhiqiang Zhu, Institute of Nuclear Energy Safety Technology, China
Assoc. Prof. Dr. Yussupova Gulbakhar Madreymovna, Department of RET, Turan University

Conference Secretary

Lec. Bekir Dursun, Trakya University, Turkey
Dr. Kayhan Celik, Gazi University, Turkey

FOREWORD

Dear Colleagues,

We are glad to meet with you in the second edition of the serial event - Interdisciplinary Conference on Mechanics, Computers and Electrics (ICMECE 2022). This first event has been organized fully virtual form due to the global pandemy. The local organizers ECERG – Electrical and Computer Engineering Research Group at Gazi University and PROJENIA welcome to all participants. Many international institutions took a part as the cooperating institutions including many international refeered academic journals.

The goal of ICMECE 2022 is to gather scientists, engineers, researchers, technicians and industrial representatives to present the cutting-edge studies on Mechanical, Computer and Electrical Systems and form an interdisciplinary academic forum to discuss the scientific and engineering issues to arrive at more complete systems for the applications of future world.

The audience on the interdisciplinary issues can be M.Sc./Ph.D. students, post graduate Students, research Scholars, post-doc scientists, senior academicians and all other academical bodies related to Mechanical Engineering, Civil Engineering, Electrical, Electronics & Communication Engineering, Computer Science & Engineering, Communication Engineering, Mechatronics, and Natural Sciences. In addition, companies focusing on the entrepreneurship and research & development can participate, too. The conference will perform traditional research paper presentations as well as the keynote talks by prominent speakers focusing on the related state-of-the-art technologies in the interdisciplinary fields of the conference.

Presently, the academia and researchers are not only pondering but also experiencing the overwhelming outcomes of interdisciplinary researches. Indeed, interdisciplinary studies are encouraged by the governments, research agencies and the academic institutions in order to create more complete products and knowledge. The intent behind such an interdisciplinary and multidisciplinary approach is to provide a common platform, where academia, industrial delegates and nominees from various government and private universities and institutions can meet, and cherish about achievements so far, as well as deliberate upon futuristic approaches. The deliberations will not only encompass all avenues of electrical, electronics, computer science and information technology but also through spotlight on positive and inadvertent impact of modern technologies our societies.

The context of the conference is fostering the research culture among academia and industry on new ideas and brainstormings. Furthermore, the intent of the organization is to help the transcendental growth, recent trends, innovations and security issues involved in the domain of communication technologies, sustainable smart electrical systems, high performance computing, big data, social media, hardware & software design, advanced software engineering, Internet of Things (IoT), e-governance etc., and their impact on societal applications.

This conference proceedings consists of interdisciplinary technical papers. Each of the papers has been reviewed by independent reviewers in a double blind environment. According to the conference registration records, the participants from 64 countries contributed at the technical meetings of the conferences. We gratitute all authors, keynote speakers, special session organizers, reviewers, session chairmen and scientific board for their precisious contribution and hope to extend these cooperation for the next events, too.

This proceeedings is delivered online via the conference website icmece.org/2022/proceedings.pdf. We would like to send our warmest greetings to all of the participants and looking forward to having your future contribution to the future events for a much green, pandemy-free and peaceful word. (01st March 2023, Ankara)

Prof. Dr. Erol KURT
Chairman of ICMECE Series
Gazi University, Technology Faculty
Department of Electrical an d Electronics Engineering
06500 Besevler ANKARA TURKEY
E-mail: ekurt52tr@yahoo.com
Tel: +90 312 202 8550 (office)

CONTENTS

Use of Machine Learning Algorithms in Fatigue Prediction

Jaromir Kaspar
Inovations department
Mubea spol. s. r. o.
Zebrak, Czech Republic
jaromir.kaspar@mubea.com

Vaclav Cvancara
Inovations department
Mubea spol. s. r. o.
Zebrak, Czech Republic
vaclav.cvancara@mubea.com

***Abstract*—Machine learning algorithms have a wide application. This paper shows their use in mechanics of solids, specifically in high cyclic fatigue of stamped parts. Forming process has an impact on part's fatigue. It is necessary to analyse the forming process if lifetime is predicted. Nevertheless, numerical simulation of stamping is time consuming and many times it is not even possible to perform it. Simplified calculation methods including machine learning algorithms can be used in these situations. Aim of this paper is to validate simplified calculation model which includes algorithms originally developed for machine learning.**

Keywords— machine learning, inverse stamping, dimensionality reduction, high cyclic fatigue

I. INTRODUCTION

Cold formed sheet metal parts are widely used especially in automotive, because this technology is suitable for mass production. The final properties of the products are influenced by forming process. Plastic strain occurs during forming and residual stress and changes of the wall thickness can also be observed.

Analysis of the production process using numerical simulation is usually a difficult task. It requires knowledge of the production process including the geometry of the production tools. The required inputs are not available many times and the time needed for the simulation can be unacceptably long, especially in optimization studies. In such situations inverse stamping can be used. The inverse stamping method is fast and simple. On the other hand, its results are less precise due to simplifications. For example, it is assumed that all forming operations are done at once and friction forces are usually neglected. Both methods, i.e. numerical simulation and inverse stamping can be used to analyse the resulting plastic strain, residual stress and changes of the wall thickness.

The influence of residual stress on high cyclic fatigue can be considered in the same way as any other constant (mean) stress. Nevertheless, strong relaxation effect occurs many times and influence of residual stress can be weak. The influence of plastic strain has been studied and several methods for its consideration in high cyclic fatigue have been developed. These methods are available in commercial fatigue programs.

The algorithm of the inverse stamping method described in literature is formulated for shells. Nevertheless, we adapted the method for solid meshes. This modification extends its range of use. Primarily, modified method enables user to analyse parts with variable thickness. The modification is described in the next section. As the inverse stamping method is fast and easy to use, it is excellent for optimization loops.

II. INVERSE STAMPING AND DIMENSION REDUCTION

Standard inverse stamping algorithm formulated for shells has three main steps [1]:

1. Three dimensional (3D) triangular shell mesh is projected onto a flat plane and in the same time each element is unfolded to be parallel with the projection plane.
2. Two dimensional (2D) finite element analysis is done for iterative improvement of the initial blank shape.
3. Post-processing phase when effective plastic strain, wall thickness or residual stress are evaluated.

Metal sheet part represented by solid tetrahedral finite element mesh can be viewed like two triangular shell meshes, the first one connected with top surface of solid mesh and the second one connected with bottom surface of solid mesh. Modified algorithm has following steps:

1. Local thickness is calculated as a distance between top mesh and bottom mesh.
2. Mid-surface shell mesh is calculated based on top triangular mesh. Each element is moved about $t/2$ in reverse outer normal direction. Symbol t denotes local thickness calculated in previous step.
3. Three dimensional mid-surface is projected onto a flat plane and also each element is unfolded to be parallel with the projection plane.
4. 2D finite element analysis is done for iterative improvement of the initial blank shape.
5. Post-processing phase when effective plastic strain, wall thickness or residual stress are evaluated.
6. Mid-surface mesh is calculated based on bottom triangular mesh.
7. Steps 3-5 are repeated for current mid-surface mesh.
8. Results of both runs are saved into common result file.

Thickness of each element is individually calculated in step 1. This enables user to easily consider the variable part's thickness.

The third step of modified algorithm has crucial impact on inverse stamping reliability and robustness. If the analysed part has walls perpendicular to the projection plane or if includes holes and other features, then the dimensional reduction can be difficult. It is not possible to remove these features from the analysed body in fatigue calculation, because stress is concentrated around them. Researches haven't paid too much attention to the dimensional reduction problem in context of inverse stamping. Dimension reduction based on geometrical approach is studied in [2]. The presence of holes and prominences limits those methods. Reduction dimension problem can be solved by using algorithms common in machine learning. The algorithms suitable for inverse stamping are: Modified Locally Linear Embedding (MLLE) [3], Hessian Locally Linear Embedding (HLLE) [4], Local Tangent Space Alignment (LTSA) [5,6], t-distribution Stochastic Neighbor Embedding (t-SNE) [7]. The algorithms are joined into one robust dimension reduction procedure as it is proposed in [8]. Algorithms MLLE, HLLE, LTSA, t-SNE are implemented in Python library scikit-learn [9,10].

Unfolded and projected element shapes are iteratively compared in the fourth step of modified algorithm. Shape of projected elements is approached to the shape of unfolded elements. Procedure is stopped when the shape change of two consecutive iterations is small enough. The output from this step is the initial blank shape.

Quantities important for fatigue calculation are evaluated in the fifth step of modified algorithm. The most important quantity is effective plastic strain φ_V, equation (1) from [11], where l_0, b_0 and h_0 denote cuboid body dimensions before deformation and index 1 denotes the same dimension after deformation.

$$\varphi_V = \sqrt{\frac{2}{3}\left[\left(ln\frac{l_1}{l_0}\right)^2 + \left(ln\frac{b_1}{b_0}\right)^2 + \left(ln\frac{h_1}{h_0}\right)^2\right]} \quad (1)$$

Also other quantities can be evaluated, e. g. local metal sheet thickness or residual stress. Evaluation is based on the difference between initial blank shape achieved in the fourth step and part's shape.

III. FATIGUE PREDICTION

A. Effect of plastic strain

Effective plastic strain initiated during forming can have significant impact on high cyclic fatigue. Method of Variable Slopes (MVS) [12] and Material Law of Steel Sheet (MLSS) [13] enable us to include this effect into fatigue calculation. Fig. 1 is expansion of strain fatigue curve into 3D. Axis with effective plastic strain φ_V initiated during forming was added. Fig. 1 shows that higher φ_V leads to longer fatigue life. Symbol N denotes number of cycles and ε_a denotes strain amplitude limit.

B. Wall thinning

Forming process changes wall thickness locally. Especially thinning can be dangerous for fatigue because thinner wall withstands lower stress. The presented method is suitable for analysis of parts made from metal sheet of variable thickness. It is also possible to use 3D scan as an input for the presented method and include part's wall thickness imperfections in the analysis.

If 3D scan is used as input for the inverse method then modification of mesh used for operational stress analysis is not needed. On the other hand, if idealized part model is used as input then the mesh used for operational stress analysis should be modified in order to include the thinning effect. Local wall thickness is evaluated in inverse stamping post-processing phase and its results are used for mesh modification. If constant volume assumption [14] is made, then thickness change can be calculated easily. Surface nodes of solid mesh are moved in normal surface direction. The move distance is related to calculated wall thickness change.

C. Residual stress

Forming is irreversible process which leads to residual stress creation. Residual stress relaxes due to cyclic loading. The highest relaxation usually occurs during first few load cycles. Stress relaxation has to be investigated experimentally which hasn't been done yet. Therefore, it is assumed in later fatigue calculation that stress relaxation is total and no residual stress is considered.

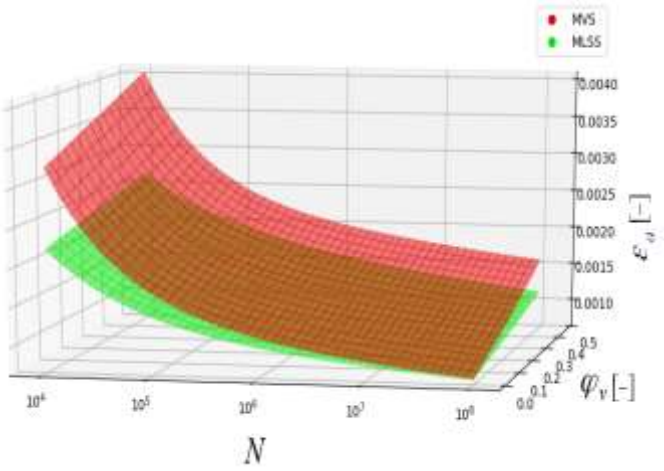


Fig. 1. Method of Variable Slope (MVS) and Material Law of Steel Sheet (MLSS)

IV. RESULTS AND DISCUSION

The calculation procedure based on inverse stamping including machine learning algorithms, operational stress analysis via FEM and high cyclic fatigue evaluation has to be validated. Therefore, experiment was set up and its results were compared with calculations. Fatigue of stabilizer clamp exposed to three points bending test was examined. Fig. 2 shows the calculation and experiment setup. The clamp was supported on its edges and alternating force (red arrow) pushed on the clamp in the middle. A piston generates the loading force.

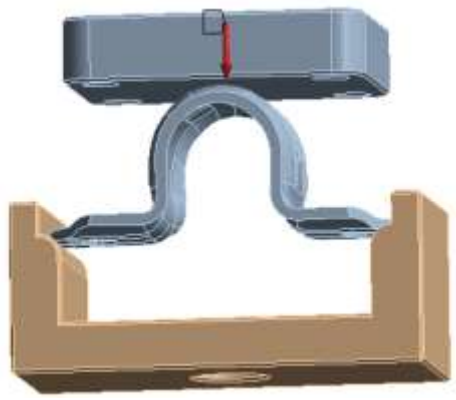


Fig. 2. Test and calculation setup.

Three of the tested specimens were scanned before test and geometrical model of clamp was created, i. e. effect of walls thinning was included in the calculation procedure.

Three loading levels were calculated and tested, see Table 1. Totally thirty specimens were tested, ten specimens for each level. Loading force frequency during the tests was 3 Hz. Test was stopped when macroscopic crack was noticed, Fig. 3. Background grid in Fig. 3 has size 10 mm.

Piston minimum and maximum displacement was recorded during testing. An example of this record in case of level 1 specimen is in Fig. 4. It is possible to estimate a time point when fatigue crack was initiated and started to propagate. Measured data was corrected by the crack growth removing.

Comparison between calculations and experiment provides Table 1 and Fig. 5. If effective plastic strain influence is completely neglected, then difference between calculation and corrected experimental data is more than 90%. Methods MLSS and MVS take into account the effective plastic strain influence. Difference between corrected experimental data and calculation based on MLSS method varies from 7% to 43%. Difference between corrected experimental data and calculation based on MVS method is 8% (level 1), 11% (level 2) and 78% (level 3). It should be kept in mind that experiment was stopped when crack had macroscopic length. On the other hand, determining point in calculation is a time point when crack is initiated. This would leaded to discrepancy if experimental data wasn't corrected.

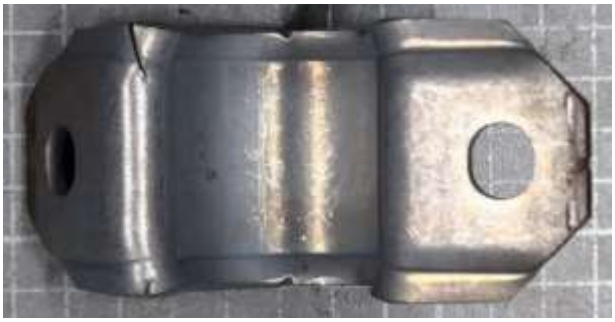


Fig. 3. Cracked clamp after high cyclic fatigue test.

All calculation models considered wall thickness change because they are based on 3D scan.

TABLE II. CLAMP LOADING AND RESULTS

Level	Mean Force (N)	Force amplitude (N)	Number of cycles until cracking				
			Experiment (uncorrected) - average	Experiment (corrected) - average	Calculation		
					Without plastic strain infl.	MLSS	MVS
1	3000	2000	143147	122626	11220	81600	112500
2	2625	1750	306578	271157	21160	154900	240960
3	2250	1500	549905	497275	43260	461000	889200

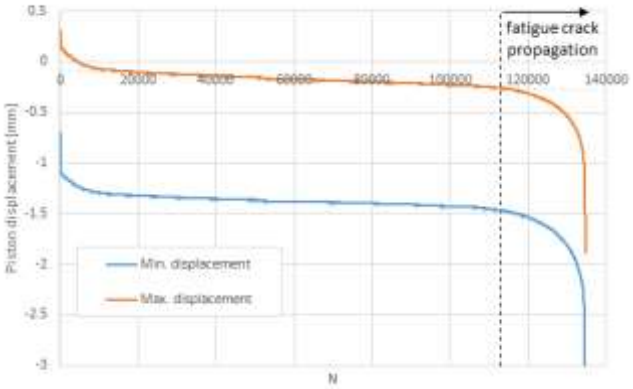


Fig. 4. Piston displacement over cycles.

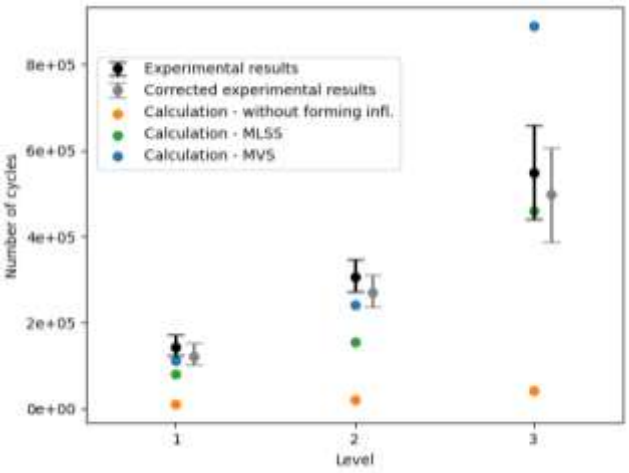


Fig. 5. Comparison of calculations and high cyclic fatigue tests.

V. CONCLUSION

Inverse stamping algorithm was introduced as an alternative to numerical simulation of forming process. If the fatigue is calculated then stress concentrators like holes and protrusions cannot be removed. This places increased demands on the inverse stamping. This model details make mesh projection difficult. Algorithms developed for machine learning can be helpful in this

situations and they increase inverse stamping robustness and reliability. Such calculation model was compared with high cyclic fatigue tests.

The comparison shows that ignoring the effect of effective plastic strain leads to huge difference between experimental and calculation results. The effect of effective plastic strain was considered via MLSS and MVS method. The consideration led to results improvement especially in case of MLSS method. This method seems to be a better choice for the practical use as well, because it provides more conservative results.

Use of inverse stamping in cold formed part's fatigue does not elongate the calculation procedure significantly. In spite of this, the impact on result accuracy can be important.

ACKNOWLEDGMENT

This paper was prepared in cooperation with Mubea, spol. s.r.o. using the laboratory for research and testing purposes.

REFERENCES

[1] J. Kaspar, P. Bernardin, V. Lasova, "Increasing the robustness of an inverse stamping algorithm," in MM Science Journal, June 2022, pp. 5684-5688.

[2] Y. Q. Guo, H. Naceur, K. Debray and F. Bogard, "Initial solution estimation to speed up inverse approach in stamping modeling" in Engineering Computations, vol. 20, no. 7, 2003, pp. 810-834.

[3] Z. Zhang and J. Wang, "MLLE: Modified locally linear embedding using multiple weights," in Advances in Neural Information Processing Systems, vol 19, January 2006, pp. 1593-1600.

[4] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," in PNAS, vol. 100, no. 10, May 2003, pp. 5591-5596.

[5] J. Wang, "Geometric structure of high-dimensional data and dimensionality reduction," Beijing: Higher Education Press, 2012. ISBN 978-7-04-031704-6.

[6] L. Honguy, C. Wenbin and S. I-Fan, "Supervised Local Tangent Space Alignment for Classification," in Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, 2005, pp. 1620-1621.

[7] L. J. P. van der Maaten and G. Hinton, "Visualizing data using t-SNE," in Journal of Machine Learning Research, vol. 9, 2008, pp 2579-2605.

[8] J. Kaspar, M. Svagr, P. Bernardin, V. Lasova and O. Sedivy, "Dimension reduction using the inverse stamping method, " in MM Science Journal, October 2021, pp. 4810-4817.

[9] W. Richert and L. P. Coelho, Building Machine Learning Systems with Python, July 2013, ISBN 978-1-78216-140-0.

[10] L. Buitinck et al., "API design for machine learning software: experiences from the scikit-learn project" in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108-122, 2013.

[11] Magna Powertrain Inc., FEMFAT 2022a User Manual. St. Valentin, January 2022.

[12] A. Hatscher, Estimation of the cyclic characteristics of steels (Abschätzung der zyklischen Kennwerte von Stählen), Claushal: Papierflieger, 2004, ISBN 3-89720-719-2.

[13] R. Massendorf, The influence of forming on the cyclic material characteristic values of fine sheets (Einfluss der Umformung auf die zyklischen Werkstoffkennwerte von Feinblech), Claushal: Papierflieger, 2000, ISBN 3-89720-413-4.

[14] J. Hu, Z. Marciniak and J. Duncan, Mechanics of sheet metal forming, Butterworth-Heinemann, 2002, ISBN 0-7506-5300-0, pp. 30.

Energy Density and Thermal Stability Evaluation of Polymer Nanocomposites for Dielectric Capacitor

Uwa O. Uyor
Department of Chemical, Metallurgical and Materials Engineering,
Tshwane University of Technology,
Pretoria, South Africa.
Department of Metallurgical and Material Engineering,
University of Nigeria, Nsukka,
UyorUO@tut.ac.za

Patricia A. Popoola
Department of Chemical, Metallurgical and Materials Engineering,
Tshwane University of Technology,
Pretoria, South Africa,
PopoolaAPI@tut.ac.za

Olawale M. Popoola
Department of Electrical Engineering, Center for Energy and Electric Power,
Tshwane University of Technology,
Pretoria, South Africa,
PopoolaO@tut.ac.za

Dada Modupeola
Department of Chemical, Metallurgical and Materials Engineering,
Tshwane University of Technology,
Pretoria, South Africa,
dadadupeola@gmail.com

***Abstract*—Polypropylene - PP is one of the commonly used polymers in the design of dielectric capacitors and power applications. However, it is associated with the shortcomings of low energy density and poor thermal stability, which place hurdles on its applications in advanced power-energy-related fields. In this study, PP nanocomposites incorporating functionalized insulative and conductive 0D (barium titanate - BT), 1D (carbon nanotubes - CNTs), and 2D (graphene nanosheets - GNs and boron nitride - BN) nanoparticles were developed to overcome these issues. The hybrid of insulative and conductive nanoparticles promoted the creation of nanocapacitors in the PP matrix with enhanced energy density and dielectric constant. The different dimensional structures of the carbon-ceramic nanoparticles contributed significantly to the enhancement of the thermal property of the PP nanocomposites. Hydrothermal and self-assembly of BN/GNs (denoted as BNG) and BT/CNTs (denoted as BTC) were employed in this study. The prepared nanoparticles were introduced into the PP matrix by first mixing with polypropylene maleic anhydrate via solution blending to promote compatibility, followed by melt mixing with pure PP matrix. The PP nanocomposites showed optimal of about 47ºC, 4972.3% and 200% increase in thermal stability, dielectric constant, and energy density relative to the pure PP respectively. With the achieved results, the developed PP nanocomposites can be used in the development of capacitors for power-energy applications.**

Keywords—Energy Density, Thermal stability, Dielectric, Polypropylene and Nanocomposites

I. INTRODUCTION

The development of dielectric nanocomposite-based polymers that have a high dielectric constant, energy density, and thermal stability has attracted research interest for use in power-related applications. Because of their high-power density and ability to charge and discharge quickly, dielectric energy materials are preferred for storing electrical energy. A significant quantity of energy can be stored and delivered by such energy storage materials in a straightforward system that can operate for millions of cycles [1]. Because dielectric capacitors have the desired power density property, they can be merged with other high-energy density systems for optimum energy storage. An innovative way to store electrical energy is through a hybrid circuit consisting of dielectric capacitors and electrochemical energy storage technologies [2]. Polymeric materials are commonly used to design dielectric capacitors due to their simplicity in construction, low weight, flexibility, chemical resistance, high breakdown strength, ease of processability, and affordability. Thus, polymer dielectrics are utilized in a variety of electrical and electronic as well as energy storage [3].

However, polymer dielectrics, as well as polypropylene – PP, have low energy density/dielectric constant and low thermal stability, which are the drawbacks that have attracted research attention for advanced power-related applications. Given that PP has high voltage endurance capability, the low dielectric constant associated with it must be enhanced to practically realize a significant improvement in its energy density [1, 4]. In line with this, various studies have modified the matrices of polymer dielectrics to enhance their suitability for advanced dielectric energy storage. For example, different ceramic reinforcements have been incorporated into different polymer matrices to increase dielectric constant [5,6]. However, it frequently takes a high percentage of ceramic reinforcements to show a noticeable improvement in the dielectric properties, which complicates processing and worsens the mechanical behaviours of such composites [7,8].

On the other hand, the potential of graphene nanosheets (GNs) and carbon nanotubes (CNTs) for enhancing the energy storage, dielectric constant, thermal, and mechanical properties of polymers has been extremely exciting [9,10]. By incorporating CNTs/GNs in a polymer system close to the percolation threshold, a dielectric constant of roughly 4500 at 1 kHz [11] and 2360 at 1 kHz [12] have been reported, which was credited to the creation of nanocapacitors in the polymer system due to the difference in the conductivity of the polymers and the

nanoparticles [11]. However, there are significant agglomerations of GNs/CNTs in polymer systems, large energy loss, and poor voltage endurance pertaining to this group of polymer nanocomposites [13]. Since the voltage withstand-ability of such dielectric nanocomposites is often low, they show a decrease in energy density, thereby limiting their usage as advanced dielectric energy materials.

For the design of capacitors, high dielectric constant, energy density and good thermal properties to withstand processing and service operational temperature are essential. Due to the fact that some properties degrade while others advance, it has proven challenging to obtain all the desired features in a single polymer dielectric material. This study aims to create polymer dielectric nanocomposites with enhanced dielectric and thermal stability for dielectric energy materials and other applications. This study used functionalized insulative and conductive 0D (barium titanate - BT), 1D (carbon nanotubes - CNTs), and 2D (graphene nanosheets - GNs and boron nitride - BN) nanoparticles to simultaneously address the challenges of low dielectric constant, low energy capacity, and poor thermal management associated with PP. The features needed for advanced dielectric capacitors' design and other energy-power related applications were improved with the developed dielectric nanocomposites.

II. EXPERIMENTAL

A. Materials

Graphene nanosheets (GNs) (essay - > 95%, O - < 3%, diameter - 2-3 μm, few nano thickness - 6-8 nm), xylene plus ethylbenzene basis (essay - > 98.5%), dopamine hydrochloride (PDA) (essay - > 98%), 3-glycidoxypropyltrimethoxysilane (GPTMS), polypropylene grafted maleic anhydride (PPMA) (0.5%, density - 0.92g/mL) and polypropylene (PP) (melt index – 4g/10min, 230oC/2.16kg, density - 0.9g/mL) were all supplied by Sigma Aldrich. Hongwu International Group in China supplied hexagonal boron nitride (BN) (assay > 99.5%, particles size < 100 nm) and multi-walled carbon nanotubes (CNTs) (essay > 98%, 10-30 nm diameter and 5-20 μm length).

B. Methodology

PDA and GPTMS were used in the surface modification of the CNTs/GNs and BN/BT nanoparticles, respectively. Typically, the process involved dispersing CNTs and GNs in beakers with distilled water and ultrasonicate for four hours at 80°C (with addition of PDA and ammonia solution in drops). While BN and BT were introduced in beakers with xylene and ultrasonicate for 4 hours at 80ºC (with addition of GPTMS). After an additional 10 hours of reaction and ultrasonication, all the suspensions were washed severally with distilled water to obtain surface functionalized CNTs, GNs, BN and BT. By hydrothermal method, the CNTs and GNs were separately mixed with BT and BN in distilled water, which are denoted as BTC and BNG respectively. To achieve homogeneous dispersion, the mixes were ultrasonically treated for four hours at 80ºC. The suspensions were then covered with aluminum foil and kept in a 140°C oven for 10 hours to allow the nanoparticles to self-assemble. For the fabrication of the dielectric nanocomposites, PPMA was dissolved in xylene and mixed with the self-assembled BT/CNTs (BTC) and BN/GNs (BNG) nanoparticles at 140°C to prepare PPMA/nanoparticles masterbatch. Each concentration of the masterbatch contained 8 wt% PPMA. PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC dielectric nanocomposites were developed by melt mixing using rheomixer at 190°C temperature and 100 rpm screw rotation speed for 10 min. The dielectric nanocomposites were then granulated and compressed for 10 minutes at 200ºC and 10 MPa using a carver presser. Pure PP was also developed using the above-described procedures for comparison.

C. Characterization and measurement

A transmission electron microscope (TEM) was used to examine the nanoparticles' microstructure (JEM-2100). A scanning electron microscope (SEM) was used to morphologically investigate the nanocomposites (VEGA 3 TESCAN). Using an LCR Meter (B&K 891), the dielectric characteristics of the nanocomposites were assessed over the frequency range of 100 Hz to 10 kHz. A Conelectric BS 3941 high voltage transformer was utilized to detect the breakdown voltage. Utilizing thermogravimetry analyzer (TGA) with TA equipment Q500 at a heating rate of 10 ºC/min in an inert atmosphere, the thermal stability of the nanocomposites was investigated.

III. RESULTS AND DISCUSION

A. Microstructural analysis

CNTs with a 1D structure are seen in the TEM image in Fig. 1a. Additionally, it exhibits a lengthy, intertwined nanotube dimension. This aids in network structural formation in a polymer matrix and improvement in the thermal property. In Fig. 1b, the TEM micrograph shows the large surface of GNs with a 2D construction. Since CNTs and GNs are high aspect ratio, large surface area, and conductive nanomaterials, they can greatly enhance polymer's thermal and dielectric characteristics. In Figs. 1c and 1d, the BN and BT nanoparticles depict stacked 2D and 0D structures respectively. While BN is a high thermal conductive and insulator nanomaterial, BT is a high dielectric constant and insulator nanomaterial. Therefore, their combination with the conductive GNs and CNTs improved the dielectric and thermal stability of the PP. The utilization of 0D, 1D and 2D nanomaterials for this investigation is thus confirmed by the TEM examination.

Fig. 2 depicts the morphological structures of the dielectric nanocomposites, which show how evenly the nanoparticles were distributed throughout the PP matrix. There was no visible aggregation of the particles in the microstructures of PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites as depicted in Figs. 2b and 2c respectively. Prior research has shown that GNs and CNTs cooperate to facilitate efficient distribution in a polymer matrix [14]. While multiple wall-to-wall interactions between CNT walls were reduced by GNs flakes, CNTs tubes reduced interactions between GNs layers.

The addition of BN and BT nanoparticles in the PP system can also be credited with contributing to the uniform morphology of the nanocomposites. The uniform dispersion of the particles in the polymer matrix was obtained due to the BN and BT further reducing the contact between individual CNTs and GNs. As a result of the larger density of BT and BN than GNs and CNTs, the dielectric nanocomposites also displayed dense morphologies, particularly for PP/3wt%BNG-1wt%BTC nanocomposite (with better microstructure). This

observed microstructure is essential in promoting the thermal property and dielectric energy of the nanocomposites. Meanwhile, pure PP showed smooth microstructure since it contained no foreign constituents as displayed in Fig. 2a.

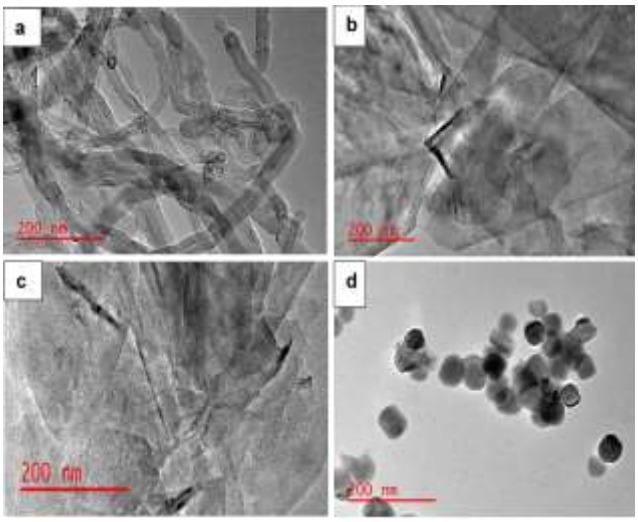


Fig. 1. TEM Images of (a) CNTs (b) GNs (c) BN and (d) BT

B. Thermal stability of the dielectric nanocomposites

The TGA curve representing the reduction in weight of the dielectric nanomaterials as the heating was increasing is presented in Fig. 3a. The figure reveals that the materials have two stages of thermal decomposition. It first exhibited relatively good thermal stability up to about 390°C, 437°C and 418°C, then thermal decomposition started with a fast slope till final degradation temperature of about 460°C, 502°C and 498°C for the pure PP, PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC respectively. The shifting of the TGA curves of the nanocomposites to the higher temperatures indicates increase in thermal stability, which is because of the thermal barriers provided by the nanoparticles [15]. The increased capability to withstand thermal energy of the composites suggests that the presence of the 0D, 1D and 2D nanoparticles used in this study have significantly stiffened the PP matrix.

Due to the uniform distribution and interfacial interaction between the nanoparticle and PP matrix, higher thermal energy was required to decompose PP matrix – nanoparticle bonds compared to the only pure PP chains. Hence, there was an increase in thermal stability of the nanocomposites. In general, optimal onset temperature of about 47°C increase was obtained for PP/1wt%BNG-3wt%BTC nanocomposite relative to the pure polymer. This enhancement in thermal stability is essential for high temperature application of the dielectric materials. The temperature at which the rate of weight loss is at highest is illustrated by the derivative of weight loss as a function of temperature in Fig. 3b. The figure shows the temperature at which maximum weight loss occurred to be around 440°C, 478°C and 472°C for the pure PP, PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites respectively. This means that the maximum weight reduction of the pure PP occurred at a lower temperature (about 38°C lower) than the developed dielectric nanocomposites.

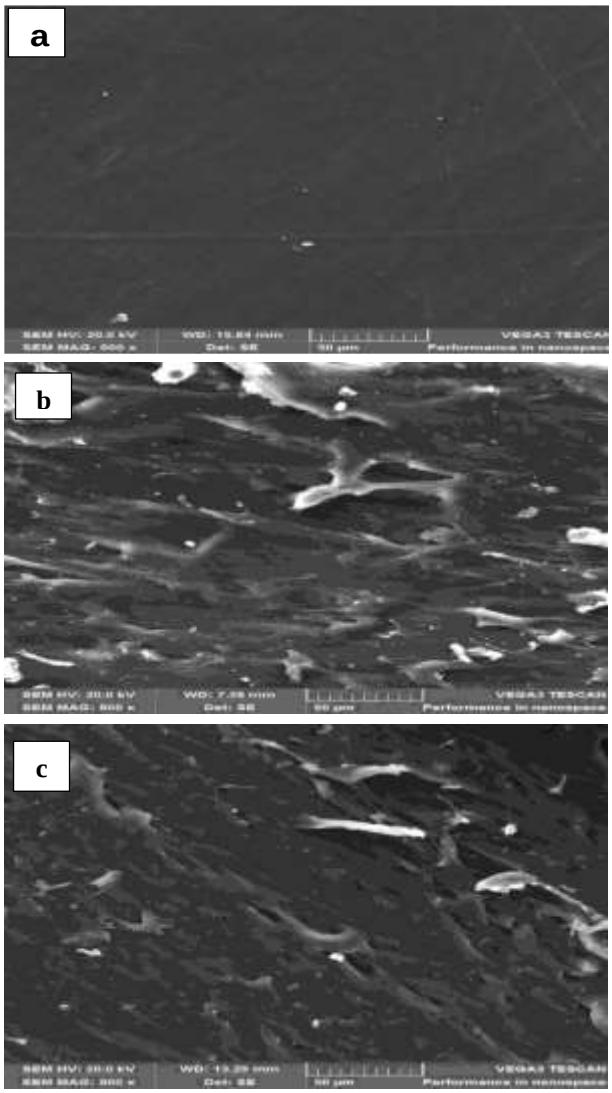


Fig. 2. SEM Images of (a) Pure PP (b) PP/1wt%BGN-3wt% BTC and (c) PP/3wt%BGN-1wt% BTC Nanocomposites.

C. Dielectric and energy density of the dielectric nanocomposites

The dielectric constant of the dielectric nanomaterials is presented in Fig. 4a. Dielectric constant is linearly proportional to its capacitance and energy stored. The figure shows that the dielectric constant decreased as frequency increases. This is because the four primary polarizations—electronic, ionic, dipole, and charge space polarization are all effective and contribute to a material's dielectric constant at low-frequency. A decrease in dielectric constant at high frequencies is caused by some polarization processes' insufficient contribution [16]. The dielectric constant improved to 93.8 and 102.5 for PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites against 2.02 for the pure PP respectively. The results suggest that the addition of the particles into the polymer matrix improved its dielectric property. This is possible since the separation of neighbouring conductive nanoparticles (GNs and CNTs) by the PP insulative layers and insulative

nanoparticles (BN and BT) can lead to the creation of nanocapacitors in the polymer system [17].

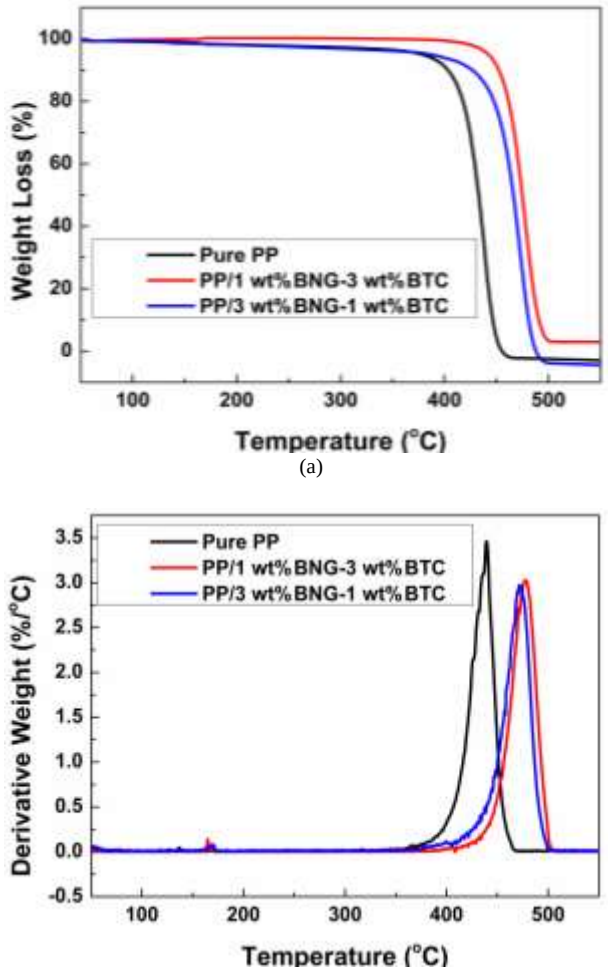


Fig. 3. TGA Curves of (a) Weight Loss and (b) Derivative Weight of the Nanocomposites.

The formation of nanocapacitors in the PP increased the dielectric constant of the nanomaterials by accumulating charges at the point of contact between the matrix and the nanoparticles, following polarization theory. Although all the developed dielectric nanocomposites displayed higher dielectric constants than pure PP, the difference was greater for PP/3wt%BNG-1wt%BTC nanocomposite compared to PP/1wt%BNG-3wt%BTC nanocomposite. This maybe because of the higher content of BNG nanoparticles (BN and GNs) which have a 2D (or layered) structure, which can form larger surface nanocapacitors in the PP matrix than BT and CNT nanoparticles, which have 0D and 1D structures respectively. Given that the nanoparticles were more evenly distributed for PP/3wt%BNG-1wt%BTC nanocomposite than PP/1wt%BNG-3wt%BTC nanocomposite, the former nanocomposite could have a larger dielectric constant than the latter nanocomposite. As a result of that, an optimal of about 4972.3% enhanced dielectric constant was achieved relative to the pure PP. This is significant and vital for the enhancement of the energy storage capacity of the dielectric nanocomposites.

This study also measured the electrical conductivity of the dielectric nanomaterials as presented in Fig. 4b. Most dielectric polymers are good electrical insulators, therefore, a low electrical conductivity of about 1.3 x 10^{-11}S/m at 100 Hz was obtained for the pure PP. The electrical conductivity increased as the frequency increases due to quick movement of charges, which favours electrical conductivity. It later increased to 4.5 x 10^{-9} S/m and 4.9 x 10^{-9} S/m for PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites respectively. Compared with the pure PP, it is about two-fold greater. The promoted electrical conductivity was because of the development of conductive structures, interconnection of the nanoparticles in the polymer [18]. This was anticipated since GNs and CNTs are conductive nanoparticles, and they have the capability to release charge carriers that can freely move in the polymer system with an increase in electrical conductivity [19]. However, the dielectric nanocomposites did not lose their insulative properties considering the range of electrical conductivity obtained. This is due to the insulative BT and BN in the PP matrix, which traps charges at the contact points of BN-GNs and BT-CNTs. Additionally, this contributed to the development of low-conductive network architectures and the limiting of charges' movement in the polymer system.

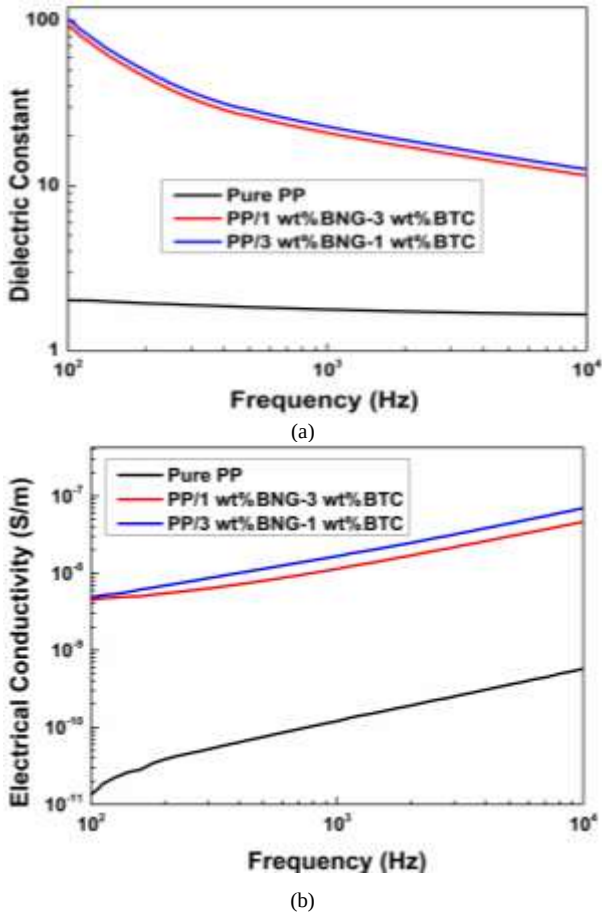


Fig. 4. (a) Dielectric constant and (b) electrical conductivity of the nanocomposites.

D. Breakdown strength/voltage and energy density of the dielectric nanocomposites

The breakdown voltage was taken from the maximum voltage at which the dielectric nanomaterials failed on the application of voltage. Breakdown strength of about 332 kV/mm was obtained for the pure polymer as shown in Fig. 5a, which confirmed the high voltage endurance of the pure PP and polymers in general [20]. However, by the incorporation of the particles into the polymer, breakdown strength decreased. This is because when electrical power was applied, the network structures formed by the nanoparticles in the matrix allowed current and charges to flow. In addition, concentrated electric field around the reinforcing phases in the polymer contributed towards the lowering of the breakdown strength of the dielectric nanomaterials due to the difference in conductivity between the matrix and the nanoparticles [21]. Hence, the nanocomposites experienced lower breakdown strength compared to the pure PP. However, appreciable breakdown voltage of about 81.6 kV/mm and 78.8 kV/mm were measured for PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites respectively.

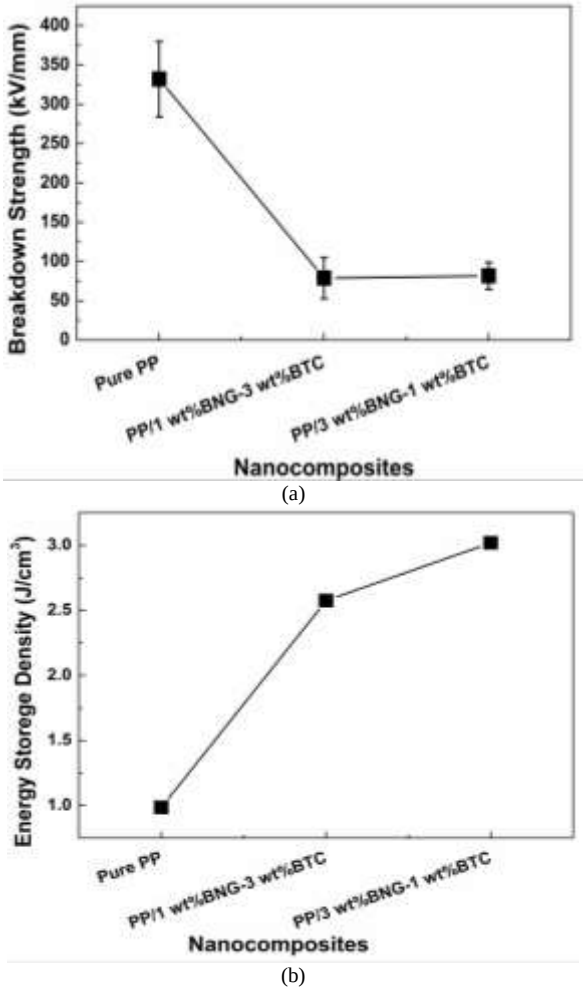


Fig. 5. (a) Breakdown strength and (b) energy storage density of the nanocomposites.

The energy density of the developed nanocomposites was much higher than that of the pure PP because of the balance obtained between the dielectric constant and breakdown voltage of the nanocomposites. For example, the pure PP has a low energy density of roughly 1.0 J/cm^3, which was enhanced to 2.6 J/cm^3 and 3.0 J/cm^3 for the PP/1wt%BNG-3wt%BTC and PP/3wt%BNG-1wt%BTC nanocomposites, respectively (see Fig. 5b). This is an optimal of about 200% increment. The low energy density of the pure PP despite its high breakdown strength is due to its low dielectric constant [20]. This indicates that both breakdown voltage and dielectric constant are essential factors in the enhancement of energy density of dielectric materials.

IV. CONCLUSION

In this study, functionalized conductive and insulative nanoparticles with various dimensional configurations such as 0D, 1D, and 2D were used to develop high-performance dielectric nanomaterials with promoted thermal response, dielectric property, and energy density. The hydrothermal and self-assembly approach was used to prepare the nanoparticles before they were used to fabricate the dielectric nanocomposites. By combining solution and melt mixing, the nanocomposites were developed. The microstructures of the nanoparticles and dielectric nanocomposites, which had uniform morphologies, were revealed by the TEM and SEM investigations, respectively. The dielectric nanocomposites showed an increase in thermal stability with onset temperature of about 47ºC increment relative to the pure polymer. The developed dielectric nanocomposites also demonstrated improvement in dielectric constant and energy density. Where about 4927.3% and 200% enhancements in dielectric constant and energy density were achieved. The processing strategy and nanomaterials' combination adopted in this study resulted in the creation of advanced dielectric nanocomposites, which can find advanced applications in dielectric energy storage and other related applications.

ACKNOWLEDGMENT

We are grateful for the support offered by the Tshwane University of Technology's Faculty of Engineering and the Built Environment and the Centre for Energy and Electric Power.

REFERENCES

[1] B. C. Riggs, S. Adireddy, C. H. Rehm, V. S. Puli, R. Elupula, and D. B. Chrisey, "Polymer Nanocomposites for Energy Storage Applications," *Materials Today: Proceedings*, vol. 2, no. 6, pp. 3853-3863, 2015, doi: 10.1016/j.matpr.2015.08.004.

[2] S. A. Sherrill, P. Banerjee, G. W. Rubloff, and S. B. Lee, "High to ultra-high power electrical energy storage," *Physical Chemistry Chemical Physics*, vol. 13, no. 46, pp. 20714-23, Dec 14 2011, doi: 10.1039/c1cp22659b.

[3] H. Peng, X. Sun, W. Weng, and X. Fang, *Energy storage devices based on polymers*, 1st ed. (Polymer Materials for Energy and Electronic Applications). Elsevier, 2017, p. 373.

[4] U. O. Uyor, A. P. Popoola, O. Popoola, and V. S. Aigbodion, "Energy storage and loss capacity of graphene-reinforced poly (vinylidene fluoride) nanocomposites from electrical and dielectric properties perspective: A review," *Advances in Polymer Technology*, vol. 37, no. 8, pp. 2838-2858, 2018.

[5] J. Li, P. Khanchaitit, K. Han, and Q. Wang, "New route toward high-energy-density nanocomposites based on chain-end functionalized ferroelectric polymers," *Chemistry of Materials*, vol. 22, no. 18, pp. 5350-5357, 2010.

[6] P. Kim *et al.*, "High energy density nanocomposites based on surface-modified BaTiO3 and a ferroelectric polymer," *ACS nano*, vol. 3, no. 9, pp. 2581-2592, 2009.

[7] Z.-M. Dang, J.-K. Yuan, J.-W. Zha, T. Zhou, S.-T. Li, and G.-H. Hu, "Fundamentals, processes and applications of high-permittivity polymer–matrix composites," *Progress in Materials Science*, vol. 57, no. 4, pp. 660-723, 2012/05/01/ 2012, doi: https://doi.org/10.1016/j.pmatsci.2011.08.001.

[8] Q. Wang and L. Zhu, "Polymer nanocomposites for electrical energy storage," *Journal of Polymer Science Part B: Polymer Physics*, vol. 49, no. 20, pp. 1421-1429, 2011, doi: 10.1002/polb.22337.

[9] F.-C. Chiu and Y.-J. Chen, "Evaluation of thermal, mechanical, and electrical properties of PVDF/GNP binary and PVDF/PMMA/GNP ternary nanocomposites," *Composites Part A: Applied Science and Manufacturing*, vol. 68, pp. 62-71, 2015, doi: 10.1016/j.compositesa.2014.09.019.

[10] W.-b. Zhang *et al.*, "Largely enhanced thermal conductivity of poly(vinylidene fluoride)/carbon nanotube composites achieved by adding graphene oxide," *Carbon*, vol. 90, pp. 242-254, 2015, doi: 10.1016/j.carbon.2015.04.040.

[11] L. Chu *et al.*, "Porous graphene sandwich/poly(vinylidene fluoride) composites with high dielectric properties," *Composites Science and Technology*, vol. 86, pp. 70-75, 2013, doi: 10.1016/j.compscitech.2013.07.001.

[12] J. Sun, Q. Xue, Q. Guo, Y. Tao, and W. Xing, "Excellent dielectric properties of Polyvinylidene fluoride composites based on sandwich structured MnO2/graphene nanosheets/MnO2," *Composites Part A: Applied Science and Manufacturing*, vol. 67, pp. 252-258, 2014, doi: 10.1016/j.compositesa.2014.09.006.

[13] Y. Li *et al.*, "Polydopamine coating layer on graphene for suppressing loss tangent and enhancing dielectric constant of poly(vinylidene fluoride)/graphene composites," *Composites Part A: Applied Science and Manufacturing*, vol. 73, pp. 85-92, 2015, doi: 10.1016/j.compositesa.2015.02.015.

[14] C. Min *et al.*, "Unique synergistic effects of graphene oxide and carbon nanotube hybrids on the tribological properties of polyimide nanocomposites," *Tribology International*, vol. 117, pp. 217-224, 2018.

[15] M. Nurul and M. Mariatti, "Effect of thermal conductive fillers on the properties of polypropylene composites," *Journal of Thermoplastic Composite Materials*, vol. 26, no. 5, pp. 627-639, 2013.

[16] N. Shukla and D. Dwivedi, "Dielectric relaxation and AC conductivity studies of Se90Cd10− x In x glassy alloys," *Journal of Asian Ceramic Societies*, vol. 4, no. 2, pp. 178-184, 2016.

[17] Z. Wang *et al.*, "Ultrahigh dielectric constant and low loss of highly-aligned graphene aerogel/poly (vinyl alcohol) composites with insulating barriers," *Carbon*, vol. 123, pp. 385-394, 2017.

[18] J.-I. Lee, S.-B. Yang, and H.-T. Jung, "Carbon nanotubes− polypropylene nanocomposites for electrostatic discharge applications," *Macromolecules*, vol. 42, no. 21, pp. 8328-8334, 2009.

[19] M. S. Cao *et al.*, "Electromagnetic response and energy conversion for functions and devices in low-dimensional materials," *Advanced Functional Materials*, vol. 29, no. 25, p. 1807398, 2019.

[20] B. Liu *et al.*, "High energy density and discharge efficiency polypropylene nanocomposites for potential high-power capacitor," *Energy Storage Materials*, vol. 27, pp. 443-452, 2020.

[21] X. Huang, P. Jiang, and T. Tanaka, "A review of dielectric polymer composites with high thermal conductivity," *IEEE Electrical Insulation Magazine*, vol. 27, no. 4, pp. 8-16, 2011, doi: 10.1109/mei.2011.5954064.

Adaptive Admittance Control for Physical Human-Robot Interaction Within Delay-Related Boundaries

Yuliang Guo
Hangzhou innovation institute
Beihang University
Hangzhou, China
guoyuliang@buaa.edu.cn

Jianwei Niu
Hangzhou innovation institute
Beihang University
Hangzhou, China
niujianwei@buaa.edu.cn

Renluan Hou
Hangzhou innovation institute
Beihang University
Hangzhou, China
jessierlhou@zju.edu.cn

Tao Ren
Hangzhou innovation institute
Beihang University
Hangzhou, China
taotao_1982@126.com

Bing Han
Hangzhou innovation institute
Beihang University
Hangzhou, China
A994055925@163.com

Xiaolong Yu
Hangzhou innovation institute
Beihang University
Hangzhou, China
yxiaolong@buaa.edu.cn

Qun Ma
Hangzhou innovation institute
Beihang University
Hangzhou, China
maqun@buaa.edu.cn

***Abstract*—Admittance control is an effective physical human-robot interaction approach, which designs the characteristics of the end-effector as a second-order system to achieve compliant performance. The performance for different admittance parameters compromise between reducing interaction effort and improving stability with stiff environment. This paper proposes an adaptive admittance control method based on an intuitive force differential index to detect unstable high-frequency oscillation and an integral adaptive law for admittance parameters. Furthermore, this paper presents the stable parameter boundaries considering system and sensor delays. Experiments are carried out on a ROKAE XB4 robot to validate the performance of unstable behavior detection and adaptive admittance control.**

Keywords— adaptive admittance control, physical human-robot interaction, delay, industrial robot

I. INTRODUCTION

An industrial robot is fast becoming a key instrument in automation, due to its overwhelming precision and efficiency. In most applications, industrial robots work at position-controlled mode and focus on the position tracking of designed trajectories. However, as the interaction between industrial robots and people/environment are inevitable to expand its utilization potential, there has been an increasing interest in the force-related algorithms of industrial robots.

For position-controlled applications, trajectory planning is done in Cartesian space and then transformed into joint space to execute. When some of the degrees of freedom in Cartesian space require a force-related control strategy, force control law replaces the corresponding position control law to form a hybrid controller [1]. However, most commercial industrial robots only provide position interface to joint actuators, and hence position-based control law to accomplish force tracking is preferred, where the position-based control law means the output of controller is position. Previous research has established various position-based force control laws, with the assistant of force/torque (F/T) sensor. The common force control law includes PI (proportional-integral) control [2] and impedance control, which treats the end-effector as a second-order mass-spring-damper system [3]. The position-based impedance control is also known as admittance control [4].

Admittance control has been widely used in force-related applications, such as polishing, assembling, physical robot interaction, etc. For physical robot interaction applications, the inertia and damping parameters can be designed, while the stiffness parameter depends on human or other environment. The stiffness is hard to determined and varies due to diverse environment. Therefore, a fixed pair of inertia and damping parameters are usually quite conservative to fit different environment stiffness [5]. However, the low-stiffness interaction performance of the conservative admittance system is unsatisfied, as it requires too much effort. The admittance parameters compromise between reducing operator's effort and improving stability with high-stiffness environment. Therefore, variable admittance control with instability detection were proposed in the previous literature. An online FFT analysis was proposed in [5], [6] to detect high-frequency oscillations in force measurements. However, the FFT usually require hundreds of data and introduce considerable delay. Moreover, the duration of the unstable behavior is uncertain, and hence a fixed window FFT may not clearly distinguish the high frequency components. Various instability indexes were proposed in [7]-[9], and the system passivity is based on energy-tank based analyses [10]. However, some of the indexes were related to acceleration, which is hard to measure. Moreover, the parameter variation in [7] could not decrease, while the parameter variations in [8], [9] were not continuous, which may result into undesired behavior.

This paper proposes a novel adaptive admittance control method based on an intuitive instability index and an integral adaptive law. To capture the high-frequency oscillations, the force differential is utilized as the instability index. The variable admittance parameters are designed to increase proportional to the absolute value of the force differential and to decrease constantly when external forces are not detected. The adaptive

law is based on integral and hence the parameter variations are continuous. Besides, this paper presents the stable parameter boundaries to facilitate the parameter selection when considering system and sensor delays.

II. CONVENTIONAL ADMITTANCE CONTROL

Admittance control, also known as position-based impedance control, is an effective force-related control strategy. Most commercial robots adopt admittance control rather than force-based impedance control, due to: (1) force control interface is not provided; (2) the precision of position tracking is assured; (3) force-based control face noncolocated problem, i.e. the F/T sensor and joint motor are not placed close to each other.

The characteristic of the end-effector is equivalent to a second-order mass-spring-damper system. The interaction with environment (human included) is considered as a stiffness system in this paper. Fig. 1 shows the equivalent system model of robot end-effector and environment. The environment stiffness is denoted by k_e, while the mass, damping and stiffness of end-effector are denoted by m, b and k, respectively. Free space and contact space can be divided according to the contact state [11], [12].

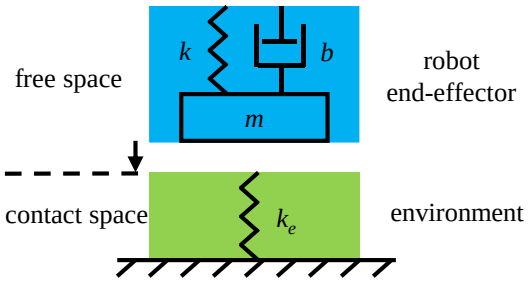


Fig. 1. The equivalent system model of robot end-effector and environment.

When a reference force is set, the end-effector is driven from free space to contact space until the contact force meets the target force. For simplicity, we assume the reference force is exerted in only one direction. Denote f_r and f_e as reference force and environment force, respectively. For simplicity, the external forces exerted by human are also treated as reference force f_r. Denote x, x_r and x_e as current position, reference position and environment position, respectively. The contact force is equal to the environment force $f_e = k_e (x - x_e)$. The admittance control law can be expressed as:

$$\Delta f = m\ddot{e} + b\dot{e} + ke \tag{1}$$

Where $\Delta f = f_r - f_e$ is the force tracking error, $e = x - x_e$ is the environment variation. In steady-state, the force tracking error Δf should be equal to 0. Therefore, the stiffness k is selected as 0 to remove the steady-state error according to (1). As a result, the control law in free space and contact space is given by:

$$\begin{cases} f_r = m\ddot{e} + b\dot{e}, & \text{free space} \\ f_r = m\ddot{e} + b\dot{e} + k_e e, & \text{contact space} \end{cases} \tag{2}$$

The Laplace transfer function is expressed as:

$$G(s) = \frac{e(s)}{f_r(s)} = \begin{cases} 1/(ms^2 + bs) & \text{free space} \\ 1/(ms^2 + bs + k_e) & \text{contact space} \end{cases} \tag{3}$$

Fig. 2 presents the schematic of ideal admittance control. The admittance control model generates position compensation for the reference position. Then the position command is realized through inverse kinematic and joint actuator. The feedback force is measured by an F/T sensor and the value is approximately the multiply of environment stiffness and the difference between environment position and real position.

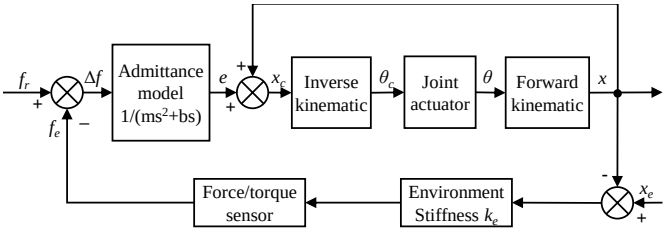


Fig. 2. Ideal admittance control schematic.

III. PARAMETER BOUNDARIES CONSIDERING DELAYS

This section extends the parameter selection of admittance control for the industrial robot from ideal model to real model considering delays. Parameter constraints are presented to ensure the stability and the performance of the real admittance control system.

A. Ideal Admittance Model

For an ideal admittance model, the transfer function (3) is used. In either space, stability is guaranteed when damping coefficient b larger than 0. In free space, the steady-state velocity for the end-effector approaching environment is calculated as $v = f_r / b$. In contact space, the system is an ideal second-order system, with a natural frequency $\omega_n = \sqrt{k_e / m}$, and a damping ratio $\zeta = 0.5b / \sqrt{mk_e}$. The ideal contact performance will not introduce overshoot when the damping ratio is larger or equal to 1. Therefore, a compromise is made only between approaching velocity in free space and damping ratio in contact space.

B. Real Admittance Model Considering Delays

The ideal admittance model in Fig. 2 cannot be realized, as the artificial mass-damper-spring system requires position, velocity and force feedback. Fig. 3 presents a typical discretized schematic of admittance control. The feedback delays include: (1) n-period communication delay between the motor driver and the controller; (2) the equivalent delay caused by low pass filter (LPF) for F/T sensor.

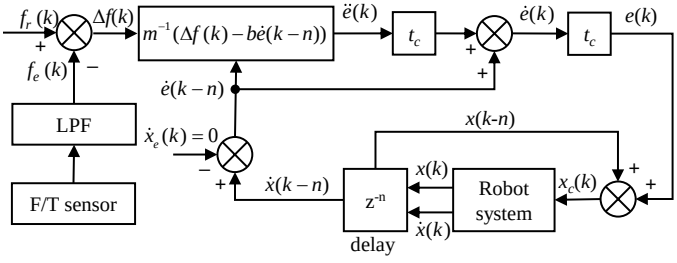


Fig. 3. A typical discretized schematic of admittance control.

1) *Free Space*

In free space, the feedback from the F/T sensor is expected to be zero. Backward Euler method is utilized for integral in this paper. The following analyses can be expanded to other discretization methods. Therefore, the equivalent transfer function block diagram from reference force to position variation is shown in Fig. 4.

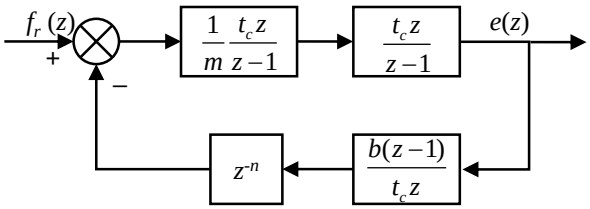


Fig. 4. Equivalent transfer function block diagram from reference force to position variation in free space.

The transfer function $G(z)$ is given by:

$$G(z)=\frac{e(z)}{f_r(z)}=\frac{m^{-1}t_c^2 z^2}{(z-1)(z^{n+1}-z^n+m^{-1}t_c bz)} \tag{4}$$

The poles should be inside the unit circle to ensure closed-loop stability. Normally, a stable continuous system is always stable after discretization using Backward Euler method. However, the inevitable delays may lead to instability. For a discrete system, stability is guaranteed when all poles are inside the unit circle. The unit circle can be mapped to the left half-plane by utilizing transformation:

$$z=\frac{1+w}{1-w} \tag{5}$$

Then the stability of the discrete system can be verified by the Routh criterion. Denote $k_b = m^{-1}t_c b$ for simplicity. The stability relies on the second term of denominator $z^{n+1} - z^n + k_b z$.

The communication delay is usually 1 period. For n=1, the poles are determined by:

$$(2-k_b)w+k_b=0 \tag{6}$$

The stability condition yields $0 < k_b < 2$. Unlike the ideal continuous-time system in previous research, this paper reveals that the discrete system considering delays presents an upper boundary for parameter k_b. As the communication delays increase, the upper boundary decreases, due to a smaller phase margin.

2) *Contact Space*

The LPF after the F/T sensor is another delay source in contact space. The F/T sensor performs a smooth waveform with an adequate cutoff frequency, where a compromise needs to be made to balance the force feedback waveform and the delay. Fig. 5 shows the z-axis F/T waveforms comparison with different cutoff frequencies for Onrobot HEX-E F/T sensor. The waveform is close to the original waveform for f_c = 500Hz, while the waveform still contains obvious disturbance even for f_c = 15Hz. It is clear that a smoother performance is achieved when f_c =5Hz or 1.5Hz. As a result, the delay caused by the low pass filter can be obvious.

Fig. 6 shows the transfer function block diagram from reference force to position variation in contact space. For simplicity, the communication delay is neglected, as it is much smaller than LPF delays.

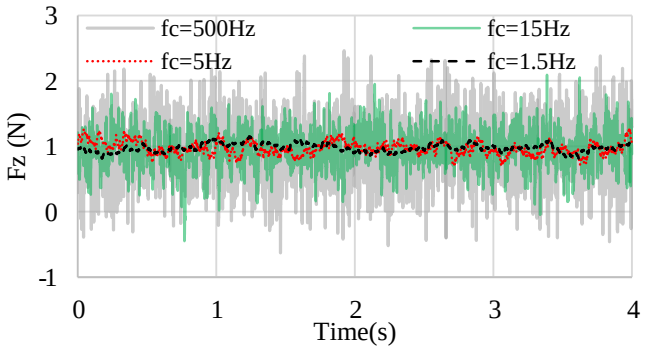


Fig. 5. Typical F/T waveforms comparison among different cutoff frequencies for Onrobot HEX-E F/T sensor.

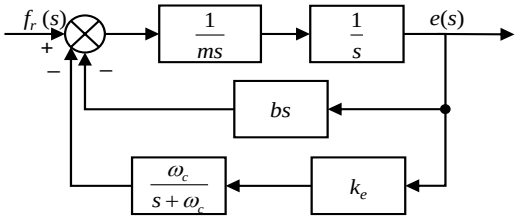


Fig. 6. Transfer function block diagram from reference force to position variation in contact space.

The transfer function $H(s)$ is given by:

$$H(s)=\frac{e(s)}{f_r(s)}=\frac{s+\omega_c}{ms^3+(m\omega_c+b)s^2+b\omega_c s+k_e\omega_c} \tag{7}$$

The Routh array is constructed as:

$$\begin{array}{lll} s^3 & m & b\omega_c \\ s^2 & m\omega_c+b & k_e\omega_c \\ s^1 & \dfrac{\omega_c b^2+m\omega_c^2 b-mk_e\omega_c}{m\omega_c+b} & 0 \\ s^0 & k_e\omega_c & 0 \end{array} \tag{8}$$

The first column should be positive, which yields:

$$b>0.5(\sqrt{m^2\omega_c^2+4mk_e}-m\omega_c) \tag{9}$$

According to (9), the lower boundary for b is affected by LPF cutoff frequency ω_c and environmental stiffness k_e. The lower boundary becomes considerably large along with stiffening physical interaction, when cutoff frequency ω_c is low and environmental stiffness k_e is large.

In conclusion, the stability of the ideal admittance control only requires $b > 0$. After considering the effect of delays, an upper boundary of b is acquired through the transfer function (4)

in free space and a lower boundary of b is acquired through transfer function (7) in contact space.

IV. ADAPTIVE ADMITTANCE CONTROL

The above section shows the boundaries of admittance control parameter. Inside the stability boundaries, the advantages of a smaller value b contain a larger approaching velocity in free space and a reducing effort during physical human-robot interaction, while the disadvantages contain a smaller damping ratio in contact space and a higher risk of instability during interaction with stiffening environment.

This paper proposes a novel adaptive admittance control method to avoid the disadvantages. The symbol of unstable behavior is the high frequency oscillations in force or position measurements. Hence, this paper proposes to utilize the differential of force measurements as an intuitive index γ to denote the high-frequency oscillations, i.e.

$$\gamma = \begin{cases} 1 & , \ \left\| \frac{dF_e}{dt} \right\| \geq \delta \\ 0 & , \ \text{otherwise} \\ -1 & , \ \left\| \frac{dF_e}{dt} \right\| < \delta \text{ and } \left\| F_{ed} \right\| < F_0 \end{cases} \tag{10}$$

where δ is an empirical threshold to detect the unstable behavior, F_{ed} is the force measurement with a dead zone and F_0 is a small force threshold. High-frequency oscillations are detected when γ equal to 1. If the force differential is smaller than δ and force measurement is close to 0, the index γ is designed as -1. Otherwise, the index is equal to 0. The index γ is utilized to facilitate the adaptive control law.

According to the instability index γ, the adaptive law $G(\gamma)$ for admittance control parameters is designed as:

$$G(\gamma) = \begin{cases} k_1 \left\| \frac{dF_e}{dt} \right\| & , \ \gamma = 1 \\ 0 & , \ \gamma = 0 \\ k_2 & , \ \gamma = -1 \end{cases} \tag{11}$$

where k_1, $k_2 > 0$ are proportional parameters to adjust the adaptive law when the instability index γ is equal to 1 and -1, respectively.

The damping parameter variation $\Delta b(t)$ is designed as an integral function of index γ:

$$\dot{b}(t) = \gamma G(\gamma), \ \Delta b(t) = \int \dot{b}(t) dt \tag{12}$$

A saturation is utilized to constraint the integral. The upper limit Δb_m makes sure that the robot interaction will not be too heavy. The lower limit guarantees the stability with regular low stiffness interaction with environment or human.

$$\Delta \bar{b} = \begin{cases} 0 & , \ \Delta b(t) < 0 \\ \Delta b(t) & , \ 0 \leq \Delta b(t) \leq \Delta b_m \\ \Delta b_m & , \ \Delta b(t) > \Delta b_m \end{cases} \tag{13}$$

The adaptive admittance parameters are given by:

$$\begin{aligned} b(t) &= b_0 + \Delta \bar{b} \\ m(t) &= \varepsilon b(t) \end{aligned} \tag{14}$$

where b_0 is the initial value of admittance damping and ε is the ratio between inertia and damping parameters. In the adaptive law, the increased damping parameter acts as an effective energy dissipater, when instability is detected. The damping parameter returns to initial value, when the external force is removed. Otherwise, the damping parameter remains unchanged. The inertia parameter varies simultaneously with the damping parameter, which is a practical design to maintain a similar system dynamics during adaption process [13]. Fig. 7 shows the overall block diagram of the proposed adaptive admittance control algorithm.

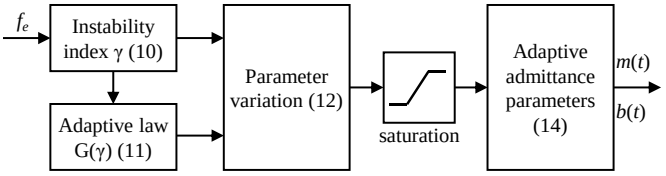


Fig. 7. Block diagram of the proposed adaptive admittance control algorithm.

V. EXPERIMENTAL RESULTS

A typical commercial 6-axis industrial robot ROKAE XB4 [14] is utilized to validate the proposed adaptive admittance control as shown in Fig. 8. The modified Denavit-Hartenberg (DH) parameters of XB4 are listed in Table I. The motor drivers are working at position-controlled mode. The motor drivers communicate with an industrial panel PC (IPC) with Linux OS through EtherCAT bus. The admittance control algorithms are developed within Matlab/Simulink and deployed to the IPC. The controlling period t_c is 1ms. An Onrobot HEX-E 6-axis force/torque sensor [15] is mounted on the end-effector with a 500Hz sampling frequency. The F/T measurements are collected in the sensor coordinate and then converted to the end-effector coordinate, where gravity terms are compensated. Robot programming can be simplified through physical interaction between the sander on the end-effector and the human or environment.

TABLE I. MODIFIED DH PARAMETERS OF ROKAE XB4

axis	α_{i-1} **(rad)**	a_{i-1} **(m)**	d_i **(m)**	θ_i **(rad)**
1	0	0	0.342	θ_1
2	-π/2	0.04	0	θ_2-π/2
3	0	0.275	0	θ_3
4	-π/2	0.025	0.28	θ_4
5	π/2	0	0	θ_5
6	-π/2	0	0.073	θ_6
7 (load)	0	0	0.168	0

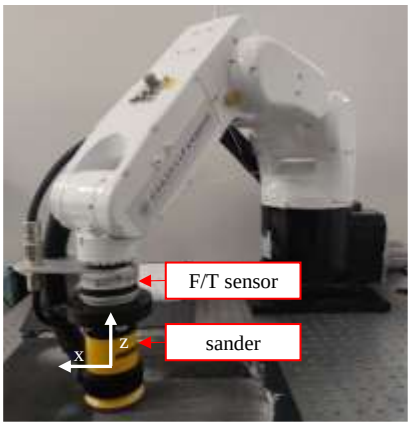


Fig. 8. ROKAE XB4 robot for force tracking experiments, with an Onrobot HEX-E force/torque sensor and a sander mounted as the end-effector.

In the following experiments, the empirical threshold δ is selected as 120. During robot interaction, a small force dead zone 1N is utilized to prevent undesired motion due to sensor offset or noise. The force threshold F_0 is selected as 0.001Nm. The proportional parameters k_1 and k_2 are 0.03 and 0.3, respectively. The initial value of damping parameter b_0 is selected as 50 and the ratio ε is set to 0.1. The upper limit Δb_m is equal to 1000.

The experiment scenario is a physical robot-human interaction of direct teaching, where the operator directly drags the robot to its working spot along z-axis. The inertia and damping parameters are 5 and 50, respectively. Fig. 9 presents the waveforms of z-axis force measurement F_z and the absolute value of its differential, together with the selected threshold δ, without adaptive admittance control. Fig. 10 presents the waveforms of the corresponding z-axis position and velocity. The low-stiffness robot-human interaction happens before 3.3s and the corresponding force differential is below the empirical threshold δ. The unstable behavior is immediately detected, once the robot contacts the high-stiffness environment surface. Although the damping parameter b_0 is beneficial to reduce the operator's effort, the end-effector bounces back from the stiff environment surface about 17mm, which is intolerable for direct teaching.

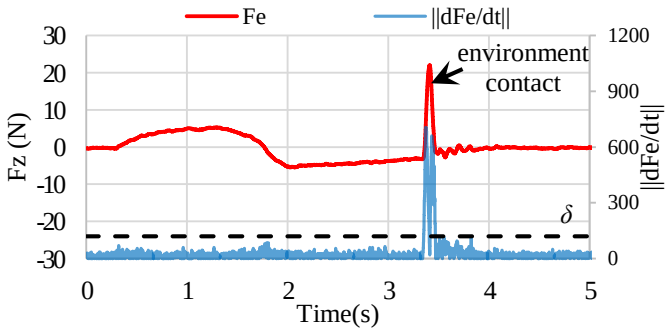


Fig. 9. Waveforms of z-axis force measurement F_z and the absolute value of its differential, together with the selected threshold δ, during a physical robot-human interaction scenario of direct teaching to approach a high-stiffness environment surface without adaptive admittance control.

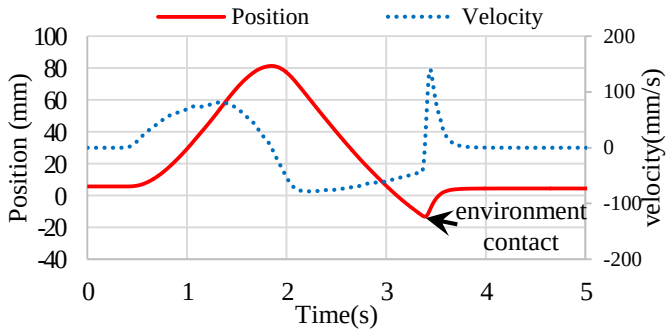


Fig. 10. Waveforms of the corresponding z-axis position and velocity without adaptive admittance control, the high-stiffness contact happens near the lowest point of the position waveform.

The proposed adaptive admittance control is implemented in the same scenario with the same initial parameters $m(0) = 5$ and $b(0) = 50$. Fig. 11 presents the waveforms of z-axis force measurement F_z and the absolute value of its differential, together with the selected threshold δ, with adaptive admittance control. Fig. 12 presents the waveforms of the corresponding position and velocity. Fig. 13 presents the synchronous waveforms of instability index γ and damping parameter variation Δb. Before 3.6s, the low stiffness robot-human interaction takes place. In this period, the force differential is always below threshold δ and the instability index is never equal to 1. Hence, the interaction effort maintain relatively small until the environment contact. The instability index is equal to 1 when the end-effector contact the high-stiffness surface. The adaptive law increases the damping parameter to attenuate the high-frequency oscillations. Therefore, the end-effector does not bounce back from the stiff surface obviously. Compared with the original method, the adaptive admittance control is better at dealing with various environment stiffness. As the adaptive law is based on integral, the damping parameter variation is continuous. The damping parameter gradually returns to initial value to facilitate the next usage, when no external force is detected and the instability index is equal to -1.

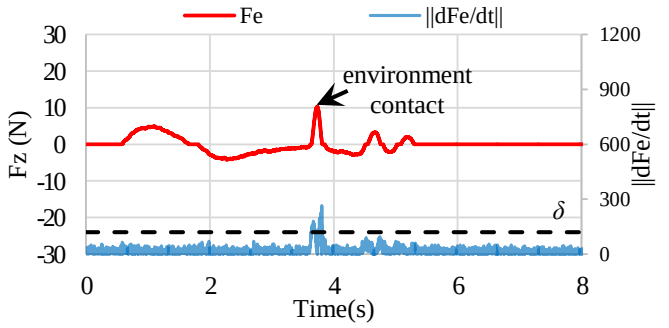


Fig. 11. Waveforms of z-axis force measurement F_z and the absolute value of its differential, together with the selected threshold δ, with adaptive admittance control..

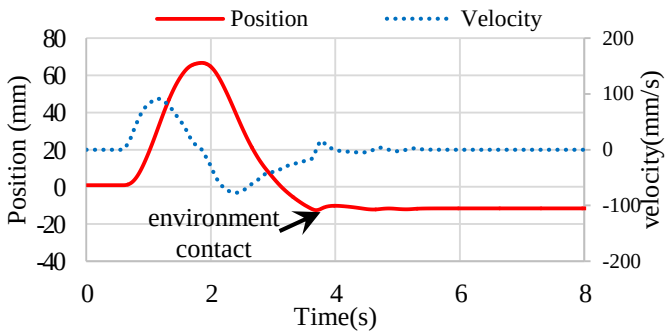


Fig. 12. Waveforms of the corresponding position and velocity with adaptive admittance control.

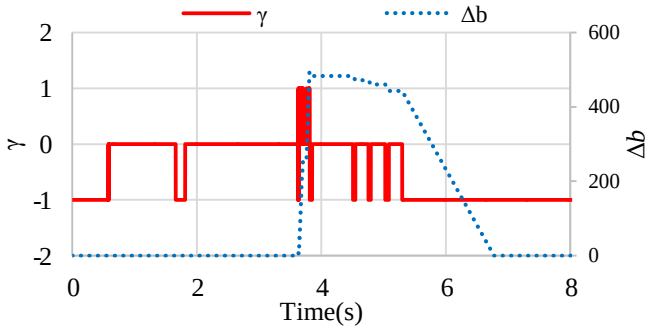


Fig. 13. Synchronous waveforms of instability index γ and damping parameter variation Δb.

VI. CONCLUSION

This paper discloses the stable parameter boundaries of admittance control for industrial robots. After considering system delay and sensor delay, this paper presents upper and lower boundaries for parameter selection both in free space and contact space. This paper further proposes an adaptive admittance control method based on a high-frequency instability index and an integral adaptive law. The experiments implemented on ROKAE XB4 presents the adaptive admittance performance of physical robot interaction from low-stiffness human to high-stiffness environment, compared with a fixed parameter algorithm.

ACKNOWLEDGMENTS

This work was supported by the Zhejiang Province Key R&D Program (2020C01026).

REFERENCES

[1] M. H. Raibert and J. J. Craig, "Hybrid Position/Force Control of Manipulators," Journal of Dynamic Systems, Measurement, and Control, vol. 103, no. 2, pp. 126-133, 1981.

[2] A. Winkler, J. Suchy, "Explicit and Implicit Force Control of an Industrial Manipulator - An Experimental Summary," in Proc. Int. Conf. methods and models in automation and robotics, 2016, pp.19-24.

[3] N. Hogan, "Impedance Control: An Approach to Manipulation: Part III—Applications," Journal of Dynamic Systems, Measurement, and Control, vol. 107, no. 1, pp. 17-24, 1985.

[4] C. Ott, R. Mukherjee and Y. Nakamura, "Unified Impedance and Admittance Control," in Proc. IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 2010, pp. 554-561.

[5] F. Dimeas and N. Aspragathos, "Online Stability in Human-Robot Cooperation with Admittance Control," *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 267-278, 2016.

[6] D. Ryu, J. Song, S. Kang, and M. Kim, "Frequency domain stability observer and active damping control for stable haptic interaction," IET Control Theory Appl., vol. 2, no. 4, pp. 261–268, Apr. 2008.

[7] C. T. Landi, F. Ferraguti, L. Sabattini, C. Secchi, and C. Fantuzzi, "Admittance control parameter adaptation for physical human-robot interaction," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2911-2916.

[8] C. T. Landi, F. Ferraguti, L. Sabattini, C. Secchi, M. Bonfè, and C. Fantuzzi, "Variable admittance control preventing undesired oscillating behaviors in physical human-robot interaction," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3611-3616.

[9] P. Cao, Y. Gan, J. Duan, and X. Dai, "Passivity-Based Stable Human-Robot Cooperation with Variable Admittance Control," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2019, pp. 446-451.

[10] F. Ferraguti, N. Preda, A. Manurung, M. Bonfè, O. Lambercy, R. Gassert, R. Muradore, P. Fiorini, and C. Secchi, "An Energy Tank-Based Interactive Control Architecture for Autonomous and Teleoperated Robotic Surgery," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1073-1088, 2015.

[11] Seul Jung, T. C. Hsia and R. G. Bonitz, "Force tracking impedance control of robot manipulators under unknown environment," IEEE Transactions on Control Systems Technology, vol. 12, no. 3, pp. 474-483, May 2004.

[12] J. Duan, Y. Gan, M. Chen, and X. Dai, "Adaptive variable impedance control for dynamic contact force tracking in uncertain environment," Robotics and Autonomous Systems, vol. 102, pp. 54-65, 2018 APR 2018.

[13] A. Lecours, B. Mayer-St-Onge, and C. Gosselin, "Variable admittance control of a four-degree-of-freedom intelligent assist device," in *2012 IEEE International Conference on Robotics and Automation*, 2012, pp. 3903-3908.

[14] ROKAE. XB4, 2021. [Online]. Available: https://www.rokae.com/pro_detail/4.html

[15] Onrobot. HEX force/torque sensor, 2021. [Online]. Available: https://onrobot.com/en/products/hex-6-axis-force-torque-sensor

Development of a New Tool for Voltage Stability Analysis in a Free and Open Source Software Package for Power System Studies

Samuel Souto de Oliveira
Faculty of Electrical Engineering
Federal University of Uberlândia
Uberlândia, Brazil
samuel.souto@ufu.br

Geraldo Caixeta Guimarães
Faculty of Electrical Engineering
Federal University of Uberlândia
Uberlândia, Brazil
gcaixeta@ufu.br

Thales Lima Oliveira
Faculty of Electrical Engineering
Federal University of Uberlândia
Uberlândia, Brazil
thales@ufu.br

***Abstract*— In recent years, several voltage stability indices have been developed and implemented in electrical power systems simulation software to help the analysis of the voltage instability problem, resulting in the creation of new computational tools capable of improving the system planning and operation. Following this proposal, the present work aims to implement three voltage stability indices within the specific software package developed at Federal University of Uberlandia and to verify their performance, when submitted to dynamic analysis. The study is applied in two cases. The first is a 3-bus radial system and the second the 14-bus IEEE system. The tests are conducted through successive increments of load on a given bus until its voltage becomes unstable.**

***Keywords*— Voltage stability indices; Power systems; Dynamic simulator; Sensitivity analysis; PSP-UFU.**

I. INTRODUCTION

According to the joint task force, IEEE/*CIGRÉ* Voltage stability refers to the ability of a power system to maintain steady voltages at all buses in the system after being subjected to a disturbance from a given initial operating condition[1]. Thus, it is essential to obtain indices that aim to facilitate the analysis of the voltage stability of an electrical system. These coefficients should be able to measure the proximity of voltage instability or collapse, the maximum amount of load for each node, and identify sensitive points to instability, helping in the planning and operation of the system. The indices proposed in this study are subjected to dynamic analysis, thus enabling a more detailed and accurate study of the electrical system. This makes it possible to verify how load changes impact voltage stability, and to what degree [1][2][3].

In this context, the present work aims to implement a new, tool called VSI (Voltage Stability Indices), in the Power System Platform of Universidade Federal de Uberlândia (PSP-UFU). In fact, it is a free and open-source software package that uses the C++ programming language and currently performs the following studies: power flow, short circuit, harmonics, transient and dynamic stability. The VSI will be developed in the context of the latter. Other computer programs such as MATACDC, MatDyn, MATPOWER, PSAT, VST, OpenDSS, PandaPower and GridCal, are cited in the technical literature as alternatives for electrical power system simulators. It is noteworthy, however, that few are used to simulate voltage collapse indicators and an even smaller number are free and open source, making the PSP-UFU a great tool to be used in the context of teaching and research [4][5].

II. FORMULATION OF INDICES

Since 1920, power systems have been monitored with the help of indices that verify voltage stability, seeking greater levels of system safety and reliability. The most recent methods applied worldwide are singular value decomposition, energy function, continued power flow, sensitivity analysis methods, bifurcations theory, minimum eigenvalue, integrated transmission line transfer index (ITLTI), etc. Each method has its unique set of benefits and drawbacks and can only be used for certain types of systems. For example, some methods are designed to analyze buses, lines, or a system, while others are designed to compare different systems. Therefore, it is important to carefully analyze each method before deciding which one to use [1][2][3][6].

The following paragraph presents three sensitive indicators that result from the relationship between increments or variations of voltage, active power, reactive power, and apparent power [7][8][9].

The first indicator, $VSI1_i = \Delta V_i/\Delta Q_{ci}$, is obtained through the ratio between the voltage decrease, DV_i, at the i-th bus, by the increase in reactive power consumed by the same bus DQ_{ci}

The second indicator, $VSI2_i = \Delta V_i/\Delta P_{ci}$, is defined as the quotient of the voltage decrease DV_i, at the i-th bus, by the increase in active power by the same bus DP_{ci}.

The third indicator, $VSI3_i = \Delta V_i/\Delta S_{ci}$, is expressed as the quotient between the decrease in voltage DV_i, at the i-th bus, by the increase in apparent power at the i-th bus. DS_{ci}.

These indicators consider that due to the maximum power transfer limit in a transmission line, the voltage on the bus should drop as the consumed power increases, therefore, close to voltage instability, a minimum load increase will lead to a large voltage dip. The sensitivity value is positive and will increase as

the load increases and will tend to infinity when the system reaches the maximum load.

III. IMPLEMENTATION OF INDICES IN PSP-UFU

The IDE (Integrated Development Environment) used to develop the new tool, VSI, was Microsoft Visual Studio 2019, which was used in conjunction with the framework WxWidgets, necessary to generate the GUI (Graphical User Interface) of the program [4][5]. Implementing indicators in the PSP-UFU is straightforward and efficient, as the program already calculates the electrical parameters to be used (voltage and power consumed on each bus), leaving the developer with the task of changing only two classes that are already present in the program. The main class is Electromechanical, where one must define the variables, calculate the indices, and point them to plotting. The second class, called Workspace, is accessed to define the plotting settings of the indices. Inside the Electromechanical class, in the SaveData() method, the indices calculations are performed. After obtaining the correct values of these, the next step is to plot the results. Still in the Electromechanical class, now in the RunStabilityCalculation() method, the indices to be plotted are defined by the user. Finally, the last step is to proceed to the Workspace class and, in the RunStability() method, to implement the code necessary to plot the indices on all buses, regardless of the size of the system created. Fig. 1 shows the flowchart of the implemented indices.

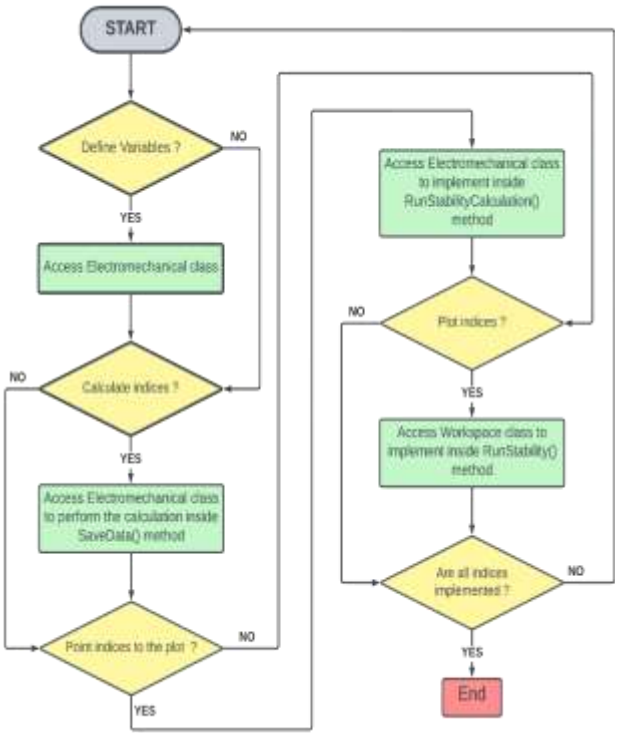


Fig. 1. Flow-chart of the implemented indices.

IV. CASE OF STUDY

To analyze the effectiveness of the indicators in predicting voltage instability of the system, the reduced 3-bus radial system (Fig. 2) and the IEEE 14-bus system (Fig. 4) were simulated. In both cases, the following considerations were made:

- All loads were modeled as constant powers.
- Load 0 was initially connected to the bus. The other loads were inserted at 1 second intervals. The power values of the incremental loads were reduced as the bus approached voltage instability or collapse.
- All incremented loads had a fixed power factor (pf), with pf = 0.920 inductive, for case-1 loads, and pf = 0.948 inductive, for case-2 loads.
- The adopted base power was 100 MVA.

A. *Case 1: Reduced 3-bus system - infinite bus*

In order to model bus 1 in Fig. 2 as infinite, the generating capacity as well as the inertia (H) were specified with very high values, while the direct axis reactance (X'd) was given a low value, and the other parameters were left at null values. Transmission line and transformer parameters were defined as: (0.031 + j0.80) p.u. and (0.001 + j0.050) p.u., respectively.

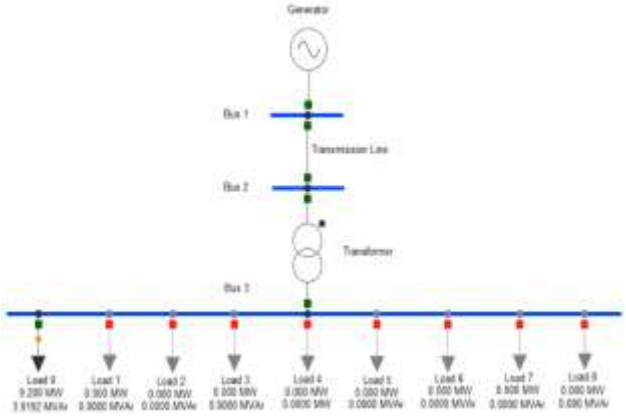


Fig. 2. Case 1-3 bus system with the infinite bus.

The analysis of Fig. 3 shows that when the load is increased on bus 3, the voltage drops until it reaches instability. Fig. 4 displays the active power generated and consumed, as well as the load. From this, it can be implied that almost every active power required by the load is being met. The two curves start practically together but they visually separate after a few seconds. In Fig. 5, it can be seen that the generated reactive power grows much more than the consumed reactive power on the load bus, revealing that the reactive impedance upstream of the load is absorbing the largest part of reactive power, causing a deficiency of reactive power in the load bus and resulting in a voltage drop. Because the line impedance is much higher than that of the transformer, it can be said that voltage collapse occurs because the line has reached the point of maximum power transfer. From Fig. 6 to Fig. 8, it is shown that the indicators start to increase, allowing to establish a relationship between their behaviors and voltage collapse. Fig. 9 points out the three normalized indicators versus the active power consumed. It is

possible to notice that the indicators increase dramatically, tending to infinity, when the maximum load limit is reached, and this proves the relationship between the indicators applied to the case in question and voltage instability, and can therefore be used to predict voltage collapse. Normalization was obtained by dividing the incremental values of each indicator by its first value, so that in the first increment all were equal to 1, enabling the comparison. The fact that the behavior of the three is the same is due to $VSI1_i$, being equal to $VSI3_i \sin\varphi$, and $VSI2_i$, being equal to $VSI3_i \sin\varphi$. As the power factor is kept constant, the terms sine and cosine disappear upon division, which is intrinsic to normalization [7].

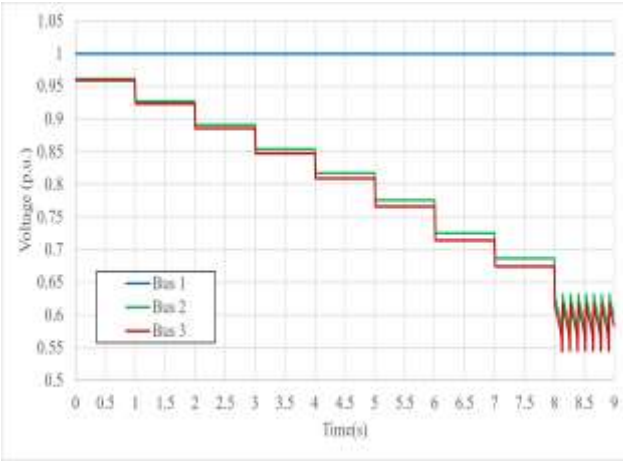


Fig. 3. Reduced 3-bus system: Bus voltages

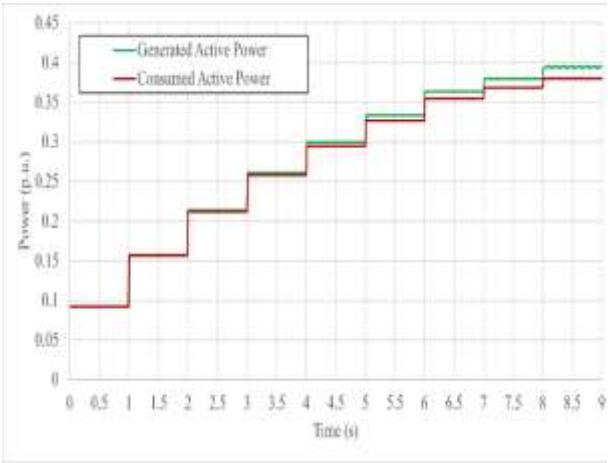


Fig. 4. Reduced 3-bus system: Comparison between the generated reactive power and the consumed reactive power by the load bus.

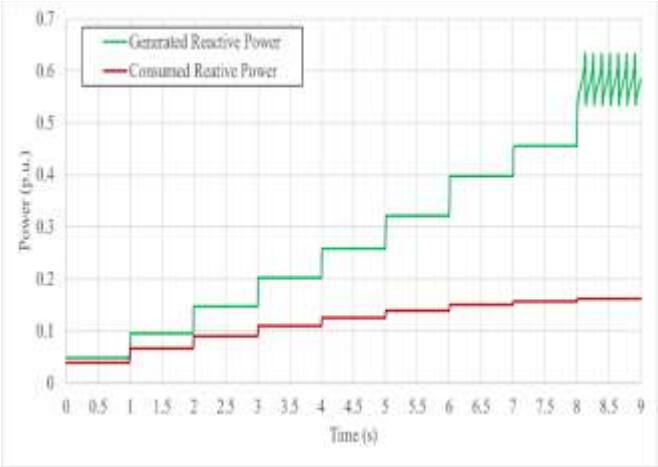


Fig. 5. Reduced 3-bus system: Comparison between the generated reactive power and the consumed reactive power by the load bus.

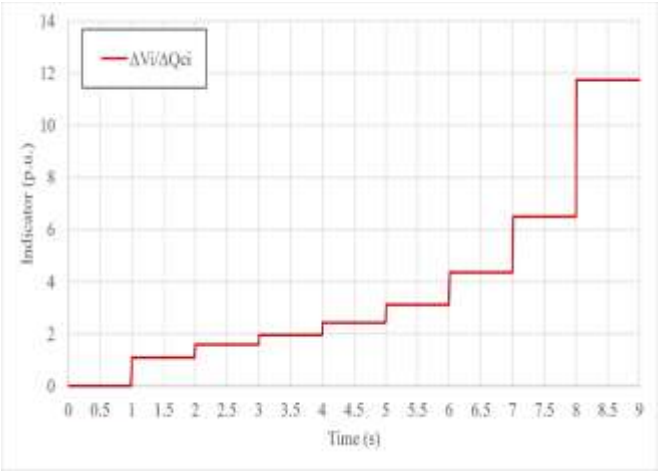


Fig. 6. Reduced 3-bus system: Indicator 1 applied to the load bus.

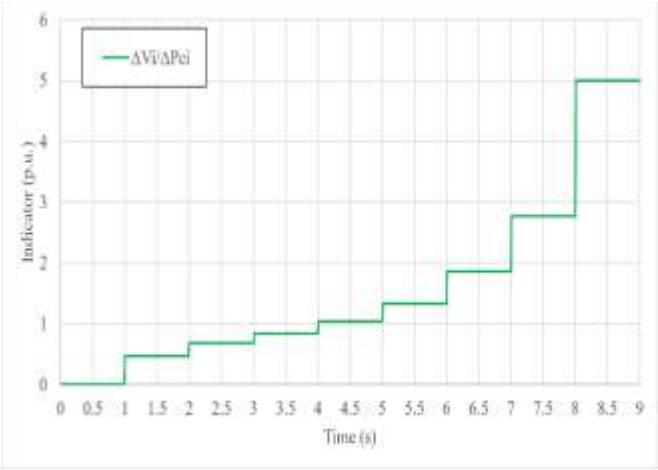


Fig. 7. Reduced 3-bus system: Indicator 2, applied to the load bus.

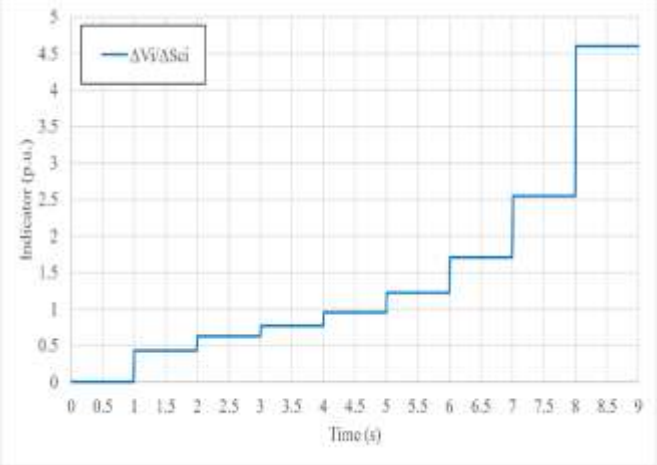


Fig. 8. Reduced 3-bus system: Indicator 3 applied to the load bus.

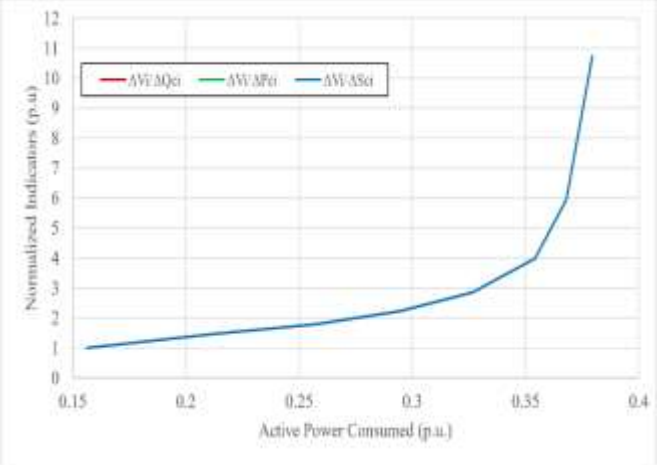


Fig. 9. Reduced 3-bus system: Comparison between normalized indicators.

B. *Case 2:* IEEE *14-bus system*

Using the IEEE 14-bus example system, shown in Fig. 10, bus 14 was randomly chosen to add loads.

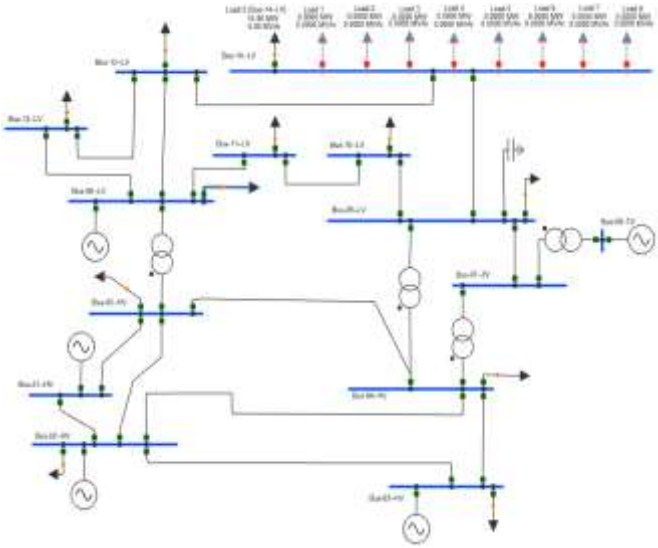


Fig. 10. Case 2 – IEEE 14-bus system.

Fig. 11 shows that the voltage on bus 14 collapses after insertion of the last load. Figs. 12 to Fig. 14 indicate that the indicators increase in value after each load is added to the bus, signaling, as in previous case, a lack of reactive power at the load bus. As previously done, the last graph (Fig. 15) shows the normalized indicators that presented a behavior similar to that explained in case 1 (Fig 9). However, it is important to mention that, since the system is meshed, the voltage collapse will be reached as soon as one of the system transmission lines reaches the point of maximum power transfer.

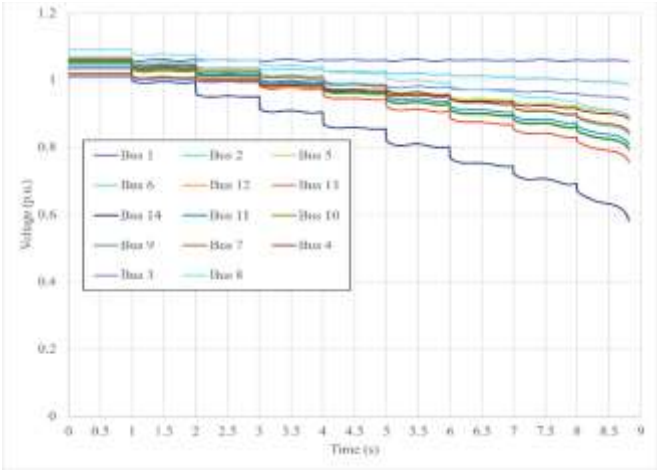


Fig. 11. IEEE 14-bus system: Behaviors of bus voltages.

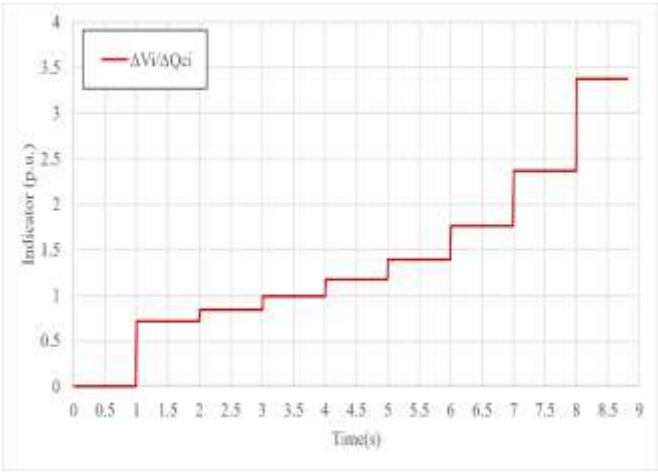


Fig. 12. IEEE 14-bus system: Indicator 1 applied to the load bus.

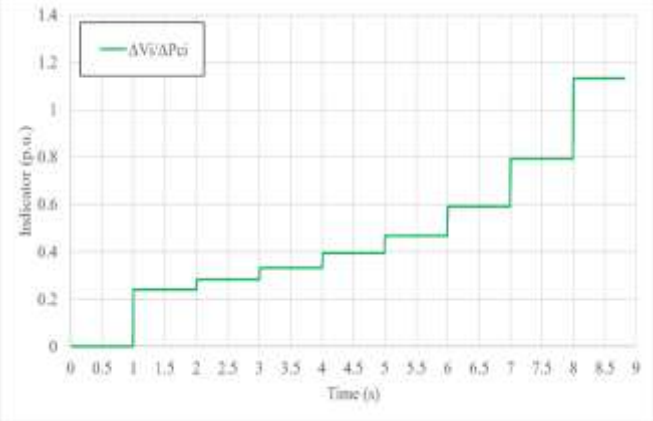


Fig. 13. IEEE 14-bus system: Indicator 2 applied to the load bus.

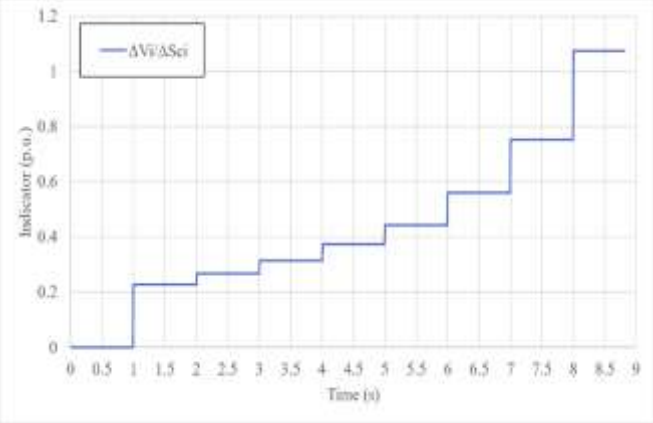


Fig. 14. IEEE 14-bus system: Indicator 3, applied to the load bus.

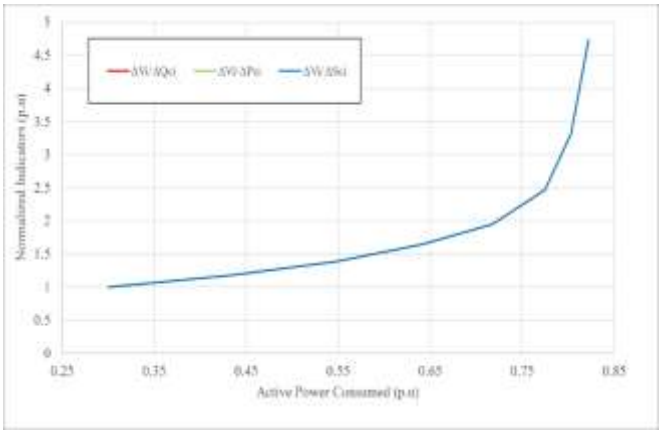


Fig. 15. The normalized indicators presenting a behavior similar to that explained in case 1 of Fig. 9.

V. CONCLUSION

This work presents the implementation of voltage stability indicators in the PSP-UFU software which can simulate the dynamics of any power system, among other functions. Thus, it was possible to implement three types of incremental, sensitive indicators. The objective of these terms is to predict the voltage collapse on a system bus with increasing load. In this context, two case studies were carried out. The first was on a radial reduced system containing 3 buses. The second was on the IEEE 14-bus system. The results have shown that all indicators performed very well in both examples systems. This conclusion came from the observation of their sensitive values which increased as the system tended to instability. However, it is noteworthy that the third indicator, $VSI3_i = \Delta V_i/\Delta S_{ci}$, has the advantage of contemplating two extreme situations: load increments whose powers are purely active or purely reactive. In contrast, indicators 1 ($VSI1_i = \Delta V_i/\Delta Q_{ci}$) and 2 ($VSI2_i = \Delta V_i/\Delta \mathrm{P}_{ci}$) can be applied to only one of these two situations.

AKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

[1] A. R. Nageswa Rao, P. Vijaya, and M. Kowsalya, "Voltage stability indices for stability assessment: a review," Int. J. Ambient Energy, vol. 42, no. 7, pp. 829–845, 2021, doi: 10.1080/01430750.2018.1525585.

[2] A. Oukennou and A. Sandali, "Assessment and analysis of Voltage Stability Indices in electrical network using PSAT Software," 2016 18th Int. Middle-East Power Syst. Conf. MEPCON 2016 - Proc., pp. 705–710, 2017, doi: 10.1109/MEPCON.2016.7836970.

[3] H. Zaheb et al., "A contemporary novel classification of voltage stability indices," Appl. Sci., vol. 10, no. 5, pp. 1–15, 2020, doi: 10.3390/app10051639.

[4] T. L. Oliveira, G. C. Guimarães, and L. R. C. Silva, "PSP-UFU: An open-source, graphical, and multiplatform software for power system studies," Int. Trans. Electr. Energy Syst., vol. 30, no. 2, pp. 1–18, 2020, doi: 10.1002/2050-7038.12185.

[5] T. L. Oliveira, G. C. Guimarães, L. R. C. Silva, and J. O. Rezende, "Power system education and research applications using free and open-source, graphical and multiplatform PSP-UFU software," Int. J. Electr. Eng. Educ., 2019, doi: 10.1177/0020720919879058.

[6] M. S. S. Danish, T. Senjyu, S. M. S. Danish, N. R. Sabory, K. Narayanan, and P. Mandal, "A recap of voltage stability indices in the past three decades," Energies, vol. 12, no. 8, pp. 1–18, 2019, doi: 10.3390/en12081544.

[7] H. S. Barbuy, A. Rocco, L. A. P. Fernandes, and G. C. Guimarães, "Voltage collapse risk associated to under-voltage capacitive compensation in electric power system operation," Am. J. Appl. Sci., vol. 6, no. 4, pp. 646–651, 2009, doi: 10.3844/ajas.2009.646.651.

[8] N. D. Hatziargyriou and T. Van Cutsem, "Indices predicting voltage collapse including dynamic phenomena," CIGRE Task Force, no. December, pp. 2–38, 1994.

[9] C. W. Taylor, Power System Voltage Stability. McGraw-Hill Book Co., 1994.

A Review Analysis of Fouling Effect on Coal Fired Boiler Efficiency

Bai Kamara
Department of Mechanical & Industrial Engineering Technology
University of Johannesburg, Doornfontein Campus
Johannesburg, South Africa.
baikamara78@gmail.com

Daramy Vandi Von Kallon
Department of Mechanical & Industrial Engineering Technology
University of Johannesburg, Doornfontein Campus
Johannesburg, South Africa
dkallon@uj.ac.za

Peter Madindwa Mashinini
Department of Mechanical & Industrial Engineering Technology
University of Johannesburg, Doornfontein Campus
Johannesburg, South Africa
mmashinini@uj.ac.za

***Abstract* - Coal fire boilers are the type of boilers that uses coal as their main fuel for combustion in the boiler furnace. Coal is widely used globally as a fuel source in boilers for heat and electricity generation because of its availability and low cost. Coal from the mine requires less or no chemical treatment, the solid coal is crushed into a fine powder known as pulverized fuel (PF) before being fed into the boiler furnace. Finely pulverized coal in the boiler furnace burns smoothly as compared to liquid and gaseous fuel. However, due to the chemical composition of the coal, there is incomplete combustion of the coal particles. These unburnt fuel substances are carried along by the flue gasses to the boiler tubes surface containing water used for steam generation. An insulating layer has been created on the boiler tube surfaces by these unburnt fuels. The formation of the insulating surface on the boiler tubes is known as fouling. The fouling effect minimises the boiler heat transfer capacity from the tube surfaces to the moving fluid inside the tube and then increases the boiler flue gas exit temperature. Fouling usually negatively impacts the boiler-designed performance efficiency. This paper was focused on reviewing the impacts of fouling on the coal-fired boiler performance efficiency and how the installation of a supervisory controlled and data acquisition (SCADA) system during coal-fired boilers' operation would minimize fouling impact and increase productivity.**

Keywords: ***Coal, boiler, combustion, pulverized fuel, fouling, unburnt fuel, efficiency*.**

I. INTRODUCTION

When coal is burned in boilers for heat energy production, the unburnt fuel is carried away by the hot flue gas in the furnace causing deposition problem on the boiler heat transfer tube surfaces. The deposition of these unburnt substances on these tubes is known as fouling [1]. Fouling on heat exchanger surfaces is known as the formation and deposition of unwanted substances on the boiler tube surfaces. Additional resistance is created resulting from the imposed foulant deposited layer that decreases the tube surface flow area, resulting in an increased volumetric flow rate and velocity. The foulant deposits on the tube surface are non-uniform hence creating a rough surface that can increase the fluid flow resistance across the surface. The fouling process is illustrated in Figure 1. Generally, the challenges posed by the effect of fouling in industrial boilers include(s) a reduction or an excessive drop in the boiler efficiency and a massive pressure drop across the heat exchanger [2]. The net fouling process can be considered or divided into two simultaneous sub-processes which include the foulant deposition and removal process as depicted in Fig. 1 [3].

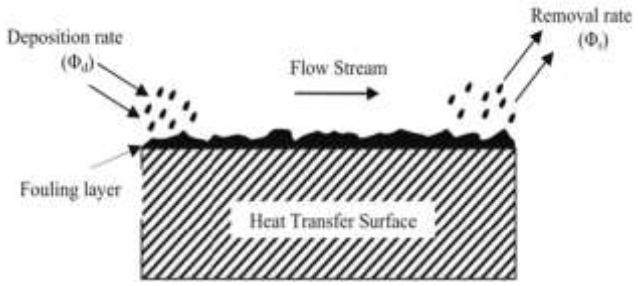


Fig. 1. The Fouling Process [3]

The fouling resistance or fouling factor is referred to as the rate of deposition growth on the tube surface, this can mathematically be represented as the foulant deposition minus foulant removal rate as illustrated in equation 1.

$$R_f = \Phi_d - \Phi_r \qquad (1)$$

Hence ϕ_d and ϕ_r are the deposition and removal rates, respectively. Fouling factor (Rf), deposition, and removal rates are referred to as thermal resistance measured in m^2K/W or the change in thickness per unit time expressed in kg/m^2s [3]. Sintered ash deposition on the heating surface of reheaters and superheaters in conventional boilers from coal combustion is known as fouling [4]. Fouling also includes the formation of deposits in some parts within the boiler not exposed to the radiant flame within the furnace, this ash deposition occurs resulting from the interactions between the suspended flue gases and the moving ash particles in regions of a decreased temperature within the boiler [5]. Ash deposition usually occurs mostly in the closely spaced convection section within the boiler tubes. A major operational concern during combustion within industrial coal-fired utility boilers is the deposits and accumulation of ash and unburnt substances on the surface(s) of the boiler heat exchangers. The fouling effect is a primary source of an essential heat loss in utility thermal power plant boilers, this impacts significantly on the boiler's efficiency and its designed operational performance. An estimated 1% efficiency loss in the thermal power plant is recorded under the boiler's normal operating conditions [6]. Fouling disproportionately might result in an increased flue exit gas temperature, and the ashing rate of deposition on the heat transfer surfaces(s) leads to continually changing conditions inside the boiler that would result in a decrease in the boiler

operational efficiency [7]. The difference in temperature that might result in high-temperature fouling ranges from 900°C to 1300°C, whereas low-temperature fouling occurs within the temperature ranges of 300°C to 900°C [8]. A survey on the fouling effect on coal-fired utility boilers in 1987 conducted by the Electric Power Research Institute (EPRI) within the United States on 91 pulverised coal boilers shows that 7% of the boilers encountered regular fouling whereas 40% undergo occasional fouling problems [9]. Recent research shows that many pulverised coal fire utilities encountered the challenge of frequent fouling. Hence, ash fouling deposition in pulverised coal boilers is a global concern [10].

II.FOULING MECHANISM AND RATE OF FOULING

The energy-generating industries that use coal-fired boilers are faced with a major challenge(s) which involves the accumulation of unwanted substances, dissolved materials, and suspended particles/ substances within flowing fluid and on the heat transfer surfaces during coal combustion [11]. This phenomenon referred to as fouling can negatively impact the equipment's operational performance by decreasing its effectiveness and thermal efficiency. This effect causes the industry serious economic losses which involve the installation of regular cleaning devices [12,13]. The fouling phenomenon in heat transfer surfaces cannot be eliminated but can be minimised to increase the plant efficiency and operationality. Fouling mitigation procedures and effective cleaning applications such as the installation of soot blowers requires an in-depth understanding of ash deposition and cleaning mechanisms [76]. There are several distinct fouling mechanisms. Mostly, the majority of these mechanisms occur spontaneously (combined) that require a different prevention application. The fouling mechanism requires or can involve the following stages.

i. Initiation period of delay: this involves the initial period before the settling of foulants on a clean heat transfer tube surface. Deposition of a relatively minimal foulant improves the surface heat transfer capacity resulting in an initial recorded negative rate of fouling.
ii. Particle formation aggregation, and flocculation: this involves the formation and settlement of solid particles on the surface of the heat exchangers. Water obtained from streams normally contains suspended solid impurities, for example, cooling water. The particle settling process is dependent on the fluid velocity and characteristics. An increase in the fluid velocity can aid in flushing the particles from the tube surface as in most cases particles may tend to stick to each other on the surface and becomes challenging to remove with an elapse of time.
iii. Transportation and migration of foulant
iv. Foulant deposition and separation process of foulant initiation attachment leading to deposit formation or nucleation. This chemical process takes place as the accumulation of solid particles or viscous tar formation near the hot tube surface.
v. Particle growth, hardening, aging, flue flow opposition increase, auto-retardation, and removal.

III. RATE OF FOULING

The occurrence of particle deposit accumulation in a phenomenal process with temperature variation in different natural, industrial, and domestic processes is referred to as fouling. The combined fouling process is represented by the fouling factor, the fouling resistance (R_f) on tube surfaces which can be calculated and obtained from section tests or assessed from the reduced heat transfer capacity of the operating heat exchanger. When unwanted materials settle on a clean surface, the surface becomes dirty. The average surface deposit load formation on a surface area at a specified time is known as the rate of fouling. The fouling rate can be represented on a fouling curve (fouling factor-time curve as shown in Figure 2) that displays the various mode of fouling regarding time. Depending on the fouling condition, the curve can be represented as linear, falling, asymptotic, or saw-tooth [3,15].

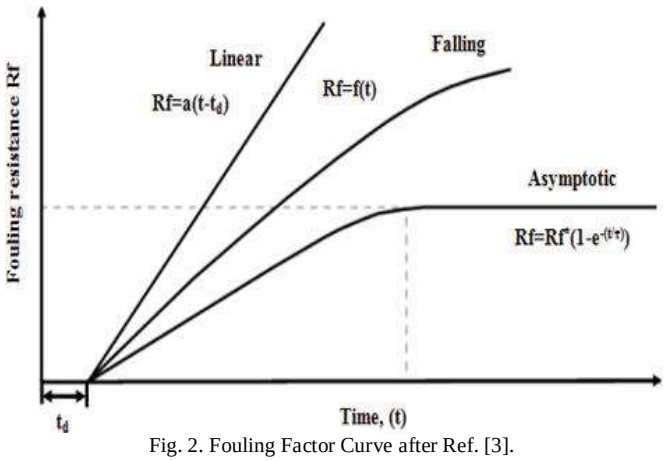


Fig. 2. Fouling Factor Curve after Ref. [3].

The initiation period and the roughness delay time (t_d): is an indication of an initial elapsed time interval where no fouling takes place. This period is an unpredictable time occurrence on the surface, it appears in a non-uniform manner with a normal scattering having average values that are dependent on a given surface or system. After fouling occurrence with the surfaces cleaned to be reused, the delay period can be normally lesser compared to when the tubes were used for the first time. It is significant to consider that the outcome of the graph of the fouling resistance against time is not dependent on the initiation period of the fouling delayed period (t_d) [3].

Saw-tooth fouling occurs when deposits on some parts of the tube surface fall off due to the critical residence period or the attainment of a critical deposit thickness as shown in Figure 3. During the boiler operations at this point, the layer formed on the tube surfaces from the foulant accumulation breaks off eventually. The foulant formation and break-off time difference occur because of pressure pulses, scaling, air trapped within the surface deposits during the boiler shutdowns, and other operational reasons. This occurrence normally happens during the system start-ups and shutdowns

or other transient periods during the equipment operation [15].

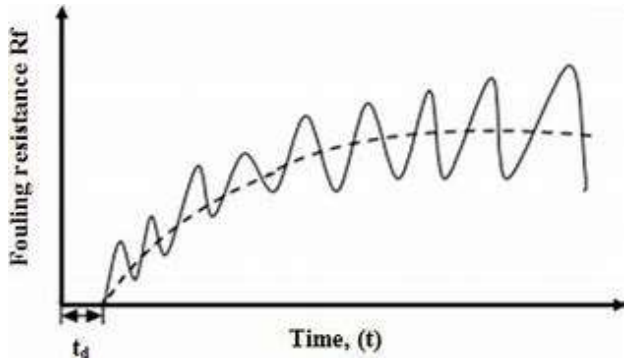


Fig. 3. Sawtooth fouling curve after [3].

Generally, for all fouling modes, the materials deposited as per the tube surface area, (m_f) are a function of the resistance fouling (R_f), foulant mass per unit volume (ρ_f), thermal conductivity (λ_f), and the foulant thickness (x_f). this is expressed as,

$$m_f = \rho_f x_f = \rho_f \lambda_f R_f \quad (2)$$

$$R_f = \frac{x_f}{\lambda_f} \quad (3)$$

The thermal conductivity of a material in the transfer is the ability of a material to transfer the heat received between layers of two fluids separated by a solid surface e.g., heat from flue gasses during coal combustion to the moving water inside the tubes in a coal-fired boiler. This is measured in watts per meter kelvin. Table 1 represents the thermal conductivities of some foulant.

TABLE 1 FOULANT THERMAL CONDUCTIVITY [16].

Foulant	Thermal Conductivity. W/mK
Alumina	0.42
Biofilm (effectively water)	0.6
Carbon	1.6
Calcium sulphate	0.74
Calcium carbonate	2.19
Magnesium carbonate	0.43
Titanium oxide	8.0
Wax	0.24

IV. ASH BUILD-UP MECHANISM

The ash build-up mechanism on the boiler tube surfaces is strongly dependent upon the presence and ash build-up of inorganic vapours and unburnt coal particles on the tube heat exchanger surfaces. The ash particles formation process involves the transportation of unburnt particles through the flue gasses to the boiler wall and heat transfer surfaces. The ask particles sticking and impaction on the heat transfer surfaces during coal combustion in the boiler is known as particle sticking. The particle sticking, impaction, and capture efficiency are related in Eqs. 4, 5, and 6 [17–18]:

$$Impaction\ (\%) = \frac{Mass\ of\ particulate\ impacting\ surface}{Total\ mass\ of\ particulate\ injected} \quad (4)$$

$$Sticking(\%) = \frac{Mass\ of\ particulate\ sticking\ surface}{Mass\ of\ particulate\ impacting\ surface} \quad (5)$$

$$Capture(\%) = \frac{Mass\ of\ particulate\ sticking\ surface}{Total\ mass\ of\ particulate\ injected} \quad (6)$$

$$= Impaction(\%)\ x\ Sticking(\%)$$

This definition occurs if the combined mass of the injected particle is calculated with the projected area facing an obstacle, considering a cylinder with diameter D [19]. Ash deposition process in boilers generally comprises major mechanisms which include the following, inertial impaction, vapour condensation, thermophoresis, and chemical reaction.

a. Inertial Impaction

This build-up process on the convective tubing occurs when combustible particles (>10μm) within the boiler furnace acquire enough inertia to move along with the flowing flue gases that settle on the boiler tubes heat transfer surfaces by inertia impact. The rate of particle deposition impact on the heat transfer surfaces is dependent on the property of flow gasses, impact angle, fluid velocity, particle size, density, and shape. The targeted geometry can be defined as the ratio of the number of particles impacting the heat transfer surface to the overall number of particles flowing towards the surface in free gas flow [20,21].

b. Thermophoresis

This involves the transportation of moving fluid particles along with the moving gasses based on the local temperature gradient. In other situations, this is a dominant mechanism for submicron particles (<10μm) occurring in a situation where the temperature difference between the tube surface(s) and the moving flue gases is greater. A more evenly distributed deposit on the tube surfaces is seen on finer thermophoresis [21].

c. Vapor condensation

This process occurs when vapours are conveyed through cool heat transfer surfaces and allowed to cool down on the surfaces with a lesser temperature compared to the flowing gases or the deposited foulants. The tube's cooling thickness is dependent on the occurrence of the inorganic substances. Particles with a size greater than 0.5 micrometers (> 0.5μm) are stickier and evenly distributed [22,23]. Inorganic vapour condensation occurs during cooling, it can be observed when supersaturated flue gas and the gas temperature within the boiler furnace dropped below the vapour dew point.

d. Chemical Reaction

The chemical reaction is a heterogenous reaction resulting in deposited particles and moving gases. This reaction is dependent on the mass of the deposit and the moving flue gas particles. The dominating inorganic reactions are often mostly alkali metals on

the particles of boiler walls. Couch (1994) stated that the chemical reaction determines whether the particles can stick together before the deposition growth process commence [24].

V. SIGNIFICANCE OF COAL-FIRED BOILERS

Electricity generation from coal originated in the early 1800s, coal from coal mines was pulverised (ground into a fine powder) in huge mills and blown into huge kettles known as boilers. Pulverised coal combustion takes place inside the boiler furnace, the heat generated from the coal combustion is transferred to the tube surfaces containing flowing water, and the water is then converted into steam. The steam from the boilers is used to rotate the blades of giant installed fans of propellers known as turbines. The turbines then rotate a copper coil wire called a rotor inside a magnet containing a stator that makes up the generator. The generator produces an electric current that is transmitted and distributed to consumers' homes and factories [25]. Coal fire boilers are designed with heat exchangers in the form of shells or tubes. The flowing water contained inside the tube(s) is converted into steam or high-pressure hot water. The primary source of heating is provided by the burning of fossil fuel inside the boiler chamber, the heated gas from the combustible fuel is ducted around the tube(s) containing water i.e., water-tube boilers, or the hot gas can pass through the tube i.e., fire-tube boilers in shell containing water, these processes primarily involve the transfer of heat from the burning fuel in heating the water into steam. Larger industrial stationary boilers were used in large power generating stations, the designed process of large industrial boilers is mostly water-tube types. The fire-tube boilers were restricted to approximately 2.4 MPa (350 psi) pressure output. They are mostly used for heating and steam process applications [26]. The design process of a water-tube natural water circulation boiler involves the production of a super-heated stem, the fuel used in this boiler is both natural gas and product gas in a process known as cogeneration. The boiler furnace ceiling consists of four burners, two of each with a gasifier line. The ignition process in the boiler starts with the natural gas, after ignition the flame changes the natural gas into product gas. The product gas produces a temperature of 850^{o}C used to vaporise the circulating water in the tube into steam at a temperature of 540^{o}C producing a pressure of 121 bar [27]. There was a growth in the global coal consumption capacity annually between the years 2000 and 2019 in which the generation capacity nearly double from 1,066GW to 2045GW. This growth rate is declining dramatically. In 2018, there was a 20GW net increase that was recorded as the smallest growth in decades. Nearly 40% of global electricity is generated from industrial coal fire plants. Globally, there has been a rise in the number of countries using coal power for electricity generation in recent times from 66 in 2000 to 80 in 2019. As 13 more countries were also planning to incorporate coal power into their electricity generation [28]. The use of coal as a primary means of heat energy production for electricity generation using boilers is still used in many countries globally. The world's total energy supply share in the year 2015 was made up of; oil (31%), coal (28.1%), Natural gas (21.6%), Biofuels and waste (9.7%), Nuclear (4.9%), Hydro (2.5%) and other (1.5%) [29]. By the end of 2015 in China, there were 565,000 fully operational industrial boilers, out of which 464,000 were coal-fired industrial boilers. Also, the electricity generation capacity in Poland in 2015 was 164.3TWh with 80.9% of this power supply generated from industrial coal-fired boilers [30]. Almost 60% of the electricity produced and supplied in India is generated is generated using industrial coal boilers. Research on new methods using different technologies is being developed in providing more thermally efficient industrial coal-fired boilers for heat and electricity production as well minimizing the impacts of global warming and climate change [31]. There was a significant upsurge in coal generation in Asia, particularly in China and India, but elsewhere it was declining especially in the United States and Europe. But coal remains the largest source of electricity globally consisting of a share of 38% worldwide [32].

VI. EFFICIENCY OF COAL FIRE BOILERS

The efficiency of a coal fire boiler can be simply expressed as the total energy output to the total energy input. This can be expressed mathematically as.

$$Boiler\ efficiency(\%) = \frac{Heat\ exported\ in\ steam}{Heat\ provided\ by\ the\ fuel} x\ 100 \quad (7)$$

The heat exported by the steam is obtained from the use of steam tables by applying results obtained from the temperature of the boiler-fed water, exported steam temperature, and flow rate. The combustible heat energy from the coal fuel is obtained from the fuel calorific value. The combined fuel calorific value can be obtained by combining the gross and net calorific values.

Combined calorific value

The calorific value of a fuel is defined as the energy it contained, coal and most other fossil fuels are made up of hydrocarbons. During ignition, the hydrogen reacts with oxygen to produce water, carbon dioxide and hydrogen gas. which is transported in the boiler furnace as steam. the combined energy from the burning coal in the boiler furnace used to evaporate the water in the boiler tubes to steam through the flue gasses is referred to as the fuel (coal) total calorific value.

Net calorific value

The net calorific value is the energy contained within the coal (fuel) before ignition. This can be used to obtain the boiler overall efficiency. This can be expressed mathematically as Net calorific value (%) ~ Gross calorific value (%) – 10%. During coal ignition in the boiler furnace, the chemical reaction taking place includes. The boiler accuracy and efficiency are dependent on the controlled amount of air as stated below.

- Excess air cools the furnace and takes away useful heat

- Insufficient air results in incomplete combustion, unburnt fuel would be carried away and hence smoke is produced inside the furnace [33].

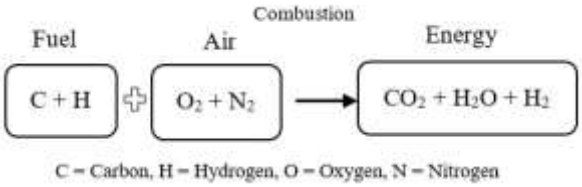


Fig. 4. Coal Fuel Combustion Reaction after Ref. [31].

Coal is presently the most widely used form of primary fuel globally. It accounts for approximately 36% of global electricity generation. Burning coal is the most cost-effective and energy-efficient means of generating electricity. Coal has traditionally dominated the electricity generation market globally for decades, although this is expected to decline as coal faces greater competition with non-conventional renewable energy generating sources. In South Africa, approximately 77% of the country's energy needs are obtained from coal whereas 53% of the electricity generated and supplied throughout the country is obtained from industrial coal-fired utilities [25]. From the findings of the research carried out by Zeitz in 1997, it was described that the thermal efficiency in utilising industrial stoker coal boilers for electricity generation is approximately 65 to 85%. There are always unavoidable losses during coal boiler operations. However, if the source of losses in the coal boilers is fully understood, this might result in the improvement of the boiler efficiency and hence there will be an increase in the energy output and hence the boiler efficiency can be increased at around 15 – 20%. The boiler efficiency is an indication of the stored energy of the coal that is converted to heat during combustion. How efficient is the boiler can be determined by its efficient combustion and heat transfer capacity to the boiler tubes containing flowing fluid (water). The boiler efficiency is a measure of the ratio of heat output to the amount of fuel (coal) used during the boiler operation [34]. It was estimated that energy generation globally from coal fire boilers represents approximately 33% of the total energy generated. The presence of coal fire boilers in many countries is used for electrical energy generation used in providing electricity for both residential and industrial consumption. Based on the coal chemical compositions (such as carbon, nitrogen, sulphur, moisture, etc). Coal in the boiler does not completely burn as in the case of liquid fuel and natural gases, these unburn substances are carried away by the hot flue gasses which then stick on the heat transfer surfaces and hence increase tube fouling resulting in a decrease in the boiler designed efficiency [35]. Large-scale industrial coal boilers are more efficient (i.e., approximately above 90%) this can be varied directly with a change in the boiler load (Rayaprolu, 2013). From the findings of Tzolakis *et* al, (from the results obtained through simulation by optimizing a 300MW lignite coal fire power plant, it was observed that increasing the boiler thermal efficiency, resulted in a lower fuel consumption and flue gas emission [34]. A study by Drosatoe *et* al shows that an improvement of 0.55% in coal fire boiler thermal efficiency decreases its fuel consumption by 2.06% (11.5t/h lignite) and reduces the CO_2 emission in the flue gases by 2.06% (4.8 t/h of CO_2). Presently, coal fire boilers need to be designed to meet the specifications required in the energy market and operate to meet the constant varying daily load changes. The boiler efficiency must be kept at a maximum level to meet the changing load consumption. This is essential from the energetic, ecological, economic, and maintenance viewpoint [35].

VII. RECOMMENDATION

Considering the essential role played by the application and use of coal-fired boilers in steam and electricity generation globally in meeting the rapidly growing global energy demand, it is recommended that coal boilers must be designed to maximize performance efficiency. Modern methods and technological approaches are to be used in minimizing the effect of fouling during coal-fired boiler operations. The use of an effective and efficient mechanical online cleaning such as the use of fitted cameras in monitoring boiler tube foulants accumulation is to be employed. acoustic cleaning devices with little maintenance cost must be installed to clean the boiler superheaters, reheaters, and economizers. Using a supervisory control and data acquisition (SCADA) system during industrial coal boilers' operation will be essential in determining the foulant accumulation, temperature, and pressure variations.

VIII. CONCLUSION

This paper has discussed the significance of coal fire boilers in electricity generation. The negative impact of the fouling effect on coal-fired boilers' performance efficiency, the fouling mechanism, and the rate of foulant deposition on the boiler heat exchanger tubes. The ash deposition on the superheaters, reheaters, and economizers in the coal boiler. The significance of coal-fired boilers for industrial applications in electricity and steam generation as well as the performance efficiency of these boilers discussed. Hence, recommendations and conclusions on the research significance are explained.

REFERENCES

[1] S. C. Stultz, and J. B. Kitto. Steam its generation and use, The Babcock & Wilcox Company, Baberton, Ohio, U.S.A. 1992.

[2] B. T. Reg. Fouling. http://thermopedia.com/content/779/ 14 February 2011. Accessed 15/08/2022

[3] M.M. Awad, Fouling of heat transfer surfaces, Mansoura University, Faculty of Engineering, Mech. Power Eng. July 2014.

[4] A.L., Robinson, H. Junker, and L.L. Baxter, Pilot-scale investigation of the influence of coal-biomass cofiring on ash deposition. Energy Fuels, vol. 16, pp. 343–355, 2002.

[5] S.C. Srivastava, K.M. Godiwalla, and M.K. Banerjee, Fuel ash corrosion of boiler and superheater tubes. J Mater Sci, vol. 32, pp. 835–849, 1994.

[6] A. Valero, and C. Cortes, Ash fouling in coal-fired utility boilers: monitoring and optimization of on-load Cleaning, Progress in Energy and Combustion Science, vol. 22, pp. 189-200, 1996.

[7] S. Kalisz, and M. Pronobis, Investigations on fouling rate in convective bundles of coal-fired boilers in relation to optimization of soot blower operation. Fuel, vol. 84, iss. 7, pp. 927–937, 2005.

[8] Q. Fang, H. Wang, Y. Wei, L. Lei, X. Duan, and H. Zhou Numerical simulations of the slagging characteristics in a down-fired, pulverized-coal boiler furnace. Fuel Process Technol, vol. 91, iss.1, pp. 88–96, 2010.

[9] R. E. Barrett, Slagging and fouling in pulverized coal-fired utility boilers. A survey and analysis of utility data, EPRI Final Report RP1891-1, Volume 1, 1987.

[10] Pena B. Soft-Computing models for soot-blowing optimization in coal-fired utility boilers, Applied Soft Computing Journal, pp. 1657-1668, 2011.

[11] H. Demasles, P. Mercier, P. Tochon and B. Thonon, Guide de L'encrassement des Echangeurs de Chaleur. France: Editions-GREThi, 2007.

[12] B. Farajollahi, Gh S. Etemad, and M. Hojjat, Heat transfer of nanofluids in a shell and tube heat exchanger. International Journal of Heat and Mass Transfer, vol. 53, pp. 12-17, 2010.

[13] S. Lalot, and H. Pálsson, Detection of fouling in a crossflow heat exchanger using a neural network-based technique. International Journal of Thermal Sciences, vol. 49, pp. 675-679, 2010.

[14] T. Pogiatzis, E.M. Ishiyama, W.R. Paterson, V.S.Vassiliadis, and D.I. Wilson Identifying optimal cleaning cycles for heat exchangers subject to fouling and ageing. Applied Energy, vol. 89, pp. 60-66, 2012.

[15] R. Jradi, A. Fguiri, C. Marvillet, and M. R Jeday. Tubular Heat Exchanger Fouling in Phosphoric Acid Concentration Process, Intech Open Books Chapter, 2014

[16] T. Kuppan, "Heat Exchanger Design Handbook", Marcel Dekker, Inc., New York, 2000.

[17] S.S. Lokare, J.D. Dunaway, D. Moulton, D. Rogers, D.R. Tree, and L.L. Baxter, Investigation of ash deposition rates for a suite of biomass fuels and fuel blends, Energy Fuels, vol. 20, pp1008–1014, 2006.

[18] B. Barker, B. Casaday, P. Shankara, A. Ameri, J.P. Bons, and J Turbomach, Coal ash deposition on nozzle guide vanes – part II: computational modelling, vol. 135, iss. 1, pp 11-15, 2013.

[19] R.S. Weber, N. Mancini, M. Mancini, and T. Kupka, Fly ash deposition modelling: requirements for accurate predictions of particle impaction on tubes using RANS-based computational fluid dynamics, Fuel, vol. 108, pp586–596, 2013.

[20] S.K. Kær, "Numerical investigation of ash deposition in straw-fired boilers: -Using CFD as the framework for slagging and fouling predictions." Institut for Energiteknik, Aalborg Universitet. 2001.

[21] L.L. Baxter, "Ash deposition during biomass and coal combustion: A mechanistic approach." Biomass and Bioenergy, vol.4, pp 85-102, 1993.

[22] H.S. Zhou, P.A. Jensen, and F.J. Frandsen, "Dynamic mechanistic model of superheater deposit growth and shedding in a biomass fired grate boiler." Fuel, vol. 86, pp.1519-1533, 2007.

[23] P.A. Jensen, "Characterization and quantification of deposits build up and removal in straw suspension fired boilers: final report." Technical University of Denmark. 2013.

[24] G R. Couch Understanding slagging and fouling in pf combustion. London: IEA Coal Research; ISBN:92-9029-240-7, 1994.

[25] Eskom2020(http://www.eskom.co.za/AboutElectricity/ElectricityTechnologies/Pages/Coal_Power.aspx 12/05/2020) Accessed 7/8/2022

[26] S. Hall, Boilers in Branan's Rules of Thumb for Chemical Engineers 5th Edition, 2012.

[27] K. Sipilä, Cogeneration, biomass, waste to energy and industrial waste heat for district heating in Advanced District Heating and Cooling (DHC) Systems. 2016.

[28] IEA2017www.iea.org/publications/freepublications/publication/KeyWorld2017.pdf. Accessed 22/7/2022

[29] P. Madejski, Thermal power plants, new trends and recent developments. Intech Open, ISBN 978-1-78923-079-6, 2018.

[30] IEA 2019. https://www.iea.org/reports/tracking-power-2019/coal-fired-power. Accessed 14/4/2022

[31] Spirax Sarco (https://www.spiraxsarco.com/learn-about-steam/the-boiler-house/boiler-efficiency-and-combustion) Accessed 26/7/2022

[32] R A. Zeitz, Ciboenergy Efficiency Handbook Council of Industrial Boiler Owners (CIBO) 6035 Burke Centre Parkway, Suite 360 Burke, Va 22015, 1997.

[33] H. O. Garc´es, J. Abreu, P. G´omez, C. Carrasco, L. Arias, A.J. Rojas, and A. Fuentes, Energy Efficiency Monitoring in a Coal Boiler Based on Optical Variables and Artificial Intelligence. IFAC PapersOnLine, vol. 50-1 pp. 904–909, 2017.

[34] G. Tzolakis, P. Papanikolaou, D. Kolokotronis, N. Samaras, A. Tourlidakis, and A. Tomboulides, Simulation of a coal-fired power plant using mathematical programming algorithms in order to optimize its efficiency. Appl Therm Eng, vol. 48, pp 256 – 267, 2012.

[35] P. Drosatos, N. Nikolopoulos, E. Karampinis, G. Strotos, P. Grammelis, and E. Kakaras, Numerical comparative investigation of a flexible lignite-fired boiler using pre-dried lignite or biomass as supporting fuel. Renew Energy, vol.145, pp. 1831 – 1848, 2020.

Bituminous Coal Ash Analysis for fouling investigation Using Induced Coupled Plasma and X-Ray Fluorescence Methods

Bai Kamara
Department of Mechanical &
Industrial Engineering Technology
University of Johannesburg,
Doornfontein Campus
Johannesburg, South Africa.
baikamara78@gmail.com

Daramy Vandi Von Kallon
Department of Mechanical &
Industrial Engineering Technology
University of Johannesburg,
Doornfontein Campus
Johannesburg, South Africa
dkallon@uj.ac.za

Peter Madindwa Mashinini
Department of Mechanical &
Industrial Engineering Technology
University of Johannesburg,
Doornfontein Campus
Johannesburg, South Africa
mmashinini@uj.ac.za

***Abstract*: During coal combustion in the boiler light ash particles are carried away along with the hot flue gases referred to as fly ash and the heavier ash particles fall below the bottom of the boiler known as bottom ash. The fly ash particles stick on the convective heat transfer surfaces causing disposition problems on the heat exchanger tubes, the deposits formed reduce the boiler efficiency. In this research, ash samples were collected from the Sasol operational site in Secunda, Mpumalanga in the Republic of South Africa, for laboratory analysis. The ash samples were analysed using induced coupled plasma and X-ray fluorescence methods in determining their chemistry. Ash transportation and deposition mechanism on heat transfer surfaces, the ash physical and chemical properties, the coal mineral transformation, and the fouling effect of the elements present within the ash samples were discussed. The results obtained in this research indicate low to medium fouling potential on the convective heat transfer tubes with low slagging potential in the boiler furnace based on the coal ash chemistry**.

Keywords: coal combustion, fly ash, boiler efficiency, fouling, slagging

I. INTRODUCTION

The derived residue that remains after the complete combustion of coal is referred to as ash. The resulting ash contents from the complete combustion of complex inorganic constituents within the coal are used as an indication of the coal quality or grade as it can be used to measure the quantity of incombustible material within the coal. During laboratory analysis, ash is derived by weighing the residue after complete combustion of approximately 1g of coal samples with proper equipment specifications under rigidly controlled mass, temperature, time, and atmospheric conditions [1]. During coal fuel combustion in coal-fired boilers, the effect of ash deposition on the boiler heat exchanger tubes would result in many operational problems which include reduced production in the boiler efficiency, frequent power plant shutdowns, reduced heat transfer, an increased operational downtime, increased soot blowing and other cleaning activities [2]. A major difficulty in a pulverised coal boiler design is achieving a higher burnout level. Managing fuel and load flexibility during the boiler operation can be challenging. The major effects of coal combustion within a boiler include slagging and fouling. Slagging occurs within the boiler where the heat exchange surfaces are directly exposed to the flame radiation area. Slag particles or deposits are mostly molten and are in a liquid state. The deposits or particles formed on heat exchange surfaces that are not directly exposed to the radiant flame area are referred to as fouling. Fouling deposits are solidified particles that are loosely bound or partially sintered and hence they are easier to clean [5]. Slagging and fouling within the boiler underwent a smooth transition, the extent to which the deposits lead to slagging or fouling is dependent on the coal type and its mineral inorganic chemical composition as well as the operating conditions of the boiler. The fly ash size formation during coal combustion is dependent on multi-modal size distribution. The larger particles originated from the coal mineral grains and the size of the fly ash formed is dependent on the coal characteristics, pre-treatment, and combustion conditions. The sub-micron fly ash particles originate mainly from homogeneous nucleation, volatilised subsequent flame coagulation within the coal organic species [3]. During a coal-fired boiler construction, critical parameters to be considered in the furnace size and height estimation are the ash softening and melting temperature(s). It is estimated that the furnace exit gas temperature should stay below the ash softening temperature [4]. Other essential factors which influence coal mineral matter transformation and ash deposition on heat transfer tubes are the boiler operating conditions, this includes the boiler load, air to fuel ratio, gas temperature, and soot blowing patterns. Slagging and fouling degree throughout the boiler vary. This variation is dependent upon the combustion environment that influences ash deposit formation on the tube heat transfer surfaces. Strongly bonded deposit formation on heat exchanger tubes reduces soot blowing effectiveness [6]. Ash deposition growth during the boiler operation changes the conditions of gas flow and temperature within the boiler. In this study, laboratory analysis in determining the elements present (i.e., major and minor) and trace enriched elements within the ash samples using induced coupled plasma (ICP-AES and MS) and XRF (X-Ray Fluorescence) analysis. The elemental analysis of the bituminous coal samples was obtained using CHNS (Carbon, Hydrogen, Nitrogen, and Sulphur) analyser.

II. LITERATURE REVIEW

During combustion, the coal mineral undergoes complex chemical and physical transformations resulting from high temperature(s), complex oxidising, and reducing environment by the flame. Ash is then produced as an unburnt residue from the combustion of the coal mineral matter. The coal inorganic constituents are made up of it physical and chemical characteristics, association, and combustion conditions resulting to solids, liquids and vapours formation [7,8]. The ash formation process during combustion is also influenced by the pulverised coal mineral particles' size distribution i.e., the smaller the coal grain the more reactive and versi versa. The core of the larger coal grain does not take part in the ash mineral transformation during the combustion [9]. It was suggested that alkalis evaporation during coal combustion is affected by the size of the coal mineral grain. Mineral matter occurrence within the coal influences its combustion behaviour. The coal grain mineral substances are either included or excluded [10]. Fe, Mg, P, and K are the clustered coal mineral particles closely associated within the coal matrix which cannot be physically separated by either crushing or milling. The loosely bonded minerals that are non-homogenously distributed within the coal are referred to as excluded or extraneous minerals. They can be easily removed during milling or crushing. They are usually distributed within the coal particles as deposits in cleavage and crack fractures or seen as dirt bands in the coal seams [10]. Excluded minerals also involve dirt obtained during coal mining and handling. During coal pulverisation, excluded minerals are given off from the coal matrix while inherent minerals remain within the coal pores. During griding, the fragmentation of the excluded coal minerals might occur as a result of the thermal shock and rapid gas evolution during the decomposition process forming fine ash particles [8]. Acute temperature gradients are observed within the decomposed coal particles caused by the thermal shock this will results in the development of strong stresses within the particle [11]. The most common mineral found in coal is quartz, with a melting point of 1723ºC and a boiling point of 2230ºC. It is the same size as the pulverised coal particles, and it is relatively a non-reactive material and occurs mostly in the form of crystals. During coal combustion, the quartz's sharp crystalline structure is preserved and hence causes erosion problems within the boiler [8,12]. The presence of sodium (Na) in ash deposition within the boiler produces low melting silicate causing slagging and fouling at an increased temperature. Fouling is also initiated by melting sodium sulphates slightly above the temperature in the reheaters and superheaters section in the boiler [13]. Fumes of reactive submicron calcium oxide (CaO) and calcium silicate cenospheres are formed from organically bound calcium [14, 15]. Calcium chloride is volatilised in the furnace radiated flames and it is then converted into sulphate in a manner similar to the sulphation process of volatilised sodium chloride (NaCl) [15]. Furnace wall and convective tubes heat transfer surface deposits are initiated by the release of calcium fumes and calcium sulphate at low temperatures during coal combustion in the boiler furnace [14,15]. There is an increased reaction rate by organic calcium with quartz and aluminium silicates in a lower melting phase formation e.g., calcium silicate within the coal matrix. A partially developed slag deposit is formed by calcium silicates e.g., gehlenite [14]. Ash particle size formation is determined by char fragmentation process during pulverised coal combustion, without fragmentation, an ash particle is formed by the combustion of a char particle. Smaller ash particles are formed by the combustion of fragmented pulverised char particle [16,17]. Smaller coal particle size distribution (PSD) and diminished mineral matter (PSD) originates from smaller ash PSD [16,17]. Due to vaporisation and condensation, fumes may be formed during coal combustion. At low temperatures, organically bounded minerals were reported to be vapourised extensively whereas more volatile species that vaporised due to local equilibrium are formed from the chemical reactions of SiO_2, AL_2O_3, CaO and MgO minerals [16,17]. Based on the generated vapour chemistry, its pressure and concentration during coal combustion. The volatilised mineral matter of the flame condenses by nucleation with small heterogenous ash particles formation on the deposits surface [18,19]. The charred structure is expected to be affected by the vapour pressure, the pressure influences the char fragmentation process, and the behaviour of the ash coalescence derived by the pulverised coal combustion. Multiple porous char particles are produced from coal that contains higher vitrinite contents that leads to smaller ash particles formation, and the coarser ash particles are produced from less porous char particles formed from coal with higher inertinite contents [7,16]. The flue gases flow pattern and ash particle size are influenced by the ash sticking ability on the convective tube surfaces and it velocity through the hot flue gases during the boiler operations.

III. COAL ASH TRANSFORMATION

Investigations on coal ash transformation during combustion had been going on for years. These studies enhance the understanding of the type of coal volatile minerals released, the minerals' chemical transformation, and the mineral interactions with the coal organic matter within the particles [20]. During combustion, the coal mineral undergoes complex chemical and physical transformations resulting from high temperature(s), complex oxidising, and reducing environment by the flame. Ash is then produced as an unburnt residue from coal combustion. Minerals within the coal are classified into extraneous and inherent particles.

a) *Extraneous particles*: these are made up approximately 90 percent by weight of the coal minerals and it can be separated from the coal organic matter by crushing before combustion, and it is usually of finer size between 4μm to 7μm (with a top size between 40μm to 70μm) more than the organic coal particles with a top size of approximately 100μm.

b) *Inherent minerals*: these particles are closely associated within the pores of the organic coal, which cannot be

separated by crushing before combustion. They are made up between 2 to 4 percent by weight of the organic coal particles. During the coal combustion process, the individual species within the coal mineral matter behave differently. There are primarily two major ash formation mechanisms during coal combustion [22].

a) Melting and reaction between the individual grain minerals within the burning coal.
b) Flame vaporisation is followed by subsequent condensation of the coal inorganic components upon the flue gases cooling.

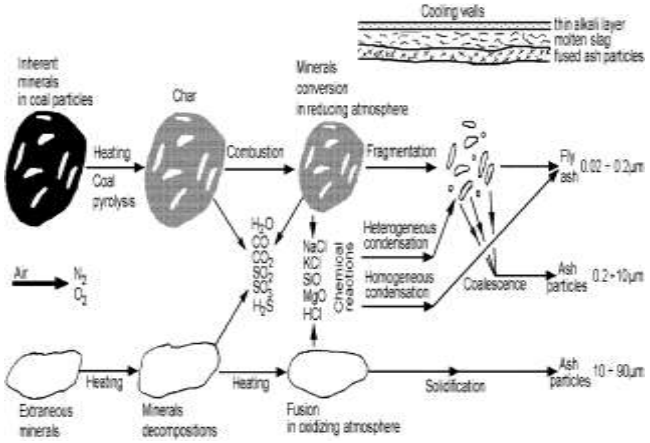


Fig. 1. Coal ash transformation process [25].

Fig. 1 represents the coal ash transformation processes in the coal combustion chamber. The particles fed into the combustor include (1) inherent mineral coal particles and (2) extraneous particles. During coal combustion in the chamber, the volatiles released are burned as char particles. During the combustion process, a temperature higher than the extraneous mineral particles is attained. The extraneous mineral particles contain less than 10 percent of the coal overall mass of the coal organic matter. Finer inherent ash particles (<0.1µm) are transformed within the char particles and is gradually released during the fragmentation process. Minerals deposition and solid phase conversion lead to the gas formation that undergoes homogenous chemical reactions with heterogenous or homogenous cooling. The inherent coal minerals fragmentation and homogenous condensation resulted in the formation of fly ash particles of sizes between 0.02µm to 0.2µm. the coalescence process of the fine mineral matter fragments forms medium size ash particles between 0.2µm to 10µm. Larger ash particles were derived from the extraneous mineral particles with sizes between 10µm to 90µm. The particle size distribution of the fly ash produced from the coal mineral matter in pulverised combustors displays bimodal characteristics [20, 23]. Havier ash particles formed originates from the extraneous minerals matter and are cantered at about 10µm, whereas the smaller fine ash particles ranging from 0.01µm to 0.1µm are formed from mineral inclusions that fuse and coalesce with one other forming ash particles below 0.1µm. a smaller fraction of the coal mineral matter vaporises and condenses subsequently during the coal combustion process [34], which forms a sub-micro fume expanding in size of approximately 0.05µm [20, 24].

IV. METHODOLOGY

In this research, the ash investigation methods used in determining the ash chemistry i.e., elements present (Major, Minor, and Trace) in the ash samples were ICP-OES, ICP-MS, and XRF, and the elemental (ultimate) analysis for the coal samples was carried out using the CHNS method.

The samples for this research were SAMPLE A refering to the bituminous coal fly ash and SAMPLE B refering to the bituminous coal bottom ash. These samples were collected from the Sasol operational site in Secunda, Mpumalanga on the 6th July 2021. The sample preparation for the ICP-OES and ICP-MS for both bituminous coal ash samples are considered. A 0.3g bituminous coal ash sample was weighed and added to a 10mL HNO_3 solution. The samples were then heated at a temperature of 180°C for 25 min and kept at this temperature for a further 10 min. The samples were then quantitatively transferred and diluted with 50mL ultrapure water. The samples were filtered, and an additional two dilutions (2X) were prepared by diluting 0.5mL to 10mL. The samples, calibration standards, and CRM were analysed.

For the XRF both bituminous coal-ash samples were heated separately at a temperature of 105°C in the open air. The samples were then placed in a glazed porcelain crucible and heated from room temperature to 930°C for 30 minutes to determine the samples' loss of ignition (LOI). 0.7g of the heated volatilised samples were fused in a borate fusion disk. The XRF analysis was carried out after the borate fusion disk preparation to determine the major element mixtures of pure chemicals (essentially oxides) and with certified reference materials using a fundamental parameters model. This allows for all combinations of elements within the range of the calibration.

For the ICP analysis, the instruments used were Spectro Arcos ICP-OES and Perkin Elmer NexION 300 ICP-MS, whereas for the XRF the instruments used were an electric fusion machine (TheOx from Claisse) and a wavelength-dispersive XRF spectrometer (MagiX PRO from Malvern Panalytical).

V. RESULTS

The following elements were discovered within the collected ash samples during the ICP-OES, ICP-MS analysis, the XRF analysis was carried out in determining the chemical composition in the coal ash samples, and the elements present in the coal samples were obtained using a CHNS analyser.

TABLE I. ELEMENTS PRESENT IN THE ASH SAMPLES

Element Present	Sample A (mg/kg)	Sample A Mass (%)	Sample B (mg/kg)	Sample B Mass (%)
Si	447300	44.7300	475300	47.5300
Al	326600	32.6600	325300	32.5300
Fe	28100	2.81000	37500	3.7500
Ca	45971.6	4.59716	58749.6	5.8750
Mg	3843.3	0.38433	3600.6	0.3601
K	774.2	0.07742	894	0.0894
Na	2066.6	0.20666	2804.9	0.2805
Ti	17600	1.76000	17100	1.7100
S	1412.7	0.14127	669.7	0.0670
Cd	0.09	0.00001	<0.02	<0.0001
Pb	19.3	0.00193	3.55	0.0004
Hg	0.1	0.00001	<0.02	<0.0001
Se	5.78	0.00058	3.19	0.0003
As	7.58	0.00076	1.88	0.0002

TABLE II. CHEMICAL COMPOSITION OF THE ASH SAMPLES

Chemical Composition	Sample A (mass %)	Sample B (mass %)
Al_2O_3	32.66	32.53
BaO	0.16	0.15
CaO	6.93	9.31
Cr_2O_3	<0.05	0.05
Fe_2O_3	2.81	3.75
K_2O	0.67	0.61
MgO	0.97	1.04
Na_2O	0.91	0.75
P_2O_5	0.35	0.28
SiO_2	44.73	47.57
SO_3	0.14	<0.05
TiO_2	1.76	1.71
LOI	7.32	2.43

TABLE III. BITUMINOUS COAL ULTIMATE ANALYSIS

Element Present	Pulverised Coal (% wt.)	Raw Coal (% wt.)
N_{ar}	1.49	1.57
C_{ar}	52.39	52.11
H_{ar}	2.96	2.99
S_{ar}	0.19	0.19
O_{ar}	42.97	43.14

*ar represent the coal on as received bases
The Oxygen (Oar) is obtained by calculation i.e.
Oar = [100 – (Nar + Car + Har + Sar)]

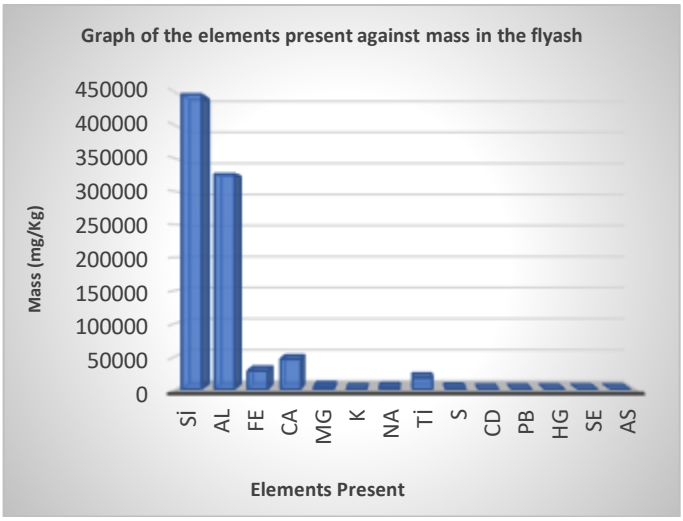


Fig. 2. Graph of the elements present within the bituminous coal fly ash.

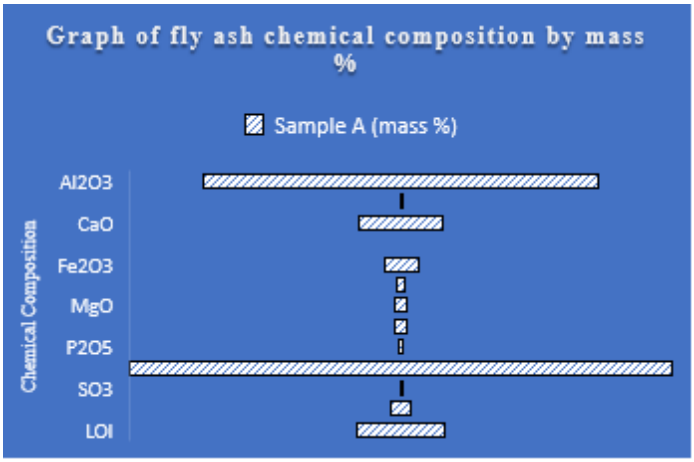


Fig. 3. Graph of the Chemical Composition in the fly ash sample

VI. DISCUSSION OF RESULTS

Table 1 represents the total number of elements present within the coal ash samples. It can be seen from the data obtained that some elements are more prevalent than others within the samples that affect the deposition effects on the boiler heat transfer surfaces. The dominant (major) elements discovered include silicon and aluminium. Elements such as iron, calcium, titanium, magnesium, and sodium are seen as minor elements, and sulphur, potassium, cadmium, lead, mercury, selenium, and arsenic were discovered in both coal ash samples as trace elements. Deposition problems in the coal-fired boiler are caused primarily by the chemical composition of the lighter fine ash particles carried away by the hot flue gases called the fly ash (i.e., Sample A). The elements present in the bituminous coal fly ash in this research are represented by mass % in Fig. 3. Data on the chemical composition of the major elements in an oxidizing atmosphere within the coal ash samples are represented in Table 2, with Al_2O_3 and SiO_2 making up approximately 80% of the total mass within the ash samples. It can be seen in Table 2 that the LOI in Sample A is almost three times greater than that in Sample B because there are more unburnt carbon contents present

in the coal fly ash as compared to the coal bottom ash. Table 3 is a representation of the elements present in the pulverised bituminous coal samples and the raw coal samples. Both coal samples consist of a greater amount of carbon and oxygen contents with a far less amount of sulphur contents. Fig. 3 is the graphical representation of the elemental composition (%mass) contained within the bituminous coal fly ash samples.

The fouling and slagging potential in the boiler furnace can be predicted using the base acid ratio index of the coal ash chemical constituents. The coal ash acidic constituents produce high-temperature ashes measured by % in the oxide i.e., (SiO_2, Al_2O_3, and TiO_2) and the ash temperature is lowered by the coal ash basic oxides % i.e., (Fe_2O_3, MgO, CaO, K_2O, Na_2O). low fusibility and high slagging potential is obtained with a base acid ratio between the ranges of 0.4 – 0.7 [27]. The base acid ratio can be obtained from Eq. 1.

$$\frac{B}{A} = \frac{(Fe_2O_3 + MgO + K_2O + Na_2O + CaO)\%}{(SiO_2 + Al_2O_3 + TiO_2)\%} \qquad (1)$$

From the data obtained in Table 2 for the coal ash Sample A, the base acid ratio is 0.15. By applying the method successfully invented by Attig and Duzy to determine the furnace slagging index (Rs) using the coal ash base acid ratio. This method has been used in identifying the slagging potential for all types of coals as expressed in Table 4. The slagging index (Rs) is calculated using equation 2 [28],

$$R_s = \frac{B}{A} \times S \qquad (2)$$

where *S* is the sulphur weight (%) in the raw coal.

TABLE IV. SLAGGING PREDICTION INDEX

Slagging Type	Slagging index Rs
Low	<0.6
Medium	0.6 - 2.0
High	2.0 – 2.6
Severe	>2.6

The slagging index (Rs) from the results obtained in this research is 0.03 which represents a low slagging potential in the boiler furnace. The ash fusibility and slagging potential in the boiler furnace are proportional to the number of alkalis (Na_2O and K_2O) present in the coal ash.

A proposed prediction for fouling (R_F) on the convective heat transfer tubes based on the quantity of sodium present in the coal ash [28,29] as shown in Table 5, the fouling prediction is obtained from equation 3 [29].

$$R_F = \frac{B}{A} \times Na_2O \qquad (3)$$

TABLE V. FOULING PREDICTION INDEX

Fouling Tendency	R_F
Low	< 0.2
Medium	0.2 – 0.5
High	0.5 – 1.0
Severe	> 1.0

From the Na_2O mass (%) in Table 2 for Sample A, the R_F value obtained in this research is 0.14 which indicates a low fouling tendency in the boiler convection tubes.

The proposed fouling index by [30] is based on the mass percentage of Na_2O in low-grade bituminous coal fly ash in the United States as shown in Table 6.

TABLE VI. FOULING PREDICTION INDEX BASED ON NA_2O PERCENT IN THE COAL ASH

Bituminous coal	
Fouling Potential	***Percentage Na_2O***
Low	< 0.5
Medium	0.5 – 1.0
High	1.0 – 2.5
Severe	> 2.5

From the data in Table 2 for Sample A the Na_2O mass (%) in the bituminous coal ash sample is 0.91, when compared to the bituminous coal ash fouling prediction index in Table 6 corresponds to a medium fouling potential on the convective heat transfer surfaces.

VII. CONCLUSION

This research article is an investigation of the ash chemistry in predicting the fouling potential of the low-grade bituminous coal used for firing the boilers at Sasol synfuels operations in Secunda, Mpumalanga, Republic of South Africa. The article starts with an introduction and a review of relevant literature on the topic. The coal ash's physical and chemical properties were discussed. The transformation processes of the coal mineral matter into ash during combustion in the boiler furnace were discussed. The various transportation mechanisms of the ash formed during coal combustion in the boiler furnace to the convective heat transfer tubes causing deposition problems in the boiler were also discussed. Ash fouling on the boiler tubes was also explained. The methods for data collection in this study which include sample origin, preparations, and the various instruments used in obtaining the results of this study was also explained. The results obtained were displayed using tables and graphical representations and were explained using relevant reviews and mathematical equations on the subject matter. A low fouling tendency prediction, medium fouling, and low slagging potential were discovered in this research based on the laboratory analysis of the coal ash constituents.

REFERENCES

[1] Q. Zhu. Coal sampling and analysis standards. IEA Clean Coal Centre (CCC/235) ISBN 978–92–9029–555–6. 2014.

[2] S.A Benson., M.L Jones., J.N Harb. Ash formation and deposition. Coal Science Technology. 1993. Vol. 20. Pp. 299–37

[3] W.R Seeker., G.S Samuelsen., M.P Heap., J.D Trolinger. The thermal decomposition of pulverized coal particles. Symposium (International) on Combustion. 1981. Vol.18. Iss. 1. Pp 1213-1226.

[4] L.A Hansen., F.J Frandsen., K Dam-Johansen., H.S Sørensen., B.J. Skrifvars. Characterization of ashes and deposits from high-temperature coal straw co-firing. Energy Fuels. 1999. Vol. 13. Iss. 4 Pp 803–816.

[5] G.R. Couch Understanding slagging and fouling in Pulverised fuel combustion. 1994. London: IEA Coal Research. ISBN:92-9029-240-7

[6] H Wang., J.N Harb. Modelling of ash deposition in large-scale combustion facilities burning pulverized coal. Progress in Energy and Combustion Science. Pp. 267–282. 1997.

[7] H.L Wee., H Wu., D.K. Zhang., D French. 2004. The effect of combustion conditions on mineral matter transformation and ash deposition in a utility boiler fired with a sub-bituminous coal, Proceedings of the Combustion Institute. Volume 30. Issue 2. Pp. 2980–2988.

[8] H.L Wee., H Wu., D.K. Zhang. Heterogeneity of ash deposits formed in a utility boiler during Pulverised Fuel combustion. Energy & Fuels. volume 21. Pp 441–450. 2006.

[9] R.P Gupta., T.F Wall., I Kajigaya., S Miyamae., Y Tsumita. Computer controlled scanning electron microscopy of minerals in coal-implications for ash deposition. Progress in Energy and Combustion Science. Volume 24. Pp 523–543. 1998.

[10] E. Raask. Mineral impurities in coal combustion, behaviour problems and remedial measures. Hemisphere Publishing, Bristol, PA. 1985.

[11] L Yan., R Gupta., T.F Wall. Fragmentation behaviour of pyrite and calcite during high-temperature processing and mathematical simulation, Energy & Fuels. Volume 15. Pp 389–394. 2001.

[12] A.P Reifenstein., H Kahraman., C.D.A Coin., N.J Calos., G Miller., P Uwins. Behaviour of selected minerals in an improved ash fusion test: Quartz, potassium feldspar, sodium feldspar, kaolinite, illite, calcite, dolomite, siderite, pyrite, apatite. Fuel. Vol 78. Pp. 1449–1461. 1999.

[13] H.B Vuthaluru., D.K. Zhang. Role of inorganics during fluidised-bed combustion of low-rank coals. 1997 Engineering Foundation Conference on Impact of Mineral Impurities in Solid Fuel Combustion (Ed. Gupta R.P., Wall T.F., Baxter L.). P. 309, Kluwer Academic Press, New York. 1999.

[14] R.W Bryers. Fireside slagging, fouling and high-temperature corrosion of heat transfer surface due to impurities in steam raising fuels. Progress in Energy and Combustion Science. Vol 22. Pp. 29–120. 1996.

[15] H.B Vuthaluru., D.K Zhang. Effect of Ca and Mg bearing minerals on particle agglomeration and defluidisation during fluidised-bed combustion of a South Australian lignite. Fuel Processing Technology. Vol 69. Iss 1. Pp. 13–27. 2001.

[16] H Wu., G Bryant., T.F Wall. Ash liberation from included minerals during combustion of pulverised coal: The relationship with char structure and burnout. Energy & Fuels. Vol 13. Issue 6. Pp 1197–1202. 1999.

[17] J.J Helble., A.F Sarofim., Influence of char fragmentation on ash particle size distributions. Combustion and Flame. Vol 76. Pp 183–196. 1989.

[18] J.C Van Dyk., S.A Benson., M.L Laumb., B Waanders. Coal and coal ash characteristics to understand mineral transformations and slag formation. Fuel. Vol 88. Pp 1057–1063. 2009.

[19] S.A. Benson. Ash Formation and Behavior in Utility Boilers: Parts 1–12. 1995 Microbeam Technologies Inc. http://www.microbeam.com/Articles/Articles.htm

[20] M Neville., A.F Sarofim. The fate of sodium during pulverized coal combustion. Fuel. Volume 64, Issue 3. Pp 384 – 390. 1985.

[21] J Tomeczek., H Palugniok. Kinetics of mineral matter transformation during coal combustion. Fuel. Vol 81. Iss 10. Pp. 1251-1258. 2002.

[22] L Zhang., Y Ninomiya., T Yamashita. Formation of submicron particulate matter during coal combustion and influence of reaction temperature. Fuel. Volume 85, Issue 10. Pp. 1446–1457. 2006.

[23] M Sadakata., M Mochizuki., T Sakai., K Okazaki., M Ono. Formation and behavior of submicron fly ash in pulverized coal combustion furnace. Combustion and flame. Vol 74, Iss 1. Pp.71-80. 1988.

[24] R.C Flagan. S.K Friedlander. 1978. Recent developments in aerosol science.

[25] J.R Qiu., F Li., Y Zheng., C.G Zheng., H.C Zhou. The influences of mineral behaviour on blended coal ash fusion characteristics. Fuel, Vol 78, Iss 8. Pp.963-969. 1999.

[26] L. L Baxter., R. E Mitchell. The release of iron during the combustion of Illinois No. 6 coal. Combustion and Flame. Vol 88, Issue 1. Pp 1-14. 1992.

[27] J.G Singer. Combustion: Fossil Power Systems, Combustion Engineering, Inc, Windsor, CT. 1991.

[28] R.C Attig. A.F Duzy. Coal ash deposition studies and application to boiler design. In: Proceedings American Power Conference. Volume 31. Pp. 290–300. 1969.

[29] E.C. Winegartner. Coal Fouling and Slagging Parameters, Research Department, ASME. p. 34. 1974.

[30] R.W. Bryers. Fireside behavior of minerals in coal. In: Symposium on Slagging and Fouling in Steam Generators. P. 63. 1987.

Machine Learning in Service of Nephrology Patients' Mortality Rate Assessment

Nevena Radović
Department of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
nevenar@ucg.ac.me

Vladimir Prelević
Clinic for nephrology
Clinical Center of Montenegro
Podgorica, Montenegro
vladimir.scopheurope@gmail.com

Milena Erceg
Department. of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
milena.zogovic@gmail.com

***Abstract*— Machine learning models are increasingly used for the examination of biomedical data. In this paper, we are employing two of them, support vector machine algorithm and *K*-means clustering algorithm, for the analysis of Montenegro's national nephrology database with the goal to provide a quality assessment of the mortality rate for patients who undergo hemodialysis treatment. Assessment accuracy of 94.12% is achieved with a support vector machine algorithm and is based on the usage of four parameters that are obtained relatively easily, during one examination of the patient. This result is supported by results obtained with *K*-means clustering algorithm.**

Keywords*— Hemodialysis; Mortality rate assessment; Malnutrition inflammation score; Classification; Clustering;

I. INTRODUCTION

Chronic kidney disease (CKD) is recognized as a growing contemporary public health issue worldwide, with cardiovascular diseases as the leading cause of morbidity and mortality in CKD patients. Hereupon, CKD may be considered today as an independent cardiovascular factor [1]. The number of patients with progression of CKD in the end-stage renal disease (ESRD) has significantly increased in the last three decades, [2].

ESRD patients could be treated with three different modalities of renal replacement therapy (RRT): hemodialysis (HD), peritoneal dialysis (PD), and renal transplantation (RT). The optimal choice of the modality of RRT determines not only the quality of patient's life but their survival rate as well. Hemodialysis is the most frequent RRT modality in ESRD patients, [3].

Aldo, it is the most frequent treatment of ESRD patients, HD correlates independently to accelerated vascular lesions through separate immune-mediated and nonimmune-mediated mechanisms. Additionally, it can be associated with heart rhythm disorder, coronary artery diseases, and adverse and fatal cardiac events, [4]. Finally, hemodynamic changes during HD are recognized as reasons for unfavorable cardiovascular outcomes, [5], [6]. Some other important contributors to high cardiovascular risk are oxidative stress and low-grade chronic inflammation. The complications specific to the group of patients in question like anemia, mineral bone disease, and malnutrition are also responsible for increased mortality, [7].

Quality assessment of factors that influence mortality and morbidity in HD patients can be beneficial in the prediction of survival rate and optimal long-term outcome, which is the main aim of this research.

II. MATERIALS AND METHODS

A. Database and Metrics

In our research, we are using the national nephrology database of Montenegro. This database contains data from approximately 87% of all HD patients in the country, which makes it safe to say that a national cohort is provided. The database itself is a result of a 2-year study (April 2018 to April 2020) and gathers data from 5 different nephrology centers throughout the country.

The records of 102 patients undergoing HD longer than 3 months were examined. Precisely, 55 female and 47 male patients, aged 40 to 91 were observed. Fig. 1 shows the distribution of basic diseases (19 different ones in total) these patients suffered from that resulted in HD as a treatment. During the analysis, data were divided into two datasets: test and training datasets of equal size.

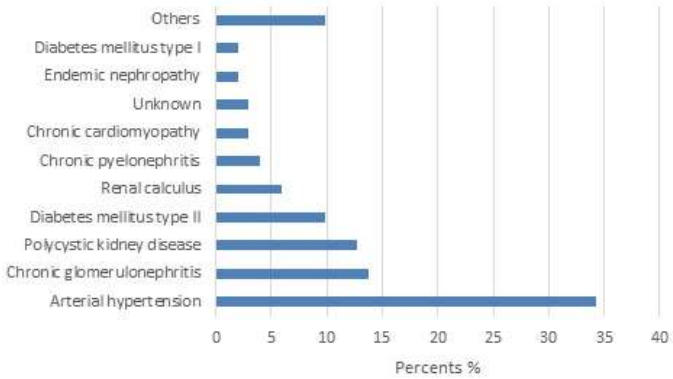


Fig. 1. Percentage share of basic diseases (that resulted with HD as a treatment) in the total number of patients from the observed database. Note that Others gathers 10 different diseases which were each diagnosed in less than 1% of patients, while Unknown refers to patients without a determined diagnosis.

Nine different parameters per patient were observed. They are listed in Table 1, along with a brief explanation of what they present.

B. Classification and Clustering Algorithms

Classification algorithms as models of Machine learning (ML) are increasingly used in different scientific and non-scientific areas with great success. Support Vector Machine (SVM), introduced by Vapnik, [8], is especially efficient in numerous fields of biomedical engineering [9] - [14]. This is the reason why we have selected it as a classification algorithm used in our analysis.

SVM determines the hyper-plane that divides two sets of linearly separable pattern vectors optimally. Precisely, that kind of hyper-plane is set in a way that provides maximal distance between the two data points that are part of two different classes, [8], [15]. In case we have sets that cannot be separated linearly, kernel functions are used to map the data to a high-dimensional feature space, where they can be linearly separated. The most common kernel function used for this purpose is the Gaussian function:

$$C(a,b) = \exp\left(-\|a-b\|^2 / 2\varepsilon^2\right) \quad (1)$$

Here, a and b are observed feature vectors, while $\|a-b\|^2$ refers to their Euclidean distance. The width of the kernel function is determined by the parameter ε.

TABLE I. PARAMETERS FROM THE OBSERVED DATABASE USED FOR ANALYSIS

Parameters	**Explanation**
I Malnutrition inflammation score (**MIS**)	Data regarding the nutritional status of a patient as well as patients' related medical history (obtained through a questionnaire, objective laboratory, and biochemical measurements), [16]
II Septum III Left ventricular mass index (**LVMI**) IV Last wall (**LW**)	Morphological cardiac parameters (obtained by echocardiography)
V Hemoglobin (**HGB**)	Parameter for the assessment of anemia (obtained through laboratory)
VI Myeloperoxidase (**MPO**) VII Total antioxidative system (**TAS**) VIII Super oxide dismutase (**SOD**) IX High sensitive C reactive protein (**hsCRP**)	Parameters for assessment of oxidative stress, [5], [6] (obtained through laboratory).

To further examine available data and additionally verify results achieved with SVM, we have used another ML model. It is the *K*-means clustering algorithm, [17], [18].

Similar to SVM, *K*-means clustering algorithm is of great importance in practical applications. The idea of this algorithm is to divide the observed dataset into *K* clusters that do not overlap. All of the defined clusters have a central point, usually called a centroid, that can but doesn't have to be a real object from the dataset. The position of the cluster's centroid determines the size and number of objects in the cluster. Usually, we are setting a criterion that enables minimization of the distance between the cluster's centroid and the dataset object. In our case, we have selected the minimization of the Euclidean distance as a criterion.

III. RESULTS AND DISCUSSION

Training of the SVM is performed based on the parameters from Table 1 in the following manner:

- Individual parameters (I to IX) were loaded into a classifier, assessments based on each of them are done and the best assessment is selected;
- Every combination of two parameters (I and II, II and III, III and IV,…) were loaded into a classifier, assessments based on each pair were done and the best assessment is selected;
- Every combination of three parameters (I, II and III; I, II and IV; I, II and V;…) were loaded into a classifier, assessments based on each triplet were done and the best assessment is selected;

  …
- All parameters were loaded into classified and an assessment is done.

The best assessments from each above-stated bullet are graphically presented in Fig. 2. Parameters combination (*x*-axis) denoted with 1 to 9 represents the following combination of parameters: 1 (I), 2 (I, II), 3 (I, II, III), 4 (I, II, III, IV), 5 (I, II, III, IV, V), 6 (I, II, III, IV, V, VI), 7 (I, II, III, IV, V, VI, VII), 8 (I, II, III, IV, V, VI, VII, VIII), 9 (I, II, III, IV, V, VI, VII, VIII, IX).

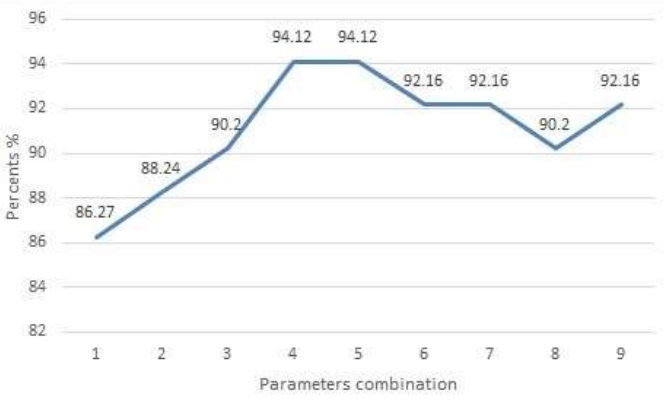


Fig. 2. Results of mortality rate assessment accuracy when applying SVM algorithm on parameters from Table 1.

Fig. 3. Distribution of errors to FP and FN during the mortality rate assessment when applying SVM algorithm on parameters from Table 1.

Mortality rate assessment accuracy (MRAA) is calculated as a ratio of the number of correct assessments and the number of the total assessments made with the SVM. Analogously, mortality rate assessment error (MRAE) is calculated as a ratio of the number of false assessments and the number of total assessments made with the SVM.

MRAEs can be made in two ways: we can falsely assess that mortality outcome will occur and it doesn't, or we can falsely assess that mortality outcome won't occur and it does. The first type of MRAE is known as false positive assessments (FP), while the second type is known as false negative assessments (FN).

MIS stands out as a parameter that individually provides the best results (MRAA=86.27%, Fig. 2). This means that solely based on this one parameter, which is obtained relatively easily and in a non-invasive manner, we can with the certainty of 86.27% say will the mortal outcome occur or not. With additional three parameters, Septum, LVMI, LW, MRAA reaches its maximal value of 94.12%. Further increase in the number of parameters will not contribute to an increase in assessment precision, Fig. 2. This is good news first for patients and then for attending physicians as well. Namely, further time and money-consuming diagnostic methods will not be necessary, which means that patients will not be exposed to additional examinations from one side, and the attending physician will faster have all necessary data for quality assessment.

In Fig. 3, the distribution of FP and FN is presented. Note that, same as in the case of Fig. 2, parameters combination (*x*-axis) denoted with 1 to 9 represents the following combination of parameters: 1 (I), 2 (I, II), 3 (I, II, III), 4 (I, II, III, IV), 5 (I, II, III, IV, V), 6 (I, II, III, IV, V, VI), 7 (I, II, III, IV, V, VI, VII), 8 (I, II, III, IV, V, VI, VII, VIII), 9 (I, II, III, IV, V, VI, VII, VIII, IX). MRAE=5.88% that is obtained in case of assessment based on the mentioned four parameters (MIS, Septum, LVMI, LW) is in 66.67% of cases consequence of falsely positive assessments. This is convenient, having in mind that FPs in this kind of analysis are "lesser evil".

To verify the above conclusions, as announced earlier in the paper, we have used *K*-means clustering algorithm. We intended to establish relations between parameters that SVM recognized as crucial ones for mortality rate assessment accuracy.

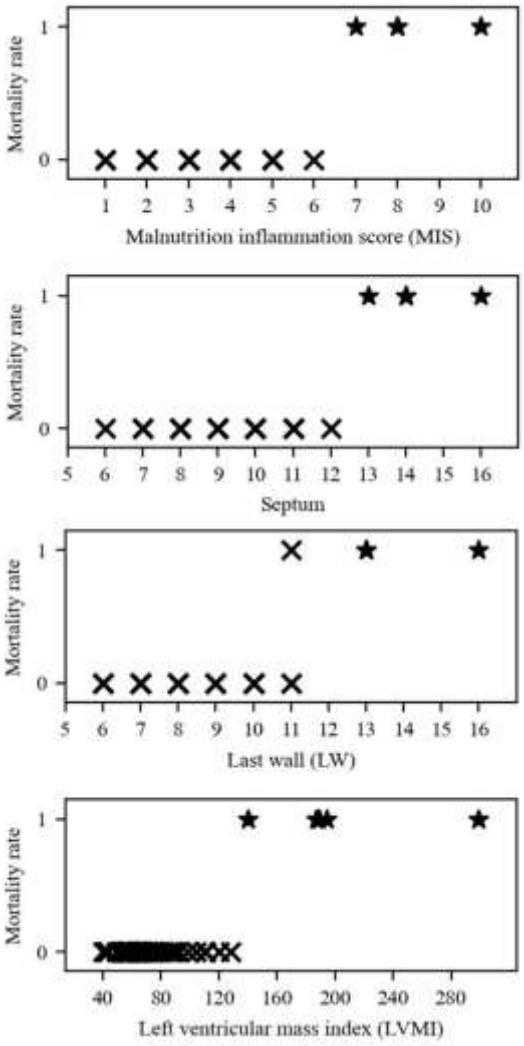


Fig. 4. The clustering of the mortality assessment results based on the individual parameters (MIS, Septum, LW, LVMI), is done by *K*-means clustering algorithm. The stars (value 1) represent positive mortal assessments, while x-es (value 0) represent negative mortal assessments.

"One-on-one" dependency between mortality rate assessment and each of the parameters individually: MIS, Septum, LVMI, LW, has been examined, and results are presented in Fig. 3. The clear distinction between two clusters (positive and negative moral outcome) can be noticed when clustering is performed regarding MIS, Septum, and LVMI. Only in the case of clustering performed regarding LW one instance of inconsistency can be noted. Namely, when LW takes the value 11, mortality outcome is placed in both clusters (positive and negative moral outcome).

Mutual interaction between mortality rate assessment and every pair of parameters: MIS, Septum, LVMI, and LW, has been examined at the end, and results are presented in Fig. 4. Again, the clear distinction between the two clusters is present in almost all cases. Precisely, only in the case of clustering performed regarding pair LW/Septum one instance of inconsistency can be noted (Fig. 5).

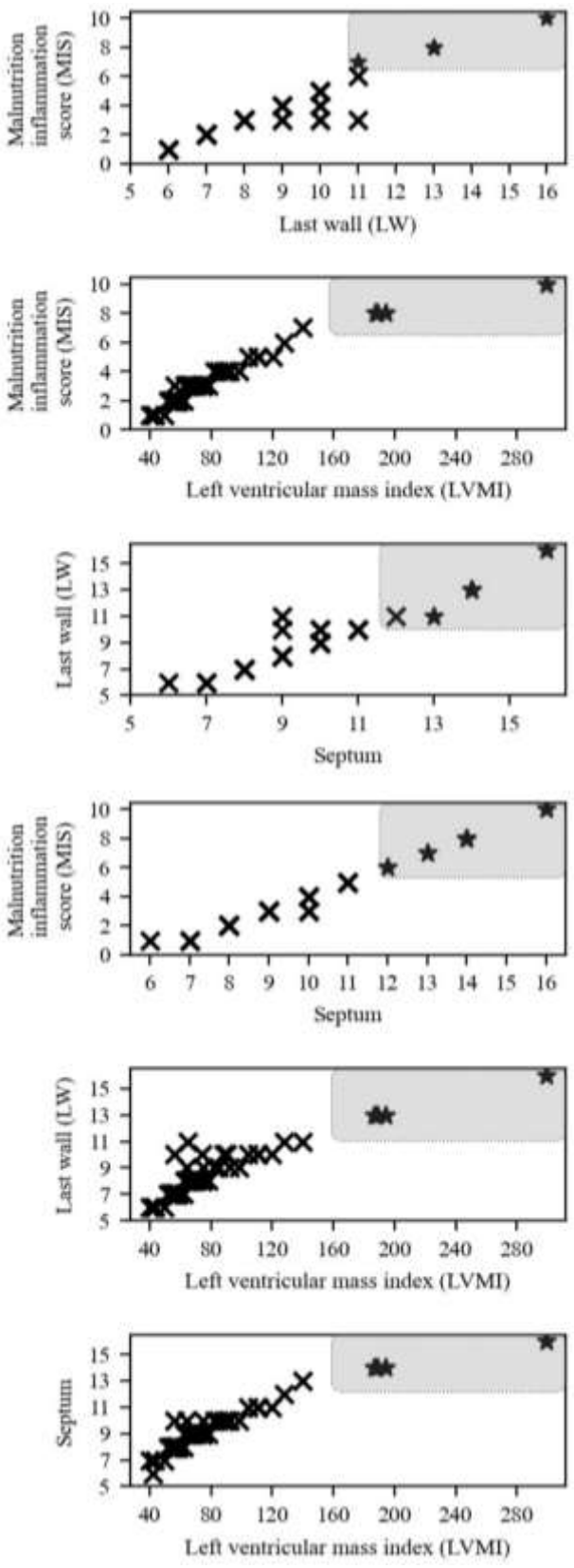


Fig. 5. The clustering of the mortality assessment results based on the mutual combination of two parameters, done by *K*-means clustering algorithm. The stars in the shaded rectangle represent positive mortal assessments, while x-es outside the shaded rectangle represent negative mortal assessments.

Obviously, results achieved with *K*-means clustering algorithm support those achieved with SVM.

IV. CONCLUSIONS

Biomedical data are traditionally analyzed with various statistical methods. However, in the last couple of years usage of ML algorithms for this purpose is showing a substantial growing trend. By using two state-of-the-art classification and clustering algorithms, SVM and *K*-means clustering algorithm, we are analyzing Montenegro's national nephrology database with the goal to provide a quality assessment of the mortality rate for HD patients. Assessment accuracy of 94.12% is achieved with SVM, and is based on the usage of four parameters that are obtained relatively easily and with (for patients) noninvasive techniques. This result is supported by results obtained with *K*-means clustering algorithm.

REFERENCES

[1] M. Cozzolino, M. Mangano, A. Stucchi, P. Ciceri, F. Conte and A. Galassi, "Cardiovascular disease in dialysis patients", Nephrol Dial Transplant, vol. 33, no.6, pp. 28–34, 2018. doi: 10.1093/ndt/gfy174

[2] G. T. Obrador, X. Rubilar, E. Agazzi & J. Estefan, "The Challenge of Providing Renal Replacement Therapy in Developing Countries: The Latin American Perspective", American Journal of Kidney Diseases, vol. 67, no. 3, pp. 499–506, 2016. doi:10.1053/j.ajkd. 2015.08.033.

[3] Z.Y. Duan, L. Jijun, G. Cai, C. Xiangmei, C. Fengkun, "Prevalence, outcome and modalities of renal replacement therapy in burn patients: a systematic review and meta-analysis", Nephrol Dial Transplant, vol.34, no. 5, 2019. gfz106.FP710, https://doi.org/10.1093/ndt/gfz106.FP710

[4] J. Jagieła, P. Bartnicki & J. Rysz, "Selected cardiovascular risk factors in early stages of chronic kidney disease", Int Urol Nephrol, vol. 52, no. 2, pp. 303–314, 2020. https://doi.org/10.1007/s11255-019-02349-1

[5] V. Liakopoulos, S. Roumeliotis, S. Zarogiannis, T. Eleftheriadis, P.R. Mertens, "Oxidative stress in hemodialysis: Causative mechanisms, clinical implications, and possible therapeutic interventions", Seminars in Dialysis, vol. 32, no. 1, pp. 58-71, 2019. doi: 10.1111/sdi.12745.

[6] Y. Zhang, P. Murugesan, K. Huang, H. Cai, "NADPH oxidases and oxidase crosstalk in cardiovascular diseases: novel therapeutic targets". Nature Review Cardiology, vol. 17, no. 3, pp. 170–194, 2020. doi: 10.1038/s41569-019-0260-8.

[7] F. Nuhu, S. Bhandari, "Oxidative Stress and Cardiovascular Complications in Chronic Kidney Disease, the Impact of Anaemia", Pharmaceuticals, vol. 11, no. 4:103, 2018. https://doi.org/10.3390/ph11040103

[8] V. Vapnik, "The nature of statistical learning theory", Berlin: Springer-Verlag, 1995.

[9] A. Onan, "Biomedical text categorization based on ensemble pruning and optimized topic modelling", Computational and mathematical methods in medicine, vol. 2018, Article ID 2497471, 22 pages, 2018, https://doi.org/10.1155/2018/2497471.

[10] Q. Li, C. Rajagopalan, G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach", IEEE Transactions on Biomedical Engineering, vol. 61, no. 60, pp. 1607–1613, 2014.

[11] N. Nuryani, S.S.H. Ling, H. T. Nguyen, "Electrocardiographic signals and swarn-based support vector machine for hypoglycemia detection", Annals of Biomedical Engineering, vol. 40, no. 4, pp. 934–945, 2012.

[12] T.T. Pham, C. Thamrin, P. D. Robinson, A. L. McEwan, P.H.W. Leong, "Respiratory artefact removal in forced oscillation measurements: A machine learning approach", IEEE Transactions on Biomedical Engineering, vol. 64, no. 8, pp. 1679–1687, 2017.

[13] J.L. Sapp, M. Bar-Tal, A.J. Howes, J.E. Toma, A. El-Damaty, J. W. Warren, P. J. MacInnis, S. Zhou, B. M. Horáček, "Real-time localization of ventricular tachycardia origin from the 12-Lead electrocardiogram", JACC: Clinical Electrophysiology, vol. 3, no. 7, pp. 687-699, 2017.

[14] R. Tao, S. Zhang, X. Huang, M. Tao, J. Ma, S. Ma, C. Zhang, T. Zhang, F. Tang, J. Lu, C. Shen, X. Xie, “Magnetocardiography based ischemic heart disease detection and localization using machine learning methods”, IEEE Transactions on Biomedical Engineering, vol. 66, no. 6, pp. 1658–1667, 2019.

[15] S. Russel, P. Norvig, “Artificial Intelligence: A modern approach”, Pearson. 3rd Edition, 2016.

[16] H. Bakkal, O. S. Dizdar, S. Erdem, S. Kulakoğlu, B. Akcakaya, Y. Katırcılar, K. Uludag, “The Relationship Between Hand Grip Strength and Nutritional Status Determined by Malnutrition Inflammation Score and Biochemical Parameters in Hemodialysis Patients”, Journal of Renal Nutrition, vol. 30, no. 6, pp. 548-555, 2020.

[17] J. E. Mac Queen, “Some methods for classification and analysis of multivariate observations.” Proceedings of the Fifth Berkley Symposium Math. Stat Prob, 1967, vol. 1, pp. 281-297.

[18] J. Han, M. Kamber, “Data mining Concepts and techniques”, 2nd edition, Morgan Kaufmann Publishers, 2007.

Control System for Quasi-Z-source Cascaded H-bridge Multilevel Inverter with PV Power Generation and Battery Energy Storage System

Pablo Horrillo-Quintero
SURET Research Group
Department of Electrical Engineering
University of Cadiz
Algeciras, Spain
pablo.horrillo@uca.es

Pablo García-Triviño
SURET Research Group
Department of Electrical Engineering
University of Cadiz
Algeciras, Spain
pablo.garcia@uca.es

Raúl Sarrias-Mena
SURET Research Group
Dept. Engineering in Automation, Elect. Comp. Arch. & Netw.
University of Cadiz
Algeciras, Spain
raul.sarrias@uca.es

Carlos A. García-Vázquez
SURET Research Group
Department of Electrical Engineering
University of Cadiz
Algeciras, Spain
carlosandres.garcia@uca.es

Luis M. Fernández-Ramírez
SURET Research Group
Department of Electrical Engineering
University of Cadiz
Algeciras, Spain
luis.fernandez@uca.es

***Abstract*—Quasi-Z-source cascaded H-bridge multilevel inverters with energy storage (ES-qZS-CHBMLIs) have considerable advantages over traditional multilevel inverters. In particular, they allow a balanced voltage level to be achieved on the DC link, and power conversion is performed in a single stage without the need for an additional DC/DC converter. Furthermore, a higher voltage gain and improved output waveform (due to the elimination of switching dead times) are attained. The battery energy storage system (BESS) integrated into each cascade converter ensures energy storage to support the renewable power, photovoltaic (PV) power plant in this case, connected to the input of this converter. This paper presents a new energy management system (EMS) for a grid-connected ES-qZS-CHBMLI with PV power generation. The EMS guarantees the correct operation of the BESS independently, limiting the state-of-charge (SOC) between the minimum and maximum safety values, and setting its maximum charge or discharge power to its nominal power, extending the useful life of the BESS. A maximum power point tracking (MPPT) based on the Perturb & Observe (P&O) algorithm ensures optimal operation of the PV power plants connected to each cascade converter. The results obtained from MATLAB-Simulink on a grid-connected single-phase configuration based on an ES-qZS-CHBMLI with three cascade qZSI, each connected to a 4.8 kW PV power plant and operating in different conditions, validate the proposed configuration and the control system.**

Keywords—Quasi-Z-source cascaded H-bridge multi-level inverter, energy storage system, energy management system, power conversion, PV power plant.

I. INTRODUCTION

The contribution of renewable sources in current power systems is significant. Furthermore, it has to continue growing in the near future in order to comply with the international commitments to reduce the emission of greenhouse gases. Therefore, research efforts that improve the energy generation and management from these sources are prominent. Among the different renewable sources, PV generation covers a relevant portion of the energy generation mix in every advanced country [1]. Typically, PV systems use a topology based on two power converters when connected to an AC grid, one DC/DC converter that allows implementing the MPPT strategy of the panels, and a DC/AC inverter at the point of common coupling. The voltage source inverter (VSI) is commonly used for this purpose [2]. In the recent years, a new topology is gaining force to substitute the aforementioned configuration. Impedance, or quasi-impedance source inverters (namely ZSI and qZSI, respectively) can be employed for the grid connection of PV generation. These converters present a modified structure with an impedance network at the DC side of a conventional VSI. This allows achieving a higher voltage boost compared to the VSI, as well as providing the MPPT capability and DC/AC conversion in a single stage [3]. Therefore, the DC/DC converter of the conventional configuration can be omitted, thus reducing cost, losses and complexity in the topology and the control. Moreover, the qZSI has the advantage of drawing a continuous current from the PV panels, in contrast with the ZSI, which demands a pulsating current that is difficult to measure and handle by the converter circuitry [3], [4].

Despite the improvements achieved with the qZSI compared to the VSI in terms of voltage boost, there are large-scale applications where the voltage gap between the PV system and the grid is excessively high for a single qZSI. In such circumstances, several converters can be arranged to build a cascaded H-bridge multilevel inverter (CHBMLI) based on the qZSI (namely qZS-CHBMLI). As indicated in [5], the cascaded configuration allows an independent MPPT for each PV string. Additionally, the system is easily scalable by adding more

cascaded modules, there is no need for a voltage transformer at the output, and the output filter is also smaller compared to other multilevel configurations [6]–[8]. The authors in [6] claim the first application of a cascaded multilevel configuration for qZSI, which merges the advantages of both concepts. For single-phase qZSI, a typical topology includes four power switches in each module [9]. When integrated in a qZS-CHBMLI, each PV string uses these modules. This maximizes the PV power generation due to the aforementioned ability to develop independent MPPT in each module of the CMI.

The intermittent generation of the PV panels is often mitigated with the use of an energy storage system that can handle power variations and provide controllable supply to the grid [10]. Electrochemical batteries are a common choice for such purpose. The use of power converters based on impedance sources offers a remarkable benefit in this sense, since they allow connecting and controlling the BESS into the impedance network of the qZSI without needing an additional DC/DC converter [9], [10]. In a qZS-CHBMLI configuration, a BESS can be integrated in each of the modules to form a qZS-CHBMLI with energy storage (ES-qZS-CHBMLI). Hence, a BESS supports the operation of each PV string, which adds a lot of flexibility and reliability to the system against changes in solar radiation or grid demand.

On the other hand, such a flexible configuration also requires paying special attention to the control and energy management among all the elements involved. In this regard, this work presents an energy management system (EMS) for an ES-qZS-CHBMLI employed for the grid-connection of PV power generation. The proposed EMS guarantees the independent operation of the BESS integrated in the inverter, maintaining their SOC in a safe range, and limiting their maximum charge and discharge power to their rated power. The regulation of the SOC and the maximum power exchange prevents the battery from premature failure.

Then the rest of the paper is structured as follows: Section II describes the system under study, section III focuses on the control system, simulation results are discussed in Section IV, and the conclusions of the study are presented in Section V.

II. SYSTEM UNDER STUDY

Fig. 1 shows the configuration of the system considered in this study. It has three modules connected in series to a single-phase grid. Each module is composed of a PV power plant, an impedance network with BESS and a single-phase VSI. This module structure allows the PV power plant to work at the maximum power point tracking according to the incident radiation, to control the active and reactive power delivered to the grid ensuring the demanded reference powers and to balance the power difference between the power demanded by the grid and the power generated by the PV power plant with the support of the BESS [10]. Furthermore, all the modules are coordinated with each other to achieve a balanced power distribution.

The input source of each module is a PV power plant with a terminal capacitor to stabilize the output voltage of the PV array, v_{pv_n} (n=1, 2, 3), which is controlled to operate the PV array at the maximum power point (MPP).

The single-phase qZSI (impedance network+VSI) boosts the DC voltage from the source (v_{pv_n}) and converts to AC voltage in a single stage. In this configuration, the traditional DC/DC boost converter used in PV power plants is replaced by an impedance network and the boost control is transferred to the VSI, the only controllable converter of the qZSI. This fact allows two operating states in the VSI: a) Non-Shoot-Through (NST) state, that is, the common mode of a VSI, where the inverter can be connected (active state) or disconnected (zero state); and b)Shoot-Through (ST) state, which is typical of the qZSI, where two switches of the same leg are connected at the same time and, thus, the inverter is short-circuited while the impedance network disconnects from the input source [11]. This state allows controlling the boost voltage according to the time the inverter stay in this state. Thus, the voltage boost ability (B) of the qZSI is:

$$B = \frac{v_{pn}}{v_{pv}} = \frac{1}{1 - 2\,D} \tag{1}$$

where v_{pn} is the output voltage of the impedance network, and D is the ST duty ratio, which is a dimensionless parameter that relates the switching cycle (T) and the interval of ST state (T_{ST}), and thus, the greater T_{ST}, the higher D and B.

The BESS is a Lithium-Ion battery connected in parallel with the capacitor C_{2_n}, without additional DC/DC converter. Its purpose is to smooth the fluctuations of the PV power plant. Therefore, the VSI controls not only the AC/DC power conversion with the power flow to the grid, but the DC/DC boost, the MPPT voltage control of the PV power plant and the power flow exchange with the BESS.

The output of the impedance network is connected to a single-phase VSI with H-bridge topology [12]. This topology has four switches in its traditional structure with two legs and two switches connected in series per leg. This converter has three possible output voltage levels: $+V_{pn,}$ $-V_{pn}$ and zero. The first two levels are the active states, where two diagonal switches are connected in the positive half-cycle and the other diagonal switches are connected in the negative half-cycle. In the last level, zero state (freewheeling mode), only the upper or the lower switches are connected and without current flow.

To achieve these three output voltage levels, a phase-shift pulse-width modulation (PS-PWM) based on Simple Boost Control (SBC) for single-phase converters is considered in this study. Thus, a modified unipolar sinusoidal pulse-width modulation (SPWM) strategy is applied to each module, where the reference modulating signal from the converter control is compared with a triangle carrier to generate the switching signals for the positive half-cycle (between 0 and $+V_{pn}$). A second modulating signal, phase shifted 180º from the first one, oversees the switching signals for the negative half-cycle (0 and $-V_{pn}$).

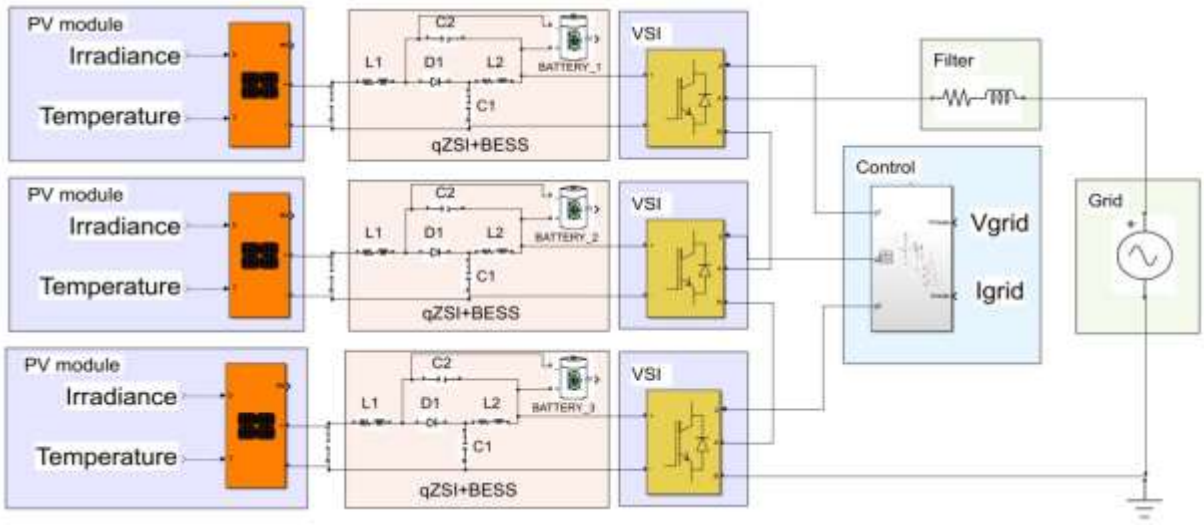


Fig. 1. Grid-connected ES-qZS-CHBMLI with PV power generation under study.

This modulation only generates the pulses for the NST states of the qZSI, whereas the SBC modulation is applied for the ST states [13]. SBC compares the carrier signal with two continuous references (V_p and V_n) defined as:

$$V_p = 1 - D \tag{2}$$

$$V_n = D - 1 \tag{3}$$

These references are limited by the maximum value of D:

$$D_{max} = 1 - M \tag{4}$$

The modules connected in series needs to be coordinated to generate the desired output voltage. A CHBMLI configuration with seven staircase levels is used in this work [14]. The operating principle is to coordinate each module with specific angles to generate the multilevel output voltage with a minimum of commutations. Thus, this configuration boosts the output voltage level, reduces total harmonic distortion (THD) and switching power losses, and increases fault tolerances. The coordination between modules with H-bridges requires the PS-PWM modulation. This natural extension of traditional PWM is very simple, in which a phase shift is applied between the carriers of the contiguous modules to achieve a staircase multilevel output voltage. In this study, a phase shifted of 60º (180º/3 modules in series) is implemented. This angle is obtained from a uniform distribution of the number of modules in half a cycle and it results less distortion and a better distribution of the average power between the modules.

III. CONTROL SYSTEM

The proposed control system is composed of a MPPT control subsystem for the PVs panels, a control subsystem for the active and reactive power injected to the grid and an EMS for the total battery power.

A. Independent MPPT control

The MPPT control subsystem is responsible for achieving the maximum power of the PV panels according to the incident radiation and temperature. It is based on the Perturb and Observe (P&O) algorithm, and its output variable is the PV reference voltage (V_{pv_n}). Each n-module has its own MPPT-n.

At that point, a PI controller is used to reach the required reference voltage on the PV panels by adjusting D_n. In this case, D_n is also employed to boost the PV voltage to a higher level. This scheme allows to obtain an independent MPPT control for each module, which implies obtaining the maximum power for different environmental conditions in the different modules.

B. Power Control

The main objective of the grid power control subsystem is to regulate the power injected to the grid according to a set reference for a balanced system. When the PV power fluctuates, the BESS will support the PV generation, balancing the excess or insufficient power.

For controlling the active power (P) and reactive power (Q), a synchronous reference frame d-q is used. The traditional d-q analysis from a three-phase system cannot be used for a single-phase system. Because of this, it is necessary to transform the three-phase current and voltage variables into orthogonal stationary α-β components, where the imaginary signal β has the same characteristics but it is delayed by ¼ period with respect to the real component α. Then, the orthogonal α-β components are converted into a rotation coordinate frame d-q.

The total BESS power is obtained as output from a PI controller, which compares the reference grid power with the measured grid power. The distribution of the total BESS power ($P_{bat,tot}$) to the n-qZSI-PV-modules($P_{bat,n}$)is carried out by the EMS, resulting:

$$P_{bat,tot} = \sum P_{bat,n} \tag{5}$$

The BESS power of each module ($P_{bat,n}$) is divided by the measured voltage of each BESS ($V_{bat,n}$) to obtain the reference BESS current in each module ($i^*_{bat,n}$):

$$i^*_{bat,n} = \frac{P_{bat,n}}{V_{bat,n}} \tag{6}$$

Using a PI controller, the reference BES current ($i^*_{bat,n}$) is compared with the measured BES current i_{bat}, whose output variable is the total power that each n-module has to contribute to the system, denoted as P^*_n. Because the three cascaded qZSI modules are connected in series, the injected grid current is the same for all of them, and its peak value is calculated as:

$$i^*_{grid} = \frac{2\,P_{tot}}{V_{grid}} \tag{7}$$

where P_{tot} denotes the sum of the total power of each module and V_{grid} is the grid voltage.

A phase locked loop (PLL) is applied to obtain the phase angle of the grid voltage, and ensure a power factor equal to unity. A current control loop is employed to balance the grid current, by adjusting the *d*-component of the grid voltage (V_d) in the *d-q* frame. The PS-PWM technique allows to derive the *d*-component of the modulation index M ($m_{d,n}$) for each module as follows:

$$m_{d,n} = \frac{2a_n V_d}{V_{pn,n}} \tag{8}$$

$$a_n = \frac{P^*_n}{P_{tot}} \tag{9}$$

$$V_{pn,n} = \frac{V_{C1}}{1 - D_n} \tag{10}$$

where $a_n$denotes the ratio between the power of each module (P^*_n) and the total power (P_{tot}), $V_{pn,n}$ is the DC output voltage of the impedance network, and V_{C1} is the voltage across the capacitor C_1 of the impedance network.

Another current control loop is implemented to balance *Q*, with a similar control scheme to *P*, but using a PI controller to compare the required and measured Q, resulting the *q*-component of *M* ($m_{q,n}$).

The shoot-through duty radio (D_n) and the α-component of the modulation index ($m_{\alpha,n}$) are combined with the phase angle of the grid voltage to produce the desired gate signals for the IGBTs of the ES-qZS-CHBMLI.

C. Energy Management System

The purpose of the EMS is to achieve a proportional distribution of the total BESS power among the n-modules, ensuring that the BESSs operate without exceeding their charge or discharge threshold values, denoted as SOC_{low} and SOC_{high}, respectively. Assuming that the total PV power is the sum of the PV power of each module ($P_{PV,tot} = \sum P_{PV,n}$), the system will operate in charging mode if the total PV power is higher than the reference active power demanded by the grid, $P_{PV,tot} > P^*_{grid}$, and in discharging mode, if $P_{PV,tot} < P^*_{grid}$. If $P_{PV,tot} = P^*_{grid}$, the n-module will operate neither charging nor discharging mode.

In the discharge state, a proportional relationship is established between $P_{bat,n}$ and BESS SOC (SOC_n), and thus, i.e., $SOC_1 > SOC_2 > SOC_3$ would mean that $P_{bat,1-ch} > P_{bat,2-ch} > P_{bat,3-ch}$, according to:

$$P_{bat,1-ch} = \frac{SOC_1}{SOC_2} P_{bat,2-ch} \tag{11}$$

$$P_{bat,1-ch} = \frac{SOC_1}{SOC_3} P_{bat,3-ch} \tag{12}$$

On the other hand, in a charge state, the proportional relationship is carried out by $P_{bat,n}$ and BESS Depth-of-Discharge ($DOD_n = 1 - SOC_n$), that is, what remains to be loaded up to the maximum SOC. $DOD_3 > DOD_2 > DOD_1$ would mean that $P_{bat,3-dis} > P_{bat,2-dis} > P_{bat,1-dis}$, according to:

$$P_{bat,n-dis} = \frac{P_{bat,tot}\, DOD_n}{\sum DOD_n} \tag{13}$$

In order to ensure the lifetime of the BESSs, in addition to establishing safety limits for charging and discharging states, the maximum power that can be absorbed or injected from the grid is limited according to its nominal power. On this premise, the maximum power that the BESS will be able to absorb or inject will therefore be the sum of the nominal power of each battery, $P_{bat,max} = \sum P_{bat,nom}$. The different modes of operation are described as follow:

- Case 1: The n-modules operate in a safe mode, that is, $SOC_{low} < SOC_n < SOC_{high}$. In this case, the distribution of power in the charge or discharge mode occurs as explained above. When a BESS reaches its nominal power, it will not be able to absorb or inject any more power, and thus, the remaining power will be distributed among the remaining n-BESSs according to their SOC or DOD in each situation.
- Case 2: When $SOC_n \geq SOC_{high}$ and the system operates in a charge state, that BESS module can no longer be charged, and thus, its power is set to zero and the total BESS power is distributed among the rest of the n-modules according to their DOD.
- Case 3: When $SOC_n \leq SOC_{low}$ and the system operates in a discharge state, that BESS module can no longer be discharged, and its power is set to zero and the needed power is provided by the rest of the n-modules according to their SOC.

IV. RESULTS AND DISCUSSION

In this section, simulation results obtained from MatLab-Simulink are presented and discussed to verify the control scheme and the EMS proposed. The system is composed of three 4.8 kW independent PV power plant, with a layout of 6 modules in parallel and 2 in series for each PV plant. The main parameters of the impedance networks are: $L_1 = L_2 = 0.56mH$, $R_{L1} = R_{L2} = 0.05\,\Omega$, $C_1 = C_2 = 11mF$. The BESS has a rated capacity of 43.63 Ah and rated voltage of 27.5 V. The carrier frequency selected for the inverter technique modulation is $f_c = 3.5\,kHz$. The initial conditions of the PV power plants are: PV power plant 1 (PV1): 800 W/m^2; PV power plant 2 (PV2): 900 W/m^2; and PV power plant 3 (PV3): 700

W/m^2. These operating conditions are changed to 600 W/m^2 at 1 s for all the PV power plants. The temperature is kept at 25ºC. The grid active power reference (P^*_{grid}) is set to 8.16 kW during the first 9 seconds (from 0 to 9 s), and it changes to 9.12 kW from 9 to 12 s. The grid reactive power reference (Q^*_{grid}) is set to 0 kVAr (unity power factor). The initial SOCs are set to: $SOC_1 = 90.01\%$, $SOC_2 = 50\%$ and $SOC_3 = 25\%$. The BESS safe conditions are selected as follows: $SOC_{low} = 15\%$, $SOC_{high} = 90\%$, $P_{bat,nom} = 1.2\ kW$.

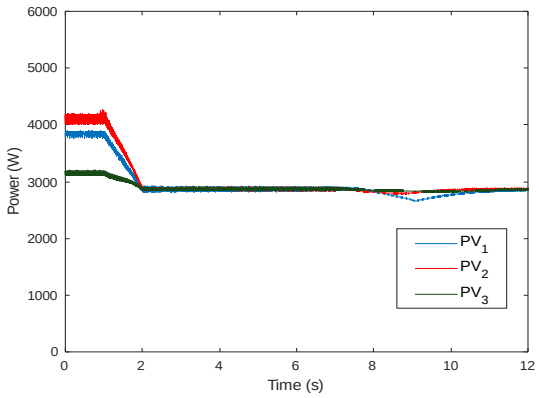


Fig. 2. Power generated by PV1, PV2, and PV3.

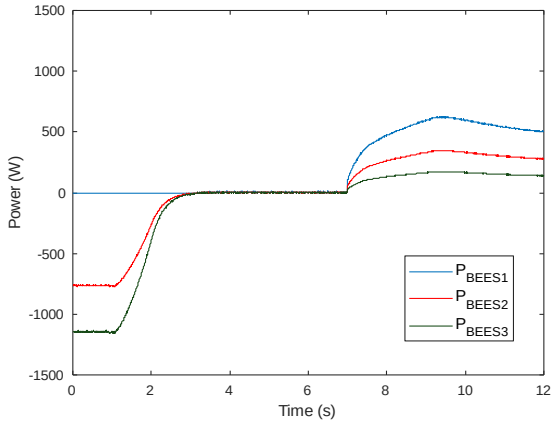


Fig. 3. BESS power: BESS1, BESS2, and BESS3.

Fig. 2 presents the power generated by each PV power plant during the simulation, showing the PV generation for different irradiances (from 0 to 2s) and for the same irradiance (from 2 to 12). Fig. 3 illustrates the power of each BESS in charge, discharge or neither charge/discharge state. Fig. 4 depicts the active and reactive power delivered to the grid (P_{grid} and Q_{grid}) and the sum of the DC power of each PV power plant at the input of the inverter and exchanged with the grid (P_{ref}) according to the system operator references. From 0 to 1 s, the system is working in charging mode because $P_{PV,tot} > P^*_{grid}$. As can be expected, $P_{bat,1-ch} = 0\ W$ because $SOC_1 \geq SOC_{high}$, and the total charging power is distributed among BESS 2 and BESS 3 according to their DOD, without exceeding the value of its nominal power, as shown in Fig. 3.

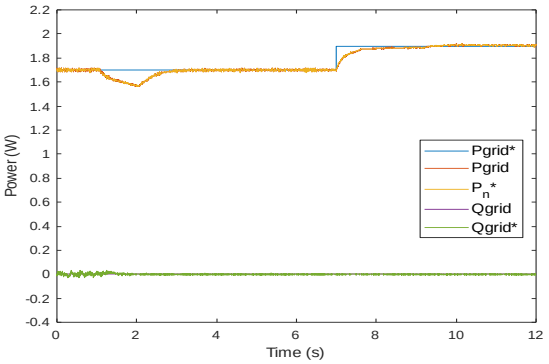


Fig. 4. Active and reactive power delivered to the grid.

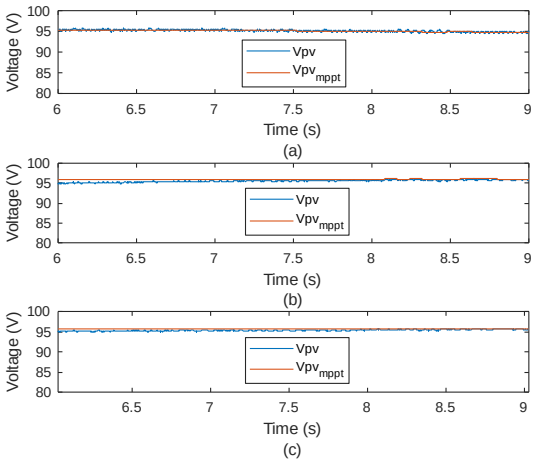


Fig. 5. MPPT (MPP voltage): (a) PV1, (b) PV2, and (c) PV3.

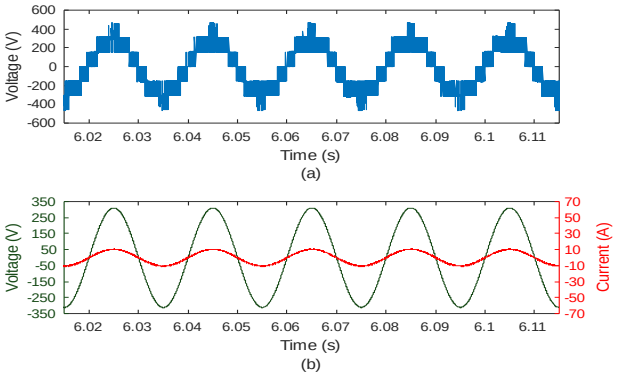


Fig. 6. (a) Seven-level output voltage of the ES-qZS-CHBMLI. (b) Grid voltage and current.

From 3 to 9 s, the PV power obviously decreases because the irradiation has gone down and each module provides 2.866 kW. The system continues working in charging mode, and thus BESS 1 still does not absorb any power ($SOC_1 \geq SOC_{high}$), and the surplus power is shared between BESSs 2 and 3. The system operates in discharging mode from 9 to 12 s because P^*_{grid} is set to 9.12 kW, and therefore $P_{PV,tot} < P^*_{grid}$. In this situation, the BESSs provide the required power, no SOCs are below its threshold value, and thus, all BESSs operate in a safe mode and the power contribution is made proportionally according to their SOC, in this case, $P_{bat,1-ch} > P_{bat,2-ch} > P_{bat,3-ch}$, as

illustrated in Fig. 4. Fig. 5 shows how the MPPT subsystem is able to ensure that each PV module operates reaching its MPP voltage (V_{mpp}), reaching at each moment the voltage that guarantees the maximum power output for each irradiation situation. Fig. 6a illustrates the seven-level output voltage of the ES-qZS-CHBMLI with 3 cascade qZSI. The grid voltage and the grid current are depicted in Fig. 6b, where it can be seen that the voltage and the current are in phase, according to unity power factor.

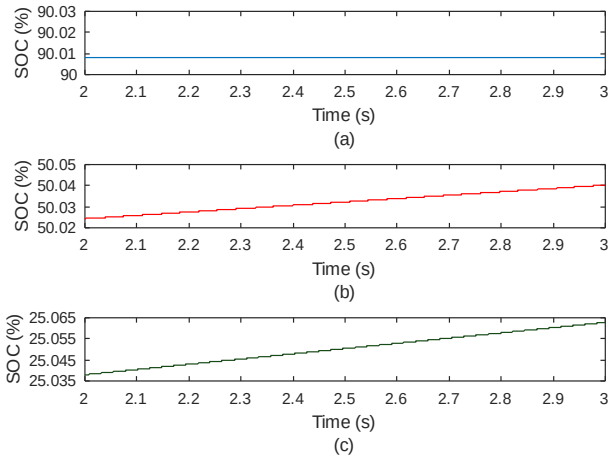


Fig. 7. BESS SOCs in charge state: (a) BESS1, (b) BESS2, and (c) BESS3.

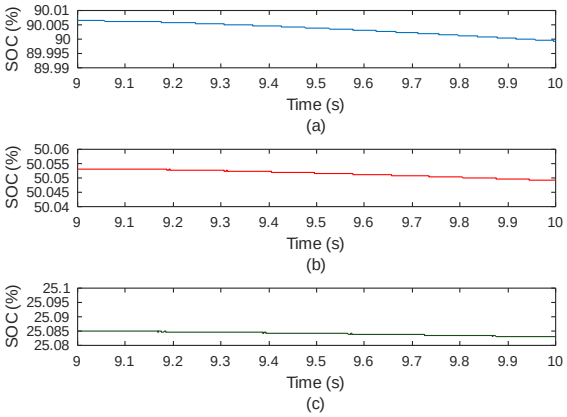


Fig. 8. BESS SOCs in discharge state: (a) BESS1, (b) BESS2, and (c) BESS3.

Fig. 7a illustrates that SOC1 does not exceed the threshold level in charging mode (90%). Figs 7b and 7c shows that the charging dynamic for BESS3 is higher than for BESS2, due to a lower SOC. In analogy, the discharging dynamic for BESS1, Fig 8a, is higher than for BESS2 and BESS3, as is depicted in Figs 8b and 8c, due to a higher SOC.

V. CONCLUSION

This paper presented a new EMS for a grid-connected ES-qZS-CHBMLI with three modules in series and PV power generation. The EMS made it possible to balance the power injected to the grid when the PV power generation fluctuated. The BESS were capable to equilibrate the excess or insufficient power, performing a proportional distribution of energy, according to their SOC or DOD. Setting threshold levels for charging and discharging, as well as limiting the maximum power that can flow through the BESS, could help to extend the lifetime of the energy storage systems. Likewise, the control scheme allowed the monitoring of the power injected to the grid based on the references ordered. The simulation results showed that the proposed control system effectively controlled the power injected to the grid under different irradiation conditions and reference setpoints while the EMS equilibrated BESS power.

ACKNOWLEDGMENT

This work was partially supported by the Regional Ministry of Economic Transformation, Industry, Knowledge and Universities of Junta de Andalucía under Grant PY20_00317 and Spain's Ministerio de Ciencia, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER) Unión Europea (UE) (grant number RTI2018-095720-B-C32).

REFERENCES

[1] International Renewable Energy Agency, "Global energy transformation: A roadmap to 2050," 2018.

[2] A. Cabrera-Tobar, E. Bullich-Massagué, M. Aragüés-Peñalba, and O. Gomis-Bellmunt, "Topologies for large scale photovoltaic power plants," *Renew. Sustain. Energy Rev.*, vol. 59, pp. 309–319, 2016.

[3] L. de Oliveira-Assis *et al.*, "Simplified model of battery energy-stored quasi-Z-source inverter-based photovoltaic power plant with twofold energy management system," *Energy*, vol. 244, 2022.

[4] B. Ge *et al.*, "An energy-stored quasi-Z-source inverter for application to photovoltaic power system," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4468–4481, 2013.

[5] O. Alonso, P. Sanchis, E. Gubía, and L. Marroyo, "Cascaded H-bridge multilevel converter for grid connected photovoltaic generators with independent maximum power point tracking of each solar array," *PESC Rec. - IEEE Annu. Power Electron. Spec. Conf.*, vol. 2, pp. 731–735, 2003.

[6] D. Sun, B. Ge, F. Z. Peng, A. R. Haitham, D. Bi, and Y. Liu, "A new grid-connected PV system based on cascaded H-bridge quasi-Z source inverter," *IEEE Int. Symp. Ind. Electron.*, pp. 951–956, 2012.

[7] S. A. Khajehoddin, A. Bakhshai, and P. Jain, "The application of the cascaded multilevel converters in grid connected photovoltaic systems," *2007 IEEE Canada Electr. Power Conf. EPC 2007*, pp. 296–301, 2007.

[8] Y. Xue, B. Ge, and F. Z. Peng, "Reliability, efficiency, and cost comparisons of MW-scale photovoltaic inverters," *2012 IEEE Energy Convers. Congr. Expo. ECCE 2012*, pp. 1627–1634, 2012.

[9] Y. Liu, B. Ge, H. Abu-Rub, and F. Blaabjerg, "Single-phase Z-source/quasi-Z-source inverters and c onverters," *IEEE Ind. Electron. Mag.*, vol. 12, no. 2, pp. 6–23, 2018.

[10] W. Liang, Y. Liu, and J. Peng, "A day and night operational quasi-Z source multilevel grid-tied PV power system to achieve active and reactive power control," *IEEE Trans. Power Electron.*, vol. 36, no. 1, pp. 474–492, 2021.

[11] Y. Li, J. Anderson, F. Z. Peng, and L. Dichen, "Quasi-ZSI for photovoltaic power generation systems," in *Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC*, 2009, pp. 918–924.

[12] E. Kabalcı, *Multilevel inverters: Introduction and emergent topologies*. 2021.

[13] F. Z. Peng, "Z-source inverter," *IEEE Trans. Ind. Appl.*, vol. 39, no. 2, pp. 504–510, 2003.

[14] R. José *et al.*, "Multilevel converters: An enabling technology for high-power applications," *Proc. IEEE*, vol. 97, no. 11, pp. 1786–1817, 2009

Automated Design of Neural Networks for FPGAs using Approximated Computing

Anatoliy Doroshenko
Dept. of Information Systems and Technologies
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine
doroshenkoanatoliy2@gmail.com

Volodymyr Shymkovych
Dept. of Information Systems and Technologies
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine
v.shymkovych@kpi.ua

Tural Mamedov
Dept. of Computing Theory
Institute of Software Systems of National Academy of Sciences of Ukraine
Kyiv, Ukraine
tural.mamedov1@gmail.com

Olena Yatsenko
Dept. of Computing Theory
Institute of Software Systems of National Academy of Sciences of Ukraine
Kyiv, Ukraine
oayat@ukr.net

***Abstract*—The facilities for automated construction and synthesis of software for programmable logic integrated circuits using high-level schemes are proposed and applied for the design of an artificial neuron. The schemes are based on algorithmic algebra and are applied for generating source code in the VHDL language, which is further executed on an FPGA. The method for designing an artificial neuron with sigmoidal activation function on field-programmable gate arrays is developed, which differs from similar approaches in that coefficients of piecewise-linear approximation of activation function are stored in memory only for positive or only for negative values of arguments. This allowed optimizing the number of utilized computing resources and increased the performance of the neural network. The proposed approach is demonstrated for developing an application with a real-time neural controller implemented on FPGA.**

Keywords—activation function, algebra of algorithms, artificial neuron, field-programmable gate array, neural network, program synthesis

I. INTRODUCTION

The technology of developing applications for field-programmable gate arrays (FPGAs) is based on a description of an algorithm in a hardware description language, such as VHDL [1], and automatic translation of the description into a specification at the level of logic tables and other functional components of an FPGA. Elementary functions in FPGA are usually implemented as separate projects or modules, which contain information about data bit rate and the internal structure of a system.

One of the tasks in hardware implementation of neural networks is a realization of an artificial neuron and its non-linear activation functions on FPGA. Existing approaches to the implementation of non-linear functions use various approximation methods [2]–[4], such as the Taylor series, a table method, piecewise-linear approximation, etc. Taylor series require numerous multiplications, and therefore are not optimal for implementation on FPGA, since the multiplication block takes a lot of resources. We chose a piecewise-linear approximation for implementing non-linear activation functions.

Programming in the VHDL language is quite difficult, so the question arises about the development of special software automation tools that would allow to efficiently generate high-performance code for programmable logic integrated circuits. In this paper, we apply the algebra-algorithmic approach for the automated design of a specification of a hardware implementation of a neural network on an FPGA. The approach uses the system of algorithmic algebra [5], under which programs are designed in the form of high-level specifications in a natural-linguistic algorithmic language. The approach is applied for developing a neuro controller implemented on FPGA for a ball-on-platform system.

In particular, this paper is related to works on the automated generation of VHDL programs [6]–[9]. In paper [6], a library for automated VHDL code manipulation and synthesis is proposed. In [7], a methodology and a tool for generating programs for FPGAs on the basis of Xilinx System Generator specifications are described. Paper [8] proposes an approach for generating program code corresponding to Moore finite state machine based on a data flow graph. Paper [9] presents a generator of VHDL programs based on P4, a domain-specific language for network devices. The main difference of our approach consists in using a natural linguistic representation of schemes in a system of algorithmic algebra for the automated design of programs for FPGAs.

The implementations of artificial neurons similar to our approach are considered in works [10], [11]. Paper [10] presents a hardwired realization of a multi-input neuron with a non-linear activation function using FPGA and VHDL for describing the system architecture. The difference of our method is that the coefficients of piecewise-linear approximation of activation function are stored in memory only for positive or only for negative values of arguments, which allows optimizing the number of used computing resources and increasing the performance of a neural network. Paper [11] proposes the implementation of an artificial neural network on FPGA using

Verilog language. The implementation uses a linear activation function unlike in our work, is slower, and uses more FPGA resources.

The controllers based on neural networks for the ball-on-plate problem are considered in [12–14]. Paper [12] proposes a feedback controller to compensate for errors resulting from the use of an approximate dynamic model in the design of the controller. In [13] and [14], a particle swarm optimization method is applied for training neural networks used in the controller.

II. ALGEBRA-ALGORITHMIC FACILITIES FOR PROGRAM DESIGN

We use natural-linguistic schemes represented in Glushkov's system of algorithmic algebra (SAA) [5] to design programs for programmable logic integrated circuits. The main objects of the SAA language are abstractions of predicates (conditions) and operators, which are divided into basic and compound. Basic predicates and operators are considered primary, atomic, indivisible constructs in SAA schemes. Compound predicates and operators are built from basic ones using logic and operator operations of SAA, in particular:

- disjunction: '*condition* 1' or '*condition* 2';
- conjunction: '*condition* 1' and '*condition* 2';
- negation: not('*condition*');
- sequential execution: "*operator* 1"; "*operator* 2";
- branching: IF '*condition*' THEN "*operator* 1" ELSE "*operator* 2";
- loop: WHILE 'condition' "*operator*" END OF LOOP.

Identifiers of conditions are delimited by single quotes, and operators are surrounded by double ones. A superposition of operations and basic elements of SAA is called an SAA scheme. The main difference between SAA and other procedural programming languages is the ability to design programs in both algebraic and natural-linguistic forms, as well as perform formal program transformations. SAA constructs for designing programs for FPGAs are considered in detail in [15].

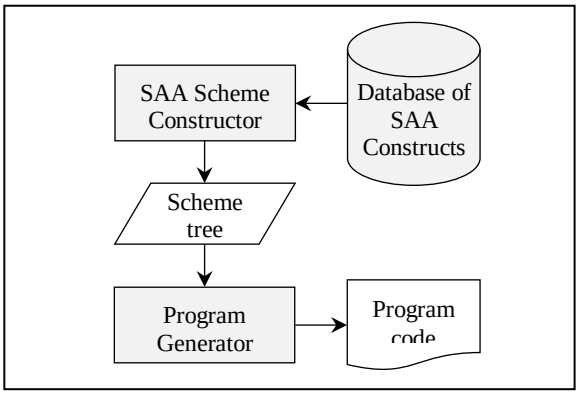


Fig. 1. The process of automated software design in the IDS toolkit.

The automated construction of SAA schemes and generation of corresponding sequential and parallel programs in target programming languages is provided by the Integrated toolkit for the Design and Synthesis of programs (IDS) [5,15]. The process of program development is shown in Fig. 1.

The main idea consists in top-down design of schemes by selecting SAA constructs from a list and adding them to a scheme tree. The descriptions of the constructs are stored in the toolkit database. Based on the designed tree, the toolkit generates program code in one of the target programming languages (C++, C#, Java, VHDL).

The following SAA scheme represents an example of a logic gate design for a Boolean expression $y = (\overline{a \wedge b}) \wedge (\overline{a \vee b}) \wedge c$ (Fig. 2). The scheme is the basis for the automatic generation of VHDL code.

```
SCHEME BOOL_EXPR

  ENTITY bool_expr IS
    PORT (
      "Input bit signals (a, b, c)";
      "Output bit signal (y)");
  END OF ENTITY;

  ARCHITECTURE arch1 of bool_expr IS
    (y <= (not(a and b)) and (not(a or b)) and c);
  END OF ARCHITECTURE

END OF SCHEME BOOL_EXPR
```

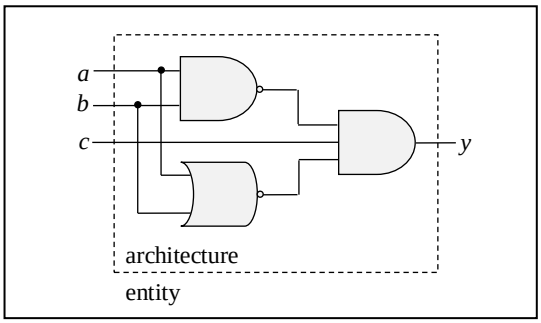


Fig. 2. A logic gate for the expression $y = (\overline{a \wedge b}) \wedge (\overline{a \vee b}) \wedge c$.

In this work, the IDS toolkit is applied for designing a scheme of a hardware implementation of a neuron on FPGA.

III. DESIGNING AN ARTIFICIAL NEURON IMPLEMENTED ON AN FPGA

In this section, we consider the model of a neuron and the FPGA design of its activation functions. The neuron is designed using high-level SAA schemes with further generation of VHDL code.

The model of a neuron works in the following way. Input signals a_{ki} enter the blocks implementing the function of synapses, each of which is characterized by its weighting coefficient w_{ki} (synapse weight). Weighted input signals are submitted to a linear summator, after which the result of summation enters the block of activation function $f(.)$, and after

respective processing is submitted to the output as a signal q_k. Generally, the activation function limits the output signal of the neuron in the range $[0,1]$ or $[-1,1]$. The model of the neuron also contains an initial shift b_k, which is added to the input signal of the activation function block. The functional scheme of an artificial neuron model of continuous type is given in Fig. 3.

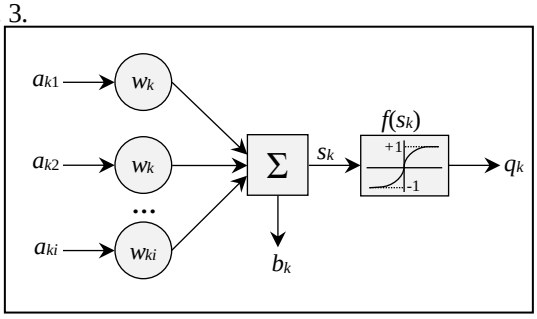


Fig. 3. The scheme of an artificial neuron.

Mathematically, the model of a neuron is described by the following dependencies:

$$q_k = f\left(S_k\right) = f\left(\sum_{i=1}^{n} w_{ki}a_{ki} + b_k\right), \qquad (1)$$

where q_k is the output signal of the *k*-th neuron; $f(.)$ is the activation function; a_{ki} are input signals of the neuron; w_{ki} is a synapse weight; b_k is a shift of the *k*-th neuron.

The activation function $f(.)$ implements non-linear transformation, carried out by the neuron. The most commonly used type of activation function is a sigmoidal function. Such functions are monotonically increasing, continuous and differential. The differential nature of sigmoidal functions is an important property of some methods of training and analysis. They also have universal approximation features [2], [4]. The peculiarity of neurons with such activation function is that they amplify strong signals much less than weak ones, since the areas of strong signals correspond to flat sections of the characteristics. This allows preventing the saturation from strong signals.

Sigmoidal functions are a class of functions described by the expression

$$f(x,k,b,T,c) = k + \frac{c}{1+be^{Tk}}, \qquad (2)$$

where x, k, b, T, c are parameters; $k,b \in R$; $b>0$; $T,\ c \in R\setminus\{0\}$.

If $k=0$, $c=1$, $b=1$, and $T=-1$, then expression (1) will be the following, which is called a "classic" sigmoid:

$$f(x,0,1,-1,1) = 0 + \frac{1}{1+1e^{-1x}} = \frac{1}{1+e^{-x}}. \qquad (3)$$

The active domain of the neuron activation function is the range of values of input parameters, where the values of the function change substantially. For the sigmoidal function, the interval [–4; 4] is used as the active domain. In this case, the function takes values in the interval (0.018; 0.982), which is 96.4% of the whole range of values.

We propose the following method of designing activation functions of a neuron on an FPGA. The input data of the method are functions described by (1) or the ones which can be written using them, and also the accuracy, with which the activation function has to be implemented. In the first stage, the activation function is examined concerning the symmetry regarding the axes. Considering the function (2):

$$f(-x) = 1 - f(x) = 1 - \frac{1}{1+e^{-x}} = 1 - \frac{1}{1+\frac{1}{e^x}} = \qquad (4)$$
$$1 - \frac{e^x}{e^x+1} = \frac{e^x+1}{e^x+1} - \frac{e^x}{e^x+1} = \frac{1}{e^x+1},$$

we obtain

$$1 - \frac{1}{1+e^{-x}} = \frac{1}{1+e^x}, \text{ or } 1 - f(x) = f(-x). \qquad (5)$$

The function $f(x)$ can be considered only for positive arguments. For its negative arguments, the value can be found by formula (3), which in turn will speed up the calculation of the function and will reduce the utilized resource of an FPGA.

In the second stage, the piecewise-linear function is defined on each of the intervals $(-\infty;x_1)$, $(x_1;x_2)$, ... $(x_n;+\infty)$ by a separate formula:

$$f\left(x\right) = \begin{cases} k_0x + b_0, & x < x_1, \\ k_1x + b_1, & x_1 < x < x_2, \\ \cdots \\ k_nx + b_n, & x_n < x. \end{cases} \qquad (6)$$

In the third stage, $f(x)$ is computed for the previously calculated value of x. The coefficients k and b are picked from memory.

Thus, based on linear formulas, we can find the approximation of the function of a sigmoidal type for any argument with a given accuracy. The developed method of designing non-linear activation functions of an artificial neuron on field-programmable gate arrays differs from [10] in the fact that coefficients of piecewise-linear approximation of activation function are stored in memory only for positive or only for negative values of arguments, which allows optimizing the number of used computing resource and increase the performance of a neural network.

The algorithm of implementation of the artificial neuron with classic sigmoidal activation function is the following.

Step 1. Synapse weights of the neuron are set. Every neuron is given a block of memory, where synaptic weights are stored.

To set the weights, the following signals are used:

- *synaddr* (synapse address) — selecting the synapse, the weight of which will be read or written by its number in a binary code;
- *synsetw* (synapse set weight) — the value of weight to be written to synapse;
- *synwren* (synapse write enabled) — when the input of this signal is 1, the neuron writes the value from *synsetw* into the selected synapse, and when the signal is 0, nothing happens.

These variables are necessary for training the neuron and neural networks.

Step 2. The values of the input vector are submitted to the inputs of the artificial neuron. The variable x is set, which is equal to the output of the summator:

$$x = \sum_{i=1}^{N} w_i \cdot a_i, \qquad (7)$$

where a_i are neuron inputs; w_i are synapse weights of the neuron.

The calculation is done using fixed-point numbers. Every number takes 16 bits (9 bits for the integer part and 7 bits for the fractional one).

Step 3. The module of the sigmoidal function argument is calculated. The variable x' is set: $x' = |x|$.

Step 4. The sigmoidal activation function is partitioned into linear pieces, and the coefficients k and b of linear equations are defined. For this, the basic model of piecewise-linear approximation with a given number of linear segments was used [16]. The partitioning and error are shown in Fig. 4. The coefficients of the linear equations are defined as follows:

$$k, b = \begin{cases} [0.234, 0.500], \text{ if } 0 \le x' < 1; \\ [0.129, 0.605], \text{ if } 1 \le x' < 2.5; \\ [0.234, 0.500], \text{ if } 2.5 \le x' < 4; \\ [0.009, 0.946], \text{ if } x' \ge 4. \end{cases} \qquad (8)$$

Step 5. Variable f is defined according to the formula:

$$f = k \cdot x' + b. \qquad (9)$$

Step 6. If $x < 0$, the values of the local variable are calculated by the formula $f = 1 - f$.

Step 7. The value of the output signal of the neuron is set equal to the value of the variable f.

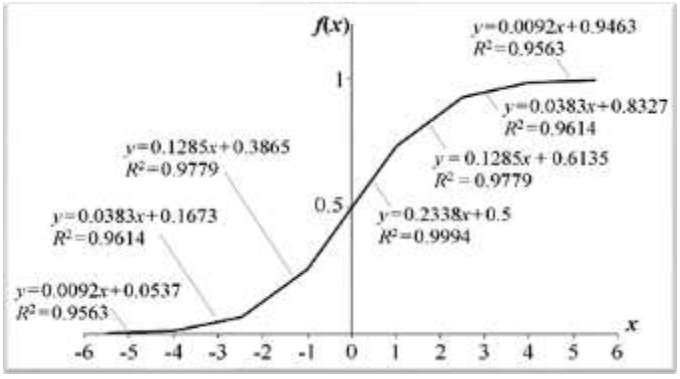


Fig. 4. Piecewise-linear approximation of the sigmoidal function.

Fig. 5 shows the classic sigmoidal function by Eq. 2 and the function implemented on FPGA based on the developed algorithm. The implemented block of the sigmoidal function on FPGA reflects the function with sufficient accuracy for further implementation of artificial neural networks.

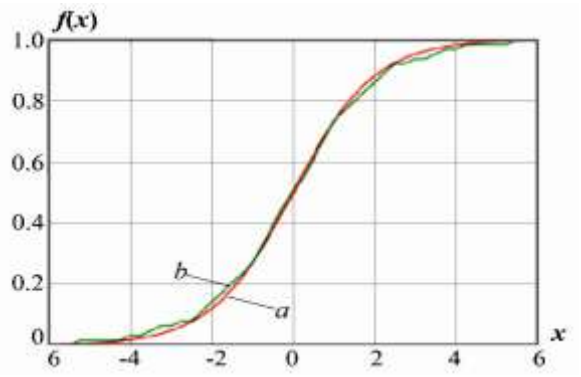


Fig. 5. The graphs of the classic sigmoidal function (*a*) and the function implemented on FPGA (*b*).

The block of the artificial neuron with four inputs is shown in Fig. 6.

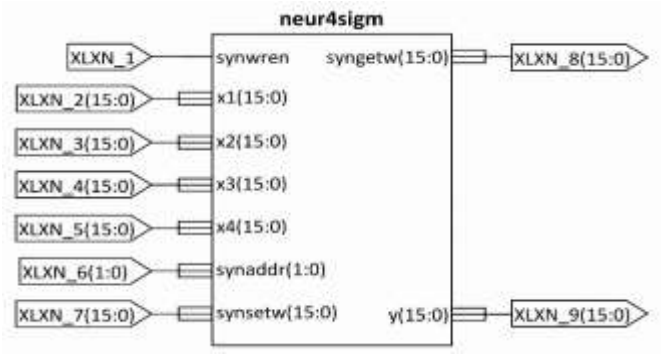


Fig. 6. The block of an artificial neuron implemented on FPGA.

It was designed in an automated way in the form of an SAA scheme using the developed IDS toolkit described in Section II. The specification of an entity in SAA is given below as an example. SAA scheme was then used for automatic generation of VHDL code.

```
ENTITY neur4sigm IS
  PORT (
    "Input logic vector signals (x1, x2, x3, x4) of range
     (15) down to (0)";
    "Input logic vector signal (synaddr) of range
     (1) down to (0)";
    "Input logic signal (synwren)";
    "Input logic vector signal (synsetw) of range
     (15) down to (0)";
    "Output logic vector signal (syngetw) of range
     (15) down to (0)";
    "Output logic vector signal (y) of range
     (15) down to (0)" )
END OF ENTITY
```

IV. EXPERIMENTAL RESULTS AND PRACTICAL APPLICATION OF THE APPROACH

The FPGA implementation of the artificial neuron with four inputs and a sigmoidal activation function using 16-bit fixed-point numbers took 672 LUTs (Look Up Tables). The performance (the total combinatorial scheme delay) of the neuron block was 75.6 ns. The absolute error was ± 0.005, and the accuracy of the implementation of the sigmoidal function was shown earlier in Section III (Fig. 4).

Table I shows the comparison of the results of the implementation of the sigmoidal function with the most similar known analogs [10], [11] on FPGAs Xilinx Spartan 6 and Xilinx Spartan 3. As can be seen, the implementation of the neural network based on the developed method and algorithm is faster and requires fewer resources of the chip, maximum deviation was also decreased. In this work, non-linear sigmoidal functions are used in contrast to linear functions used in [10].

The proposed approach was applied for developing a neural controller for balancing a ball on a platform in real time with the hardware and software realization on FPGA (Fig. 7). The platform regulates the position of a ball to bring it to an equilibrium state by tilting along horizontal axes (*x* and *y*) using two servo motors. The ball position is fixed by a video camera.

TABLE II. THE COMPARISON OF RESULTS OF THE IMPLEMENTATION OF THE NEURON'S SIGMOIDAL ACTIVATION FUNCTION ON VARIOUS FPGAS.

Work	paper [10]	this work	paper [11]	this work
FPGA series	Xilinx Spartan 6		Xilinx Spartan 3	
FPGA resource (LUT)	108	60	336	75
Performance, ns	22.12	17.5	120.1	30.2
Maximum deviation	0.50%	0.45%	0.556%	0.45%

A simplified representation of a physical model of the movement of the ball on the platform which actuators is shown in Fig. 8, where L is the distance from the middle to the edge of the platform (0.35 m); r is the distance from the center of the ball to the end of the platform; α is the platform rotation angle; d is the servo motor extension length (0.05 m); θ is the servo motor rotation angle.

Fig. 7. The stand for balancing the ball on the platform.

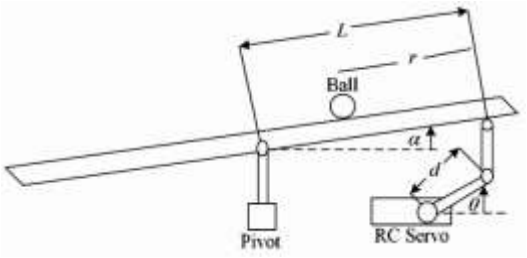


Fig. 8. Physical simulation of the ball movement on the plate.

For controlling the motion of a ball according to a tilt angle, a neural controlling system with an inverse model of a controlled object and feedback was developed (Fig. 9). A three-layer artificial neural network with two neurons in the input layer, eight neurons in the hidden layer, and two neurons in the output layer was chosen since this topology replicates the motion of the ball with the smallest error. Initial training of the neural network was performed on data obtained in Ref. [17] at modeling a PID controller.

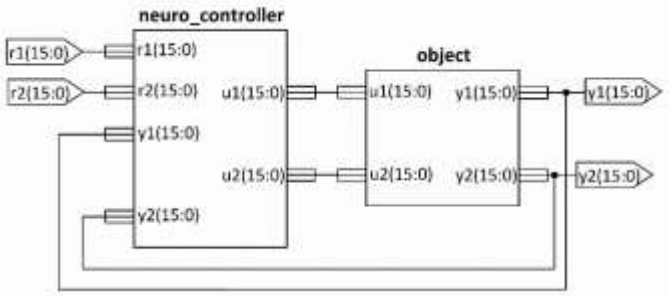


Fig. 9. The scheme of the neural controller implemented on FPGA.

The results of the simulation obtained on a platform with the neural network controller and traditional PID controllers are presented in Fig. 10 and Fig. 11. As the graphs show, the neuro controller adapts and eliminates all assumptions and uncertainties in modeling and calculations, as well as to changes in its operating conditions, for example, when changing the parameters of the ball, to the introduction of disturbing influences into the system. As a result, the ball is set to a given point faster and with less deviation.

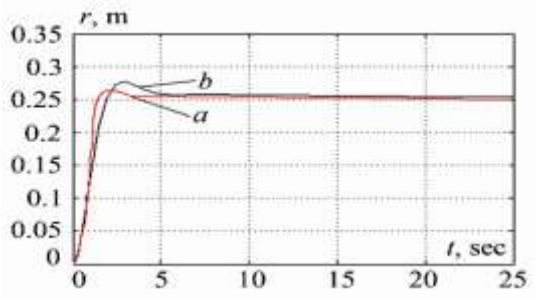


Fig. 10. Graphs of the dependence of the ball position r from time: (*a*) on a platform with neuro controller, (*b*) on a platform with regular PID controllers [17].

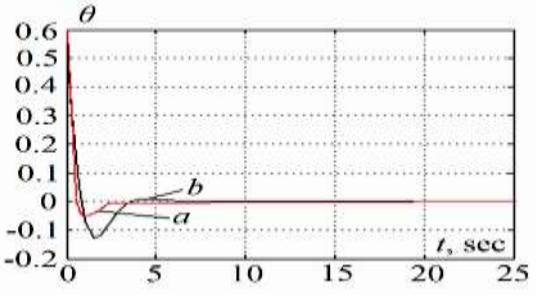


Fig. 11. Graphs of the dependence of the angle of rotation θ of the servo motor from time: (*a*) on a platform with neuro controller, (*b*) on a platform with regular PID controllers [17].

V. CONCLUSION

The tools for automated design and generation of programs for FPGAs based on the algebra-algorithmic schemes have been developed and applied for the automated design of an artificial neuron. The schemes are used for generating source code in the VHDL language, which is further executed on an FPGA. The method for designing an artificial neuron with sigmoidal activation function on FPGAs was developed, which differs from similar approaches in that coefficients of piecewise-linear approximation of activation function are stored in memory only for positive or only for negative values of arguments. This allowed optimizing the number of used computing resources and increasing the performance of the neural network. The developed approach was applied for developing a system with a neural controller for balancing a ball on a platform implemented on FPGA.

REFERENCES

[1] A. P. Godse and D. A. Godse, VHDL Programming: Concepts, Modeling Styles and Programming. Seattle, USA: Amazon Digital Services LLC, 2020.

[2] A. R. Barron, "Universal approximation bounds for superposition of a sigmoidal function," IEEE Trans. Inform. Theory, vol. 39, pp. 930–945, 1993.

[3] J.-Y. Jhang, K.-H. Tang, C.-K. Huang, C.-J. Lin, and K.-Y. Young, "FPGA implementation of a functional neuro-fuzzy network for nonlinear system control," Electronics, vol. 7, no. 145, pp. 1–22, Aug. 2018.

[4] D. Costarelli and R. Spigler, "Approximation results for neural network operators activated by sigmoidal functions," Neural Networks, vol. 44, pp. 101–106, Aug. 2013.

[5] A. Doroshenko and O. Yatsenko, Formal and Adaptive Methods for Automation of Parallel Programs Construction: Emerging Research and Opportunities. Hershey, PA: IGI Global, 2021.

[6] C. Pohl, C. Paiz, and M. Porrmann, "vMAGIC – automatic code generation for VHDL," International Journal of Reconfigurable Computing, vol. 2009, pp. 1–9, Jul. 2009.

[7] P. Martín, E. Bueno, Fco. J. Rodríguez, O. Machado, and B. Vuksanovic, "An FPGA-based approach to the automatic generation of VHDL code for industrial control systems applications: a case study of MSOGIs implementation," Mathematics and Computers in Simulation, vol. 91, pp. 178–192, May 2013.

[8] D. R. F. de Bulnes and Y. Maldonado, "VHDL code generation as state machine from a data flow graph," In Proc. 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC 2016), 2016, pp. 1–6.

[9] P. Benáček, V. Puš, H. Kubátová, and T. Čejka, "P4-To-VHDL: automatic generation of high-speed input and output network blocks," Microprocessors and Microsystems, vol. 56, pp. 22–33, Feb. 2018.

[10] K. Goel, U. Arun, and A. K. Sinha, "An efficient hardwired realization of embedded neural controller on System-On-Programmable-Chip (SOPC)," International Journal of Engineering Research & Technology, vol. 3, no. 1, pp. 276–284, Jan. 2014.

[11] S. Singh, S. Sanjeevi, V. Suma, and A. Talashi, "FPGA implementation of a trained neural network," IOSR Journal of Electronics and Communication Engineering, vol. 10, no. 3, ver. III, pp. 45–54, Jun. 2015.

[12] A. Mohammadi and J.-C. Ryu, "Neural network-based PID compensation for nonlinear systems: ball-on-plate example," Int. J. Dynam. Control, vol. 8, pp. 178–188, 2020.

[13] M. Shaheer et al., "Control of a ball-bot using a PSO trained neural network," In Proc. 2nd International Conference on Control, Automation and Robotics (ICCAR 2016), 2016, pp. 24–28.

[14] K. Han, Y. Tian, Y. Kong, J. Li, and Y. Zhang, "Tracking control of ball and plate system using an improved PSO on-line training PID neural network," In Proc. 2012 IEEE International Conference on Mechatronics and Automation, 2012, pp. 2297–2302.

[15] A. Doroshenko, V. Shymkovych, O. Yatsenko, and T. Mamedov, "Automated software design for FPGAs on an example of developing a genetic algorithm," In Proc. 17th International Conference "ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer" (ICTERI 2021), 2021, pp. 74–85.

[16] E. Camponogara and L. F. Nazari, "Models and algorithms for optimal piecewise-linear function approximation," Mathematical Problems in Engineering, vol. 2015, pp. 1–9, Jul. 2015.

[17] V. Shymkovych, V. Samotyy, S. Telenyk, P. Kravets, and T. Posvistak, "A real time control system for balancing a ball on a platform with FPGA parallel implementation," Technical Transactions, vol. 5, pp. 109–117, 2018.

Electric Bus Battery Degradation Simulation

Paula Zenni Lodetti
Sustainable Energy Center
CERTI Foundation
Florianópolis – SC, Brazil
pzl@certi.org.br

Jessica Ceolin de Bona
Sustainable Energy Center
CERTI Foundation
Florianópolis – SC, Brazil
jdb@certi.org.br

Flavio Junior de Faveri
Sustainable Energy Center
CERTI Foundation
Florianópolis – SC, Brazil
fjf@certi.org.br

Marcos Aurelio Izumida Martins
Sustainable Energy Center
CERTI Foundation
Florianópolis – SC, Brazil
mlz@certi.org.br

Rodolfo Sabino de Moura
Electrical Mobility Engineering
EDP Brazil
São Paulo – SP, Brazil
rodolfo.sm@edp.com

Jorge Gustavo Schmidt
Sustainable Energy Center
CERTI Foundation
Florianópolis – SC, Brazil
jorgesch07@gmail.com

***Abstract*—Electric mobility is a hot topic within academia, industry and politics. In particular, electric buses have shown numerous advantages for transporting people without increasing the level of greenhouse gases and their effects on global warming. However, a point of attention is the battery degradation of these vehicles over time. Thus, this paper presents the results of the modeling and simulation of a system composed by an electric bus and its components, the charging station and the battery degradation of two routes, in order to estimate the electric bus battery life.**

Keywords— electric bus, battery degradation, smart mobility

I. INTRODUCTION

Passenger transport can be divided into 3 categories: urban, road and charter. The transport of passengers on a charter system consists of an intermediary modality between collective public transport and individual private transport. It is a service intended for the transport of people with a common origin or destination, such as company employees or tourists. This transport can be through cars, vans or, more commonly, buses that share the characteristics of road buses [1]. Meanwhile, electric-powered buses have showed a less polluting and more energy-efficient alternative than their diesel counterparts, with adoption in several cities around the world [2].

In this context, the Brazilian R&D project named "Development and Pilot Implementation of a Technical and Business Model of Recharge Infrastructure for Electric Bus Fleets" intends to operationalize a commercial electric bus for chartering employees of two major industries in the region of Vitoria (ES) in Brazil. The project consists of an electric bus, with a battery capacity of 324 kWh, and four AC charging stations, with 88 kVA power, operating in different business models. The pilot will allow the evaluation of heavy-duty electric vehicles for private passenger transport. The business models for fleets of electric charter buses developed in the project is presented in [3].

As with other electric vehicles, the battery is one of its most important subsystems, along with the charging system. However, studies place the lifetime of bus batteries (and electric vehicles in general) in the range of eight years, depending on the direction, vehicle capacity and even topography of the place of use [4] [5]. Comparing to the lifetime of buses (15-20 years), this indicates the use of at least two batteries during the lifetime of the vehicle. Therefore, the work introduced by this paper presents the results from an electric bus battery degradation simulation performed for two different routes taken by the chartered bus fleet analyzed by the project.

The routes used to validate our tests are both near the metropolitan region of Vitoria, capital of the state of Espírito Santo in Brazil. The first one is from a bus station called Viação Águia Branca (VAB) to the Samarco unit in Anchieta (located in south of Vitória). The second one is from another bus station called Rodoviária João Neiva (RJN) to Suzano industry, in the north of the capital.

Battery degradation affects directly the economic viability of these applications during its lifetime [6]. Given that, [6] presents an extensive experimental degradation data for lithium-ion battery cells from three different manufactures. [7] states that when used in the field of vehicle power batteries, lithium-ion batteries may be slightly overcharged due to the errors in the Battery Management System (BMS) state estimation. This can lead to problems such as battery performance degradation and battery stability degradation.

The paper is divided into five sections. Section I is the introduction. Section II presents the system modeling (electric bus, charging station, degradation). Initial tests are described in section III and the results are presented in section IV, through the calculation of the state of health. Finally, section V concludes the paper with future works.

II. SYSTEM MODELING

In the automotive applications context, the use of HIL (Hardware In the Loop) systems for validation and verification of the power train control module (PCM - Power Control Module) is very common. The PCM is a unit that coordinates several components of a vehicle, being understood as a set of control subsystems instead of a single controller [8], [9].

In an EV, the two most important subsystems are the electric motor feeder and the charging system. Therefore, both systems were modeled using the Typhoon HIL software in order to make it possible to carry out tests for the project. This section is intended to describe the modeling adopted and the knowledge obtained during its implementation.

A. Electric Bus

1) *Inverter:* The chosen model for controlling the torque and/or frequency of an electric motor was the Induction Motor with Closed Loop Control (IFOC), because it is stable throughout the motor RPM range and because it meets current testing needs. Also, EV model of Typhoon HIL [10] employs the same system as this model. The electrical model of the induction motor is based on state space representation and described by equation (1).

$$\frac{d}{dt}\begin{bmatrix}\lambda_{ds}\\ \lambda_{qs}\\ \lambda_{dr}\\ \lambda_{qr}\end{bmatrix} = \begin{bmatrix}\frac{-L_r}{L_sL_r-L_m^2}R_s & 0 & \frac{L_m}{L_sL_r-L_m^2}R_s & 0\\ 0 & \frac{-L_r}{L_sL_r-L_m^2}R_s & 0 & \frac{L_m}{L_sL_r-L_m^2}R_s\\ \frac{L_m}{L_sL_r-L_m^2}R_r & 0 & \frac{L_s}{L_sL_r-L_m^2}R_r & \omega_r\\ 0 & \frac{L_m}{L_sL_r-L_m^2}R_r & -\omega_r & \frac{L_s}{L_sL_r-L_m^2}R_r\end{bmatrix}\begin{bmatrix}\lambda_{ds}\\ \lambda_{qs}\\ \lambda_{dr}\\ \lambda_{qr}\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}v_{ds}\\ v_{qs}\end{bmatrix}$$

$$\begin{bmatrix}i_{ds}\\ i_{qs}\\ i_{dr}\\ i_{qr}\end{bmatrix} = \begin{bmatrix}\frac{L_r}{L_sL_r-L_m^2} & 0 & \frac{-L_m}{L_sL_r-L_m^2} & 0\\ 0 & \frac{L_r}{L_sL_r-L_m^2} & 0 & \frac{-L_m}{L_sL_r-L_m^2}\\ \frac{-L_m}{L_sL_r-L_m^2} & 0 & \frac{L_s}{L_sL_r-L_m^2} & 0\\ 0 & \frac{-L_m}{L_sL_r-L_m^2} & 0 & \frac{L_s}{L_sL_r-L_m^2}\end{bmatrix}\begin{bmatrix}\lambda_{ds}\\ \lambda_{qs}\\ \lambda_{dr}\\ \lambda_{qr}\end{bmatrix} \quad (1)$$

The mechanical model of the engine implemented by Typhoon HIL uses Newton's Second Law for rotation, as described in equation (2), where ω_m is the mechanical speed of the rotor, J_m is the moment of inertia of the machine, T_e the electrical torque, T_{load} the load torque (tilt, roll and air resistance) and D_f is the coefficient of friction [10] [11].

$$\frac{d\omega_m}{dt} = \frac{1}{J_m}(T_e - T_load - D_f\omega_m) \quad (2)$$

2) *Motor Rectifier and Controller:* The bus used for the project has an onboard 80 kVA charger and does not support DC fast charging. Thus, a rectifier that also works as an inverter was included in the simulation in order to simulate AC recharges but also allow vehicle to grid (V2G) operation.

To implement the rectifier, the "Battery Inverter" block from the "Microgrid" library of Typhoon HIL was used. This model takes a vector of inputs with five parameters: On, Mode, f_{ref}, V_{ref}, P_{ref} and Q_{ref}.

Internally, the component implements a two-level inverter with control through dq components and PI controllers. It also has a contactor on the AC side of the device that is kept open until synchronization is complete (5% voltage tolerance).

The system was tested with the rectifier modeled with transistors and with the following parameters: Battery voltage: 800 V; Filter inductance: 240 µH; Cut-off frequency: 700 Hz; Switching frequency: 5 kHz; Execution step (electrical model): 1 µs; Execution step (control loop): 100 µs.

Once some instability problems were resolved, the model demonstrated satisfactory behavior and was considered to be sufficiently accurate for the design simulations.

3) *Battery:* As part of the vehicle's powertrain, the battery has a significant impact on an EV's performance, and as such, it is interesting that its modeling is as accurate as possible. The chosen electrical model is shown in Figura 1. The model seeks to represent the battery nonlinearities through impedances that vary with the *SoC* and with the battery temperature [12], [13], [14], [15].

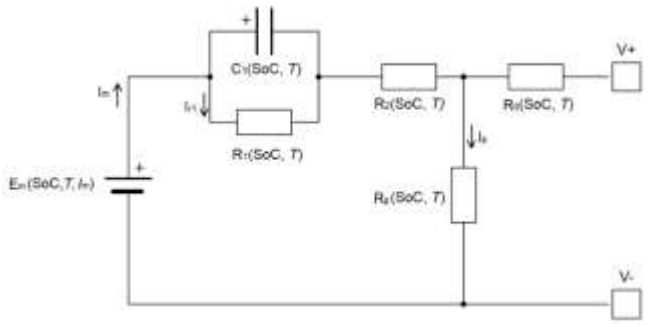


Fig. 1. Battery electric model

The RC block in Figura 1 simulates battery electrolyte saturation, series and shunt impedances simulate ohmic and self-discharge losses. The internal voltage, electromotive force or EMF (Electromotive Force), of a battery varies with the *SoC*, discharge current and temperature. The EMF curves can be described by equations (3)-(6), obtained totally or partially from [13] and [15]. More details about the equations, see works [13] and [15].

$$E_m(SoC,T,i_m) = E_{bat}(e_0 + E_m(T) + E_m(i_m) + E_m(SoC)) \quad (3)$$

$$E_m(T) = e_1(1 - SoC)\tan^{-1}(\sinh(e_2(T - T_{ke}))) \quad (4)$$

$$E_m(i_m) = -e_3\frac{i_m}{SoC} \quad (5)$$

$$E_m(SoC) = e_4 e^{-e_5(1-SoC)} \quad (6)$$

To calculate the battery temperature, the used model is described by equation (7). Where T is battery temperature, in Kelvin; t is time, in seconds; Q_{irr} is irreversible heat flow. Here, only the losses in resistors R_0, R_1, R_2 and R_p are considered, however a more complete model would consider parasitic reactions; Q_{rev} is heat flux due to reversible charge (negative) and discharge (positive) reactions; Q_{amb} is heat flux to the external environment; C_t: thermal capacitance, obtained experimentally or calculated by the product of the mass and the specific heat of the battery, $C_t = m.cp$.

$$\frac{dT}{dt} = \frac{\dot{Q}_{irr} + \dot{Q}_{rev} - \dot{Q}_{amb}}{C_T} \quad (7)$$

That is, it gives the net heat flux at each simulation step. Integrating this value over time allows the battery temperature to be estimated. This thermal model is described in [16]. A widely used indicator to measure the accumulated wear of a battery is the state of health (*SoH*), an indicator of the health status of the device. This indicator is used as a definition of EoL (End of Life), in which the battery is considered to have reached its useful life for EVs when the SoH decreases below a certain value, e.g. 80%.

The objective of the modeling is to simulate the electrical behavior as a function of *SoC* and *SoH*. Thus, the proposed model must consider that the available inputs are different from those that a battery management system Battery Management System (BMS) would use.

What is proposed is the calculation of the variation of *SoH* based on achievable cycle curves of a battery. One of these curves is illustrated by Figure 2, which shows the influence of the depth of discharge (DoD) on the number of reachable cycles. Note that the greater the discharge depth and temperature, the smaller the number of achievable cycles, that is, the shorter the battery life.

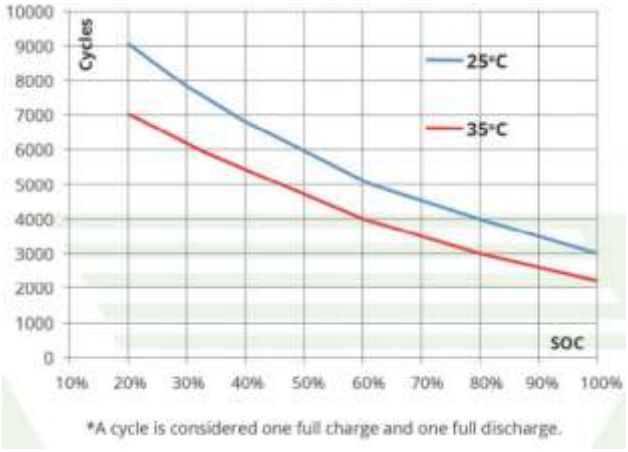


Fig. 2. Battery electric model after Ref. [17]

In Fig. 2, a cycle corresponds to an event where the battery is discharged, starting at 100% *SoC*, and recharged again to 100%. As already mentioned, the battery EoL setting is often taken as the *SoH* being reduced to 80%.

In Ref. [18] a degradation model for energy storage systems based on the curves from Figure 2 is proposed. What is proposed here is a modified version of this model so that it is possible to calculate a variation of *SoH* based on the information of *SoC*, current and battery temperature. The proposed model is synthesized by equations (8) and (9), which aggregate the variation of *SoH* caused by four aging mechanisms: depth of discharge (DoD), SoC, temperature and current. In addition, it is noted that one of the assumptions of this model is that battery degradation is the same when discharging and recharging.

$$\Delta SoH[t] = \frac{1}{2}\left|\frac{1}{NC(1-SoC[t])} - \frac{1}{NC(1-SoC[t-1])}\right| \qquad (8)$$

$$NC(x) = \frac{h_0 - h_1|T - T_{0s}|}{|i_m|^{h_2}} \cdot |x|^{-h_3} \cdot e^{-h_4|x|} \qquad (9)$$

Where ΔSoH[t] is the variation in battery health state in iteration t; NC(x) is the function that describes the number of reachable cycles. This function is obtained by setting x=DoD and fitting the parameters to graphs like the one in Figure 2; $h_{(0\text{-}4)}$: coefficients obtained experimentally or by curve fitting; T_{0s}: reference temperature, generally 298.15 K. Thus, the value of *SoH[t]* in each iteration is obtained through the sum of ΔSoH[t], according to equation (10).

$$SoH[t] = SoH_0 - 0{,}2\sum_{0}^{t}\Delta SOH \qquad (10)$$

Where SoH_0 represents the battery health status at the beginning of the simulation and the multiplicative factor 0.2 is used because the datasheet curves are considered to define EoL as 80% *SoH*.

The *SoH* value directly influences the battery SoC calculation. The state of charge is calculated by integrating the main branch current Im and taking into account the current storage capacity of the battery, a technique called Coulomb Counting, such that:

$$SoC[t] = SoC_0 - \frac{E_{bat}}{3600M_{bat}}\sum_{0}^{t}\frac{I_m}{SoH[t]}T_s \qquad (11)$$

where *SoC[t]* is the estimated state of charge at iteration t; SoC_0: initial battery charge state, between 0 and 1; M_bat: nominal battery storage capacity [Wh]; T_s: simulation step [s].

This way of counting ensures that the indicated SoC will always correspond to the current effective battery capacity. This contrasts with forms of calculation in which the SoC is given in relation to the battery's rated capacity. In the latter, the battery that had already been worn out would never reach a 100% charge. As a response to the internal resistance to aging, it was simply defined that the value of R_0 would be divided by the *SoH*:

$$R_0' = \frac{R_0}{SOH} \qquad (12)$$

With R'_0 being the value of internal resistance actually used.

B. Charging Station

The Typhoon Control Center was used for the modeling and simulation of the charging station. It has a schematic editor and a SCADA interface. The charging station was modeled from its main components: input circuit breaker, contactor, electrical magnitude meter and charging connector. The intelligent center in charge of managing the recharge was modeled in a programming block in C language.

The station's circuit breaker was modeled as a three-pole switch commanded by an algorithm structured in C-block. The algorithm takes as input the circuit breaker's on and off command, the rated current, a reset command, and the currents that circulate in each phase instantaneous and with a delay of 200 ms. From this, we verified at each iteration if the current exceeds a limit stipulated by the rated current, which characterizes an overcurrent.

The contactor was modeled as a three-pole switch, native to the Typhoon HIL schematic editor, commanded automatically from the moment the connector is activated (simulating the connection of the connector to the vehicle socket). Its command is executed by a C-block algorithm.

The electrical magnitude meter is modeled as a series of measurements of current, voltage, powers, power factor and frequency, both in RMS values and in instantaneous values. All these meters are native to the Typhoon HIL schematic editor.

The charging station connector, as well as the contactor, is also modeled with a three-pole switch commanded by the user from the simulation on the SCADA panel. This switch plays an important role in the simulation, as its on or off signal is sent, as feedback, to C-block that executes the algorithm that makes the contactor close. In this way, the vehicle charging starts.

In this version of the simulation, C-block, in addition to automating the closing of the station's contactor, also implements a logic that makes it possible to change the station's charging states.

For the simulation to be possible, it is still necessary to simulate the electrical network that feeds the station and a model of the vehicle. In this case, the electrical network is modeled as a three-phase source and a network impedance. The vehicle is modeled as a variable RL load with an initial power factor of 0.95, which can also be varied during the simulation. For that, we used variable resistors and

inductances available in the Typhoon HIL schematic editor and a C-block to operationalize the necessary calculations that result in the resistance and inductance values.

Other devices also involved in the charging system, such as RCD (Residual Current Device), circuit breaker and surge protector were also modeled.

C. Battery Degradation

The consumption control parameters of the bus simulated in HIL are: acceleration, braking and inclination, this combination defines a route being traced. For this case, average speed and slope were used as primary parameters.

The chosen route for the initial tests was the VAB-Suzano, the average speed for each section was obtained through a web platform eCalc [19], where it is possible to insert the route and vehicle parameters to obtain the average speed, consumption, *SoC* and elevation profile. The data from this route were acquired and processed to be inserted into the HIL. After, a dynamic table which uses a clock and a look-up table to acquire data through a CSV file, was used because of the infinite derivatives (raw data). The angle of inclination was calculated based on the distance traveled and the difference in elevation between the points. The acceleration was defined in steps of 0.5%, between 0 and 100%, taking the current speed and the target as a reference.

During the periods when the speed was reduced abruptly, the brake was applied in the simulation by 75%, so that the current speed returns to the reference speed level. When the vehicle just slows down steadily, the regenerative brake is engaged.

For comparison, over a 4-minute simulation there was a 1.9% error between the simulation in Typhoon HIL and eCalc. This error is reduced over time, reaching 1% in 10 minutes of simulation, this is due to the initial acceleration error at vehicle start-up.

III. INITIAL TESTS

The model tests carried out at the HIL mainly evaluated the aspects carried load, speed, energy consumption per kilometer and the power required by the engines. They directly influence the vehicle's autonomy.

The two routes were initially simulated by eCalc to acquire average speed values. These values along with braking, tilting and charging start triggers were added to the HIL model by the dynamic table.

A. Gross Transported Weight

One of the initial tests that were carried out to compare the autonomy of the vehicle was the carried load, considering two scenarios:

- 1st Scenario: 16.5 tons - base vehicle weight (15 tons) + 20 people weighing 75 kg
- 2nd Scenario: 18 tons - base vehicle weight (15 tons) + 40 people weighing 75 kg

Considering a 74 km stretch (data via HIL SCADA data logger), the *SoC* data are consolidated in Figure 3.

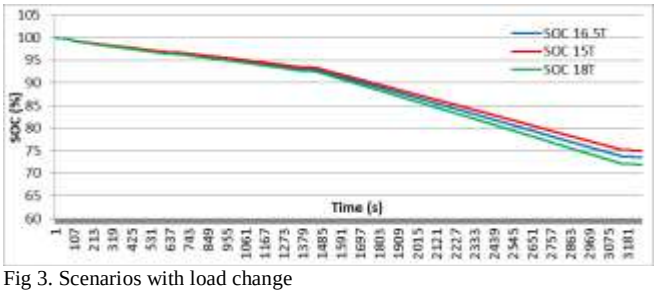


Fig 3. Scenarios with load change

Regarding the final *SoC* value for each scenario, there is a difference of 1.44% (15,000 kg to 16,500 kg) and 1.62% (16,500 kg and 18,000 kg). Respectively, the autonomy of each scenario will be 297, 281 and 265 km total.

B. Route 1 – VAB to Samarco

This route has 73 km per stretch, consumption less than a third of the autonomy expected for the electric bus. A charging station is planned for this route close to the final destination (Samarco). In this way, the bus was simulated to leave the VAB with 100% *SoC*, arrive destination and drop off the passengers. Then, proceed to the charging station and stay for 8 hours and 45 minutes, until it is time to return to Samarco and pick up the passengers to return them to VAB. The considered gross weight was 30 people weighing 75 kg. The *SoC* graph of this process is in Figure 4.

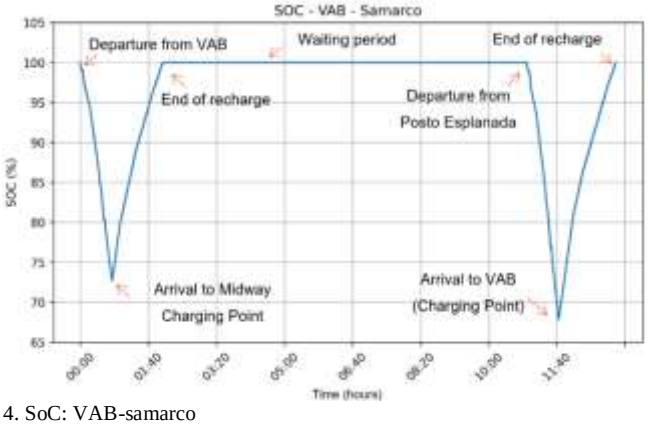


Fig. 4. SoC: VAB-samarco

In the first stretch of the trip (outbound), energy consumption was 90.06 kWh, with an average of 0.815 km/kWh and 264 km of autonomy. The second stretch had 100.86 kWh of energy consumption, 0.735 km/kWh on average and 238.14 km of autonomy. The average slope for this stretch is positive, with an additional expense due to this variable.

In terms of speed (Figure 5), in the first stretch the maximum speed was 82 km/h and the average was 66.50 km/h. In the second one, the maximum speed was 83 km/h and the average, 67.47 km/h. The reduction in vehicle autonomy between the two stretches is due to the increase in the average vehicle speed, representing a higher energy consumption per kilometer.

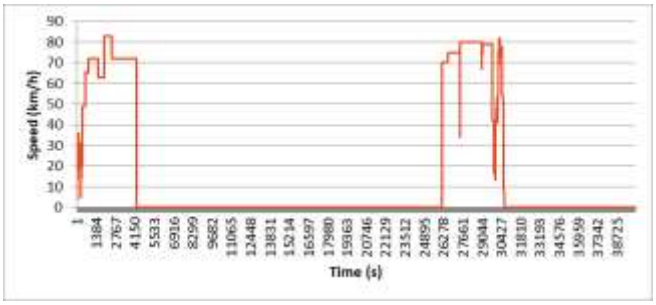


Fig. 5. Speed: VAB-samarco.

The first recharge lasted 122 minutes with an average recharge power of 44.32 kW. The second recharge, 120 minutes and 50.43 kW of average recharge power. The station's power was 80 kW throughout the period. This reduced power is a result of the lithium battery charging characteristic.

C. Route 2 – RJN to Suzano

This route has 53 km per stretch, about 106 km total, representing a consumption of less than one third of the expected range for the electric bus. For this route, there is only one station in RJN. The bus was simulated to leave the station with 100% *SoC*, arrive destination and drop off the passengers. After that, staying in place for 7 hours until the time to return to the bus station. The considered gross weight was also 30 people weighing 75 kg. The *SoC* graph of this process is in Figure 6.

Fig. 6. SoC: RJN-suzano.

In the first stretch of the trip (outbound) energy consumption was 64.42 kWh, with an average of 0.823 km/kWh and 266 km of autonomy. The second stretch had 68.42 kWh of energy consumption, 0.775 km/kWh on average and 251 km of autonomy. This stretch had a positive slope variation, requiring more energy consumption and, therefore, increasing energy consumption.

In terms of speed (Fig. 7), in the first stretch the maximum speed was 83 km/h and the average was 75.12 km/h. In the second one, the maximum was 83 km/h and the average, 67.47 km/h. The reduction in autonomy is due to the increase in average speed, representing a higher energy consumption per kilometer.

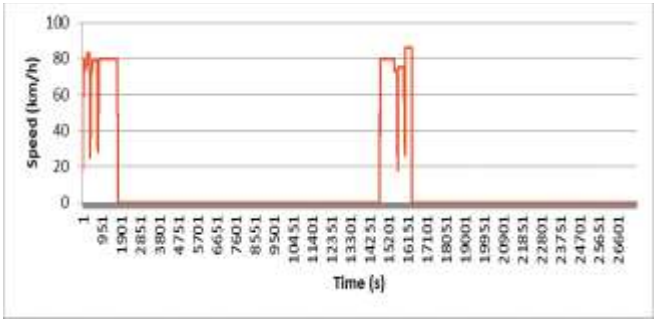


Fig. 7. Speed: RJN-suzano.

The single recharge of this route lasted 183 minutes and had an average recharge power of 42.57 kW (station power was maintained at 80 kW throughout the period).

IV. STATE OF HEALTH CALCULATION ALGORITH

The State of Health (SoH) calculation was developed with the data acquired through the simulations of the electric bus routes. In II-43, the electric vehicle lithium battery model and the calculation characteristics for implementation in HIL modeling are discussed. According to this, two different *SoH* algorithms were elaborated for continuous application through data repetition of the routes. The objective was to acquire the *SoH* curve with two different patterns to be used in the electric bus business model.

The limitation of HIL is that the simulation takes a long time for the *SoH* reduction to be visible. Therefore, the data obtained in the simulations were manipulated in Python and expanded to a sufficient number of routes in order to reduce the bus's operating capacity.

In II-43, the algorithm proposes the variation of the *SoH* based on the information of *SoC*, electric current and battery temperature, being synthesized with a set of equations to describe the reduction of the nominal capacity of the cells, with the *SoH* directly influencing the *SoC*. This set of equations were implemented in Python with the cell input data that were introduced in the simulation.

It is important to highlight the role of cell aging and its expected impact on outcomes. Battery degradation depends on both usage cycles and age. The work presented in [20] demonstrates that more than 75% of the storage capacity disappears due to this factor. This factor is called calendar aging and has the greatest impact on electric vehicles that spend most of their life cycle unused, parked. In this scenario, the great impact will be the cycle aging, due to electric buses constant use on routes.

The cycle is defined as charging or discharging the cells, so while using the vehicle it will be discharging. The EoL (End of Life) was determined according to the above assumption and the dependence of the useful life on the cycle depth, that is, the Depth of Charge (DoD) is notable. DoD is how much the battery is discharged before being charged to 100% again. In this case, with the battery being discharged to 0% *SoC* and then charged again, it is assessed as the worst case for battery life. An example Cycle Aging chart is in the Figure 8.

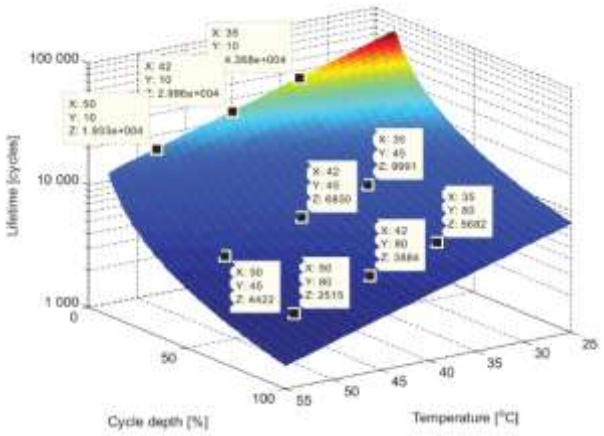


Fig. 8. Example of cycle ageing [20].

V. RESULTS

The Python algorithm was implemented using the simulations performed in HIL as input data. The number of simulated cycles was proportional to the arrival at 80% of *SoH* for each route.

A. *Route 1 – VAB to Samarco*

On this 146 km route, the vehicle will leave with 100% *SoC* from VAB and head to Samarco, where it will arrive with around 73% *SoC* and will charge again to 100% during the 8 hours and 45 minutes of waiting for departure. In Figure 9 it is possible to verify the influence of the SoH on the SoC over the cycles. This Figure shows the consumption of the route as shown in Figure 4, but with the different curves describing the behavior with low SoH. Note that the slope of these curves becomes more prominent due to the smaller amount of energy available.

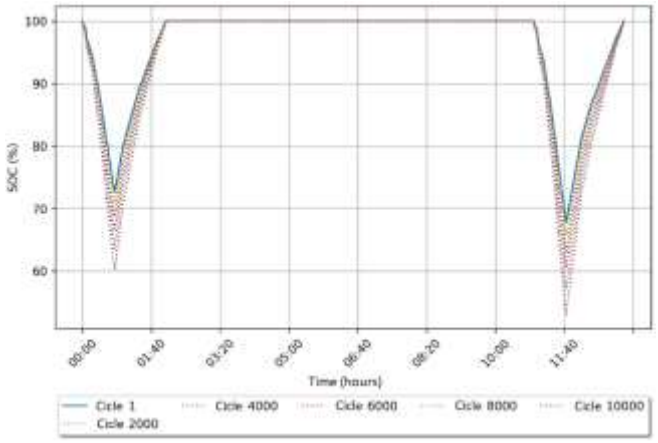


Fig. 9. SoC with SoH influence: VAB to Samarco.

Fig. 10 shows the *SoH* curve in relation to the number of cycles/routes. On this route, the battery undergoes two partial cycles, with DoD (Depth of Discharge) of 30%, but in the graph it is described as just one cycle for better comparison with the results of the other route (RJN-Suzano). The number of cycles developed until reaching 80% *SoH* is 9,487, totaling about 1,385,102 km of useful life, much higher than the 500,000 km warranty provided by the bus. According to [21], it is common for lithium cell manufacturers to define the amount of 80% as EoL (End of Life) of the battery, then moving on to its second life that can be used as a battery in microgrids, for example. One of the main reasons is the battery degradation speed which increases significantly from 80%.

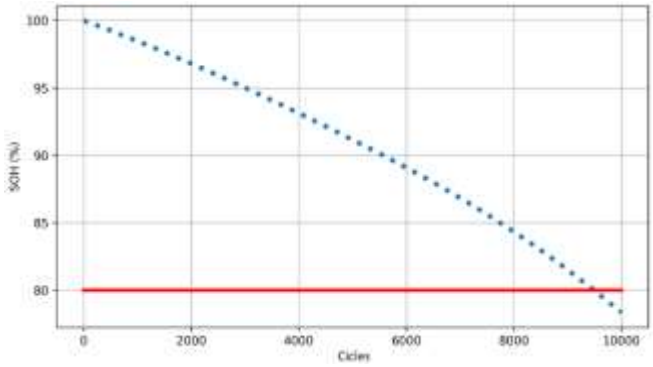


Fig. 10. SoH: VAB to Samarco.

The *SoH* directly impacts the amount of energy stored in the lithium cells, with a direct proportion to the vehicle's autonomy, so if the battery bank has 324 kWh, according to the bus datasheet, at 80% of *SoH* it will have 259.2 kWh. Figure 11 shows the range of the vehicle according to the range previously calculated in section 4.2, with the beginning SoH equal to 100%) being equivalent to the average between the ranges found, 251.07 km, in 80% the range of 200.86 km.

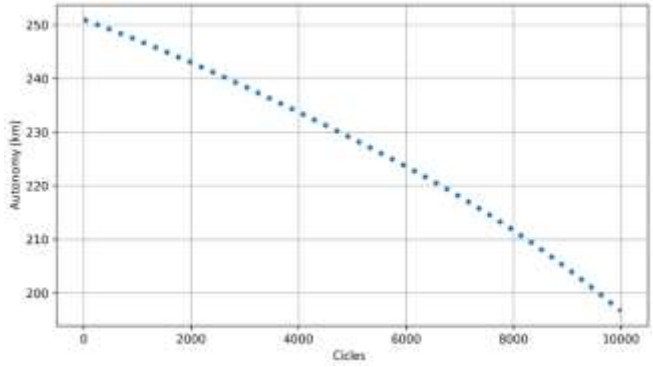


Fig. 11. Autonomy: VAB to Samarco.

B. *Route 2 – RJN to Suzano*

On this 106 km route the vehicle will leave with 100% SoC from Rodoviária de João Neiva and head to Suzano, where it will arrive with about 80% SoC and will wait for 7 hours to leave. Departing from Suzano and loading only in RJN again. Figure12 shows the SoC of the vehicle and the influence of the SoH over the cycles on it, based on Figure 6.

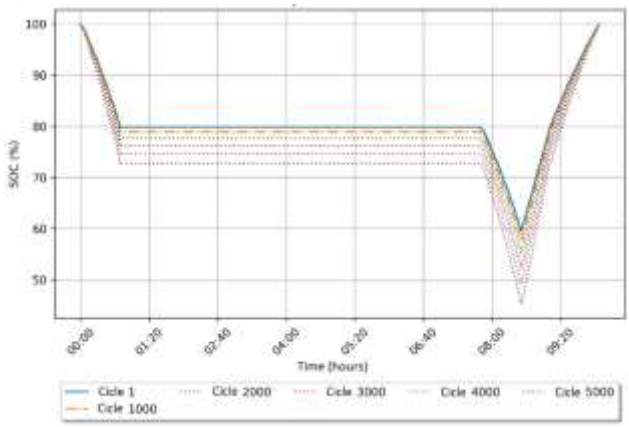


Fig. 12. SoC with SoH influence: RJN-suzano

Fig. 13 shows the *SoH* curve in relation to the number of cycles/routes. On this path, *SoC* reaches 60% with 100% *SoH*, a DoD of 40%, describing in each cycle a complete route between João Neiva, Suzano and return.

The number of cycles developed until reaching 80% *SoH* is 4,394, totaling about 465,764 km of useful life, less than the 500,000 km warranty provided by the bus. In one year, there are about 260 working days, totaling about 260 cycles per year and, therefore, equivalent to almost 17 years of useful life, if two cycles were performed per day, it is equivalent to 8 years and 164 days. Battery bank EoL is achieved much earlier using this route, according to the *SoH* algorithm used, which is consistent with Figure 8 which demonstrates the correlation between DoD and battery life.

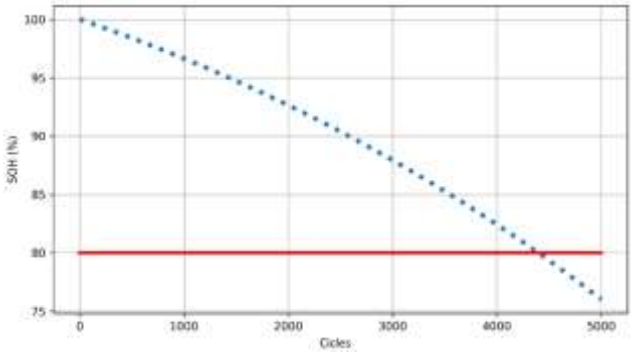


Fig. 13. SoH: RJN-suzano.

In Fig. 14, the vehicle autonomy is displayed according to the autonomy calculated earlier, using a safety margin, it is possible to assume that the battery bank continues to be used below 80% of *SoH*, but the degradation would reach what is called “knee”, a point where degradation is no longer linear and becomes exponential. It is possible to verify that the *SoH* derivative increases with its reduction, even though this algorithm is not ideal for calculating the *SoH* curve after linearity.

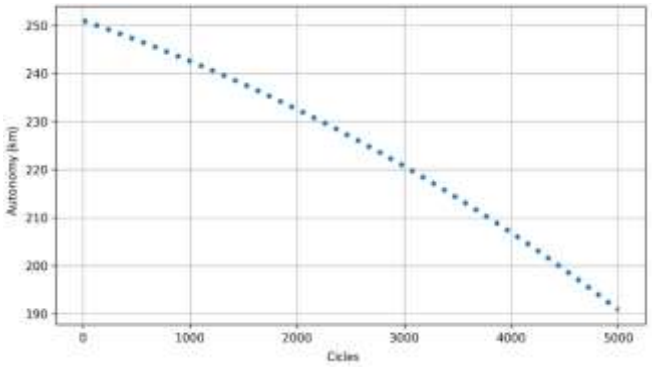


Fig. 14. Autonomy: RJN-suzano.

VI. CONCLUSION

In the simulation, it was possible to verify the reduction of the energy capacity of the battery according to the number of cycles developed according to two different routes. The route between VAB and Samarco has better viability due to the lower DoD, with a lower impact on the *SoH* of the battery.

According to the simulation, it would be possible to complete the route 2.16 times more compared to the RodoviáriaJoão Neiva and Suzano bus station. In general terms, because it is a simulation without the use of data from the electric bus itself, such as the telemetry of these trips and the undisclosed data about the electrical characteristics of the bus battery and motor, there is an error associated with these results. The proportion of the difference between the two routes may not be as significant as evidenced in these simulations, but the result must be evaluated based on other scenarios.

In addition, one of the facts that make the result of the VAB-Samarco superior, is its mid-way recharge. This highlights the importance of having a good network of stations, something essential for electric mobility in public transport.

ACKNOWLEDGMENT

The authors thank the Research Program and Technological Development of the electricity sector regulated by ANEEL and EDP Brazil for the financial support to the project. This work is part of the project under number PD-0064-1054-2019.

REFERENCES

[1] C. N. do Transporte. (2017) Transporte rodovi´ario de passageiros em regime de fretamento. [Online]. Available:https://cnt.org.br/transporte-rodoviario-passageiros-regime-fretamento

[2] S. Bus. (2019) Electric bus, main fleets and projects around the world.[Online].Available:https://www.sustainable-bus.com/electric-bus/electric-bus-public-transport-main-fleets-projects-around-world/

[3] L. S. Dos Santos, C. Q. Pica, R. S. de Moura, P. R. Belin, M. A. I. Martins, J. C. de Bona, and L. H. Cruz, “Business models for charter electric bus fleets,” in 2022 Intermountain Engineering, Technology and Computing (IETC). IEEE, 2022, pp. 1–5.

[4] EVANNEX. (2019) Just how long will an ev battery last? [Online]. Available: https://insideevs.com/news/368591/electric-carbattery-lifespan/

[5] A. Franca, J. A. Fernandez, C. Crawford, and N. Djilali, “Assessing the impact of an electric bus duty cycle on battery pack life span,” in 2017 IEEE Transportation Electrification Conference and Expo (ITEC), 2017, pp. 679–683.

[6] A. Gailani, R. Mokidm, M. Al-Greer et al., “Analysis of lithium-ion battery cells degradation based on different manufacturers,” in 2020 55th International Universities Power Engineering Conference (UPEC). IEEE, 2020, pp. 1–6.

[7] X. Wu, Y. Chen, X. Han, J. Du, T. Wen, and Y. Sun, “Analysis of performance degradation of lithium iron phosphate power battery under slightly overcharging cycles,” in 2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI). IEEE, 2020, pp. 75–79.

[8] A. Cars. (2017) What is a powertrain control module (pcm) in cars? [Online]. Available: https://drivinglife.net/powertrain-control-modulepcm-cars/

[9] S. Raman, N. Sivashankar, W. Milam, W. Stuart, and S. Nabi, “Design and implementation of hil simulators for powertrain control system software development,” in Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251), vol. 1. IEEE, 1999, pp. 709–713.

[10] T. H. A. Library. (2020) Electric vehicle. [Online]. Available: https://www.typhoon-hil.com/documentation/typhoonhil-application-notes/References/electric vehicle.html

[11] ——. (2020) Three-phase induction machine with squirrel cage. [Online]. Available: https://www.typhoon-hil.com/documentation/typhoon-hil-schematiceditor-library/References/three-phase induction machine.html

[12] R. Jackey, M. Saginaw, P. Sanghvi, J. Gazzarri, T. Huria, and M. Ceraolo, “Battery model parameter estimation using a layered technique: an example using a lithium iron phosphate cell,” SAE Technical Paper, vol. 2, pp. 1–14, 2013.

[13] T. Huria, M. Ceraolo, J. Gazzarri, and R. Jackey, “High fidelity electrical model with thermal dependence for characterization and

simulation of high power lithium battery cells," in 2012 IEEE International Electric Vehicle Conference. IEEE, 2012, pp. 1–8.

[14] R. A. Jackey, "A simple, effective lead-acid battery modeling process for electrical system component selection," SAE Transactions, pp. 219–227, 2007.

[15] O. Tremblay, L.-A. Dessaint, and A.-I. Dekkiche, "A generic battery model for the dynamic simulation of hybrid electric vehicles," in 2007 IEEE Vehicle Power and Propulsion Conference. Ieee, 2007, pp. 284–289.

[16] C. Ziebert, A. Melcher, B. Lei, W. Zhao, M. Rohde, and H. Seifert, "Electrochemical–thermal characterization and thermal modeling for batteries," in Emerging Nanotechnologies in Rechargeable Energy Storage Systems. Elsevier, 2017, pp. 195–229.

[17] N. Batteries. (2017) Nv14 specifications. [Online]. Available: https://www.neovolta.com/wp-content/uploads/2020/02/NV14-NV24-Specifications-All022020.pdf

[18] L.Rhode, "Battery wear cost applied to bess optimization in microgrids," Energy Systems, 2020.

[19] M. Muller. (2021) evcalc - charge real range. [Online]. Available: ¨ ecalc.ch/evcalc.php

[20] M. Swierczynski, D.-I. Stroe, A.-I. Stan, R. Teodorescu, and S. K. Kær, "Suitability of the nanophosphate lifepo 4/c battery chemistry for the fully electric vehicle: Lifetime perspective," in 2014 Ninth International Conference on Ecological Vehicles and Renewable Energies (EVER). IEEE, 2014, pp. 1–8.

[21] P. Venugopal, "State-of-health estimation of li-ion batteries in electric vehicle using indrnn under variable load condition," Energies, vol. 12, no. 22, p. 4338, 2019.

Numerical Investigations Applied to Chemical Reaction-Diffusion Cycles Induced by Temperature Gradients

Mohammed LOUKILI
Avignon University
INRAE, UMR408 SQPOV
Avignon, France
mohammed.loukili@univ-avignon.fr

Raphael PLASSON
Avignon University
INRAE, UMR408 SQPOV
Avignon, France
raphael.plasson@univ-avignon.fr

Ludovic JULLIEN
Département de chimie, École normale supérieure,
PSL University, Sorbonne University
Paris,France
Ludovic.Jullien@ens.psl.eu

Abstract— **This work revolves around the analytical and numerical investigation applied to a general chemical reaction-diffusion system. The aim is to shed light on the phenomenology associated with the onset of a steady cycle of chemical reaction-diffusion, sustained in a nonequilibrium state by an externally imposed temperature gradient, leading to continuous transformation fluxes occurring in opposite direction in spatially separated regions. The corresponding reaction-diffusion equations result from the coupling of the equation describing first order chemical reactions—whose kinetics follow the Arrhenius' law—with chemical diffusion, when a temperature gradient is imposed as a form of Heaviside. An analytical study was first performed in the case of a simple two-compartment model. This model was then compared with numerical modeling of reaction-diffusion system using XMDS2 open source code. A validation test could show a good agreement between the analytical and the numerical approaches. Furthermore, the effect of different thermodynamic parameters associated with the problem was assessed, focusing on the intensity and spatial distribution of the induced chemical fluxes.**

Keywords— Diffusion, chemical reaction cycles, thermo-diffusion, Arrhenius' law, reaction-diffusion

I. INTRODUCTION

Biological systems rely on their striking ability to efficiently convert the energy from their environments for sustaining extremely organized chemical systems. Such systems are a wonderful inspiration for designing novel artificial dynamic systems. Then the challenge is to understand how chemical reactions can be connected to an external source of energy, establishing continuous cycles of chemical transformations sustained in nonequilibrium steady states. At term, this can lead to the design of interconnected reaction cycles, leading to the establishment of chains of energy transfer directed towards chemical processes of interest [1].

An often overlooked source of energy relies in the existence of thermal gradients. A spatial difference of temperature necessarily implies the establishment of a thermal diffusion flux from hot to cold regions. Chemical reaction can then be coupled to this thermal energy flux, leading to the establishment of chemical reactions cycles. Typically, the kinetics and thermodynamics of reactions depend on the temperature. This will lead to differential transformation fluxes in spatially separated regions, leading in turn to the establishment of chemical diffusion, and thus of chemical reaction-diffusion cycles. Such a very general behavior can be used to tap on thermal gradient energy for designing thermochemical systems of interests.

Firstly, harvesting waste heats is considered as a hot issue nowadays. The scientific community has increasingly concentrated its emphasis on creating the cheapest and cleanest energy possible. Furthermore, the consumption acceleration of non-renewable resources for the production of energy has led to an urgent requirement to increase the energy production from renewable resources. Eventually, the quantity of energy wasted in the atmosphere is quite high in various areas. To make use of this enormous energy, efficient and low-cost thermal energy collecting technologies are required. For sake of explanation, the appearance of a temperature differential between the electrodes, or the entropy difference between the two sides of the redox process, causes a potential difference in the thermo-electrochemical (TEC) and keeps the current circulating. Then, it could be used as an alternative device design showing increasing promise for the conversion of low-grade thermal energy. Furthermore, thermocellscan continuously generate electrical energy when a temperature gradient is present, without producing emissions or consuming any materials. It is considered as one of the most cheaply cost-effective methods for converting low-grade waste heat into power [2,3]. In addition, low-grade waste heat's main benefit is its widespread availability and abundance, which is typically continuous in nature. Waste thermal energy, is plentiful in industrial waste streams, transportation, and geothermal operations [4]. For this reason, thermo-electrochemical cells are considered as a potential new technique for gathering low-grade thermal energy and transforming it into a sustainable source of energy [5]. While the development of thermal energy harvesting devices has mostly concentrated on semiconductor-based solid-state devices [6], their commercialization has been hampered by high prices, and relatively low efficiencies and, poor long-term reliability [7]. Moreover, thermoelectric devices made of semiconductor materials create potential differences at the range of μVK^{-1} , which restricts their effectiveness at

temperatures close to ambient. Then, thermo-electrochemical cells, or thermocells, are an alternative device design for converting low-grade thermal energy that shows growing potential when a temperature differential is present, thermocells may continually create electrical energy without emitting any pollutants or using any resources. These thermocells, when based on a redox-active electrolyte, may create potential differences of the order ofmVK^{-1}. On other hand, the most notable benefit of this technology is its design and scale flexibility. In addition, thermal energy may be gathered in a variety of ways, from collecting lost heat from industrial pipes or waste streams to using body heat as a power source [8].

Secondly, reducing CO_2 emission and alleviating global warming is an important and challenging deal [9-10]. Carbon dioxide is engaged in several proton exchange reactions when dissolved in water, leading to hydrogenocarbonate and carbonate ions. For example, concentration gradients of the different hydrogenocarbonate species could be engineered for generating *pH* gradients that could in turn be used for generating electricity [11]. Inversely, it has been shown that rich non-equilibrium reaction-diffusion cycles can be established when carbonated water is submitted to stationary *pH* gradients [12]. Coupling thermal gradients to proton exchange reaction shall similarly lead to the establishment of *pH* gradients, and thus to CO_2 gradients within a thermochemical cell, enabling its thermal pumping.

Finally, thermal gradients have been successfully be used for controlling reaction as complex as polymerization reaction. These chemical transformations, implying the assembly of a potentially very large number of elementary building blocks, can in turn lead to the formation of very different compounds. Processing this transformation in non-equilibrium states can enable to select and control in a rich way [13]. It was for example possible to generate very long chains of RNA in temperature-gradient sustained non-equilibrium systems, with lengths that would have been impossible in uniform temperature [14].

The goal of this research paper is revolved around numerical modeling of reaction-diffusion system where the chemical kinetics is described by Arrhenius' law. We will focus on the onset of stationary non-equilibrium reaction-diffusion cycles, feeding energy from an external temperature gradient. For sake of clarity, the goal is to model a partial differential equation (PDE), which can simultaneously simulate many thermo-diffusion applications. A simple generic chemical transformation will be considered as a test model. For these purposes, this paper is depicted into four sections. After introducing the goal behind of this research in the first section, the second section is dedicated to problem statement where the phenomenology investigated, the numerical and theoretical method used are clearly presented. Next, the third section is dedicated to results and discussions. Finally, the last section is devoted to conclusion and perspectives.

II. PROBLEM STATEMENT

A. *Chemical system*

Let us consider a simple system consisting in two chemical components A_1 and A_2, engaged in a single reaction written as

$$A_1 \underset{k_-}{\overset{k_+}{\rightleftharpoons}} A_2 \tag{1}$$

At temperature *T*, the kinetic parameters associated with the reaction are expressed following the Arrhenius equation:

$$k^- = A.e^{-\frac{E_a}{RT}} \tag{2}$$

$$k^+ = A^+.e^{-\frac{E_a^+}{RT}} \tag{3}$$

E_a and E_a^+ are the activation energies, and A and A^+ the pre-exponential Arrhenius factors for, respectively, k^- and k^+. The thermodynamic equilibrium constant can then be expressed as:

$$K = \frac{k^+}{k^-} = K^{\#}.e^{-\frac{\Delta H}{RT}} \tag{4}$$

with $K^{\#} = A^+/A$ and $\Delta H = E_a^+ - E_a$.

Assuming that we observe temperature variations $T=T^0+\delta T$ around T^0, equations (2)-(3) comes down to:

$$k^- = k^0.e^{\frac{Ea}{RT^0}*\frac{\delta T}{\delta T+T^0}} \tag{5}$$

$$K = K^0.e^{\frac{\Delta H}{RT^0}*\frac{\delta T}{\delta T+T^0}} \tag{6}$$

with

$$k^0 = A.e^{-\frac{Ea}{RT^0}} \tag{7}$$

$$K^0 = K^{\#}.e^{-\frac{\Delta H}{RT^0}} \tag{8}$$

where k^0 and K^0 being respectively the kinetic and thermodynamic equilibrium constants of the reaction (1) at $T=T^0$.

If the reaction (1) is an elementary step, the kinetics can be expressed as

$$v^+ = k_+.a_1 \tag{9}$$

$$v^- = k_-.a_2 \tag{10}$$

where v^+ and v^- are the kinetic rates of the forward and backward reactions, respectively, and a_1 and $a_2$are the concentrations of A_1 and A_2 respectively.

Assuming this chemical system to be submitted to non-uniform conditions in the sole *x*-direction, the governing equations associated with this reaction are expressed as

$$\begin{cases} \frac{\partial a_1}{\partial t} = D_1 \frac{\partial^2 a_1}{\partial x^2} - v^+ + v^- \\ \frac{\partial a_2}{\partial t} = D_2 \frac{\partial^2 a_2}{\partial x^2} + v^+ - v^- \end{cases} \tag{11}$$

where D_1 and D_2 are respectively the diffusion coefficients of A_1 and A_2.

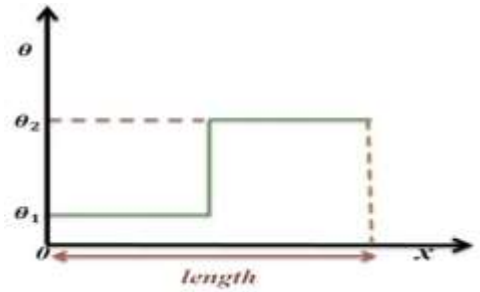


Fig.1. Temperature profile.

B. Spatial environment

We suppose that the system is submitted to a gradient of temperature with a Heaviside shape (see Fig.1). We suppose that the exchanges of energy with the surrounding environment are sufficiently efficient for assuming a steady temperature profile. Finally, we consider a closed system, with the following initial and boundary conditions:

$$a_1(\forall x, t=0) = a_2(\forall x, t=0) = 0.5\, mol/m^3 \quad (12)$$

$$\frac{\partial a_1}{\partial x}\Big)_{x=0,L} = \frac{\partial a_2}{\partial x}\Big)_{x=0,L} = 0\, mol/m^3 \quad (13)$$

C. System reduction

In order to minimize the number of parameters associated with the problem studied, the chemical reaction (1)is reformulated in dimensionless form as,

$$U_1 \underset{\kappa^-}{\overset{\kappa^+}{\rightleftharpoons}} U_2 \quad (14)$$

By reformulating the system of equations (1),the dimensionless variables are adopted as;$u_i = \frac{C}{C_0}$,$\lambda = \frac{x}{l_0}$,$\tau = \frac{t}{t_0}$, $\delta_i = D_i \frac{t}{l_0^2}$, $\kappa^+ = \mathrm{k}^+ . t_0$, $\kappa^- = \mathrm{k}^- . t_0$, $\varepsilon_+ = \frac{E_a^+}{RT^0}$, $\varepsilon_- = \frac{E_a^-}{RT^0}$, $\hbar = \sqrt{K^0}$, $\rho = c_0 t_0 k^0 \hbar$, $\Delta H = E_a^+ - E_a^-$, $E_a = E_a^-$, $\theta = \frac{T-T_0}{T_0} = \frac{\delta T}{T_0}$.

By defining γ and β as:

$$\gamma = \frac{\varepsilon_+ - \varepsilon_-}{\varepsilon_+ + \varepsilon_-}, \quad \beta = \frac{\varepsilon_+ + \varepsilon_-}{2} \quad (15)$$

$\varepsilon_+$and ε_- can be rewritten as:

$$\varepsilon_+ = (1+\gamma)\beta \quad (16)$$

$$\varepsilon_- = (1-\gamma)\beta \quad (17)$$

Then, the governing dimensionless equations associated with the reaction-diffusion problem are expressed as:

$$\begin{cases} \frac{\partial u_1}{\partial \tau} = \delta_1 \frac{\partial^2 u_1}{\partial \lambda^2} - \vartheta^+ + \vartheta^- \\ \frac{\partial u_2}{\partial \tau} = \delta_2 \frac{\partial^2 u_2}{\partial \lambda^2} + \vartheta^+ - \vartheta^- \end{cases} \quad (18)$$

δ_1 ,δ_2 are respectively dimensionless diffusion coefficients associated with the system of equations (18), Furthermore, the dimensionless concentrations of the species are expressed as $[U_1] = u_1$, $[U_2] = u_2$. Finally, the dimensionless thermodynamic parameters are written as:

$$\vartheta^+ = \kappa^+ u_1 \quad (19)$$

$$\vartheta^- = \kappa^- u_2 \quad (20)$$

$$K = \frac{\kappa^+}{\kappa^-} \quad (21)$$

$$\kappa^+ = \hbar . \rho . e^{\left(\frac{\theta}{1+\theta}\right).(1+\gamma)\beta} \quad (22)$$

$$\kappa^- = \hbar^{-1} . \rho . e^{\left(\frac{\theta}{1+\theta}\right).(1-\gamma)\beta} \quad (23)$$

The dimensionless chemical systems can thus be fully described as a function of $u_1(\lambda,\tau)$,$u_2(\lambda,\tau)$ and $\theta(\lambda,\tau)$, λ and τ being respectively the dimensionless space and time parameters. This system can be described in term of the parameters $\hbar$, ρ, γ and β being respectively linked to the equilibrium constant ($K^0 = \hbar^2$), the kinetic constant ($\rho = c_0 t_0 k^0$for $\hbar = 1$), the energetic dissymmetry between the two compounds ($\varepsilon_+ = \varepsilon_-$ for $\gamma = 0$), and the average activation energy ($\beta = \frac{\varepsilon_+ + \varepsilon_-}{2}$).

D. Compartment model

The investigation will cover chemical reaction-diffusion cycles for different values of the thermodynamic parameters, and assess the transformation fluxes in spatially separated regions. In order to formulate the theoretical framework associated with the chemical reaction-diffusion cycle, the reaction-diffusion system can be simplified as a two compartment model, corresponding to the limit case where each spatial zone with uniform temperature can be considered as chemically homogeneous. The compounds U_1 and U_2 are thus considered to be present in each compartment, while the chemical reaction occurs differently in each compartment, leading to different concentrations. The chemical diffusion can then be modeled as a simple transfer of each compound from one compartment to the other (see Fig. 2).

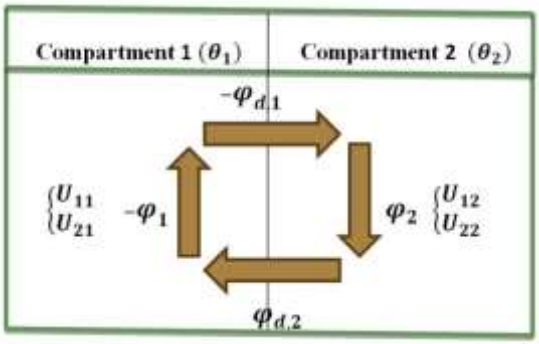


Fig. 2. Establishment of a reaction-diffusion cycle in the two compartments model; $U_{i,j}$ is the compound U_i in the compartment j; $\varphi_j$the chemical flux in compartment j; $\varphi_{d,i}$ the diffusion flux of U_i from compartment 2 to 1.

In that compartment model, the system of equations describing the problem comes down to:

$$\begin{cases} \frac{\partial u_{11}}{\partial \tau} = +\varphi_{d,1} - \varphi_1 \\ \frac{\partial u_{12}}{\partial \tau} = -\varphi_{d,1} - \varphi_2 \\ \frac{\partial u_{21}}{\partial \tau} = +\varphi_{d,2} + \varphi_1 \\ \frac{\partial u_{22}}{\partial \tau} = -\varphi_{d,2} + \varphi_2 \end{cases} \quad (24)$$

where φ_1 and φ_2 are the chemical fluxes expressed at each compartment, and $\varphi_{d,1}$ and $\varphi_{d,2}$ the diffusion flux of each compound from compartment 2 to compartment 1:

$$\varphi_1 = k_1^+ u_{11} - k_1^- u_{21} \quad (25)$$

$$\varphi_2 = k_2^+ u_{12} - k_2^- u_{22} \quad (26)$$

$$\varphi_{d,1} = \delta(u_{12} - u_{11}) \quad (27)$$

$$\varphi_{d,2} = \delta(u_{22} - u_{21}) \quad (28)$$

k_1^+, k_1^-, k_2^+, $k_2^-$are the kinetic parameters associated with the reaction expressed at each compartment, and δ the kinetic parameter exchange of compounds between the two compartments. In the steady state, this leads to a circular reaction-diffusion flux $\varphi = -\varphi_{d,1} = -\varphi_1 = \varphi_{d,2} = \varphi_2$.

Moreover, the system being closed, the sum of all concentrations is constant:

$$u_{11} + u_{12} + u_{21} + u_{22} = 4\,u_0 = 2 \quad (29)$$

$$\varphi = \frac{\delta(\kappa_1^+\kappa_2^- - \kappa_1^-\kappa_2^+)}{\delta(\kappa_1^+ + \kappa_1^- + \kappa_2^+ + \kappa_2^-) + \kappa_1^+\kappa_2^+ + \kappa_1^+\kappa_2^- + \kappa_1^-\kappa_2^+ + \kappa_1^-\kappa_2^-} \quad (30)$$

$$\varphi = \frac{\delta(K_1 - K_2)}{\delta\left(\frac{1+K_1}{\kappa_2^-} + \frac{1+K_2}{\kappa_1^-}\right) + 1 + K_1 + K_2 + K_1K_2} \quad (31)$$

From this latter equation, two necessary conditions for the establishment of a reaction-diffusion cycle are easily identified. A first condition is the possibility to observe chemical exchange between the two compartments, i.e. $\delta \neq 0$. A second condition is to obtain different equilibrium constants due to the different temperatures, i.e. $K_1 \neq K_2$. This expression can be expressed as a function of the set of parameters described in equations (32-38). Assuming the temperature is symmetrical between both compartments such as $\theta_1 = -\theta_2 = \theta$, and with the approximation that the temperature variation is weak ($\theta \ll 1$), then the thermodynamic and kinetic parameters can be expressed as:

$$\kappa_1^+ = \hbar\,.\,\rho\,.\,e^{\left(\frac{-\theta}{1-\theta}\right).(1+\gamma)\beta} \approx \hbar\,.\,\rho\big(1 - \theta\beta(1+\gamma)\big) \quad (32)$$

$$\kappa_2^+ = \hbar\,.\,\rho\,.\,e^{\left(\frac{\theta}{1+\theta}\right).(1+\gamma)\beta} \approx \hbar\,.\,\rho\big(1 + \theta\beta(1+\gamma)\big) \quad (33)$$

$$\kappa_1^- = \hbar^{-1}\,.\,\rho\,.\,e^{\left(\frac{-\theta}{1-\theta}\right).(1-\gamma)\beta} \approx \hbar^{-1}\,.\,\rho\big(1 - \theta\beta(1-\gamma)\big) \quad (34)$$

$$\kappa_2^- = \hbar^{-1}\,.\,\rho\,.\,e^{\left(\frac{\theta}{1+\theta}\right).(1-\gamma)\beta} \approx \hbar^{-1}\,.\,\rho\big(1 + \theta\beta(1-\gamma)\big) \quad (35)$$

$$K_1 = \hbar^2 e^{-2.\theta.\gamma\beta} \approx \hbar^2(1 - 2.\theta.\gamma.\beta) \quad (36)$$

$$K_2 = \hbar^2 e^{2.\theta.\gamma\beta} \approx \hbar^2(1 + 2.\theta.\gamma.\beta) \quad (37)$$

The expression of the chemical flux is then reduced to:

$$\varphi = \frac{-4.\delta.\theta.\gamma.\beta.\rho.\kappa^2}{(1+\kappa^2)[2.\delta\kappa + \rho(1+\kappa^2)]} \quad (38)$$

This expression can be further simplified depending on the relative values of δ and ρ.

$$\delta \ll \rho \Rightarrow \varphi = -4.\delta.\theta.\gamma.\beta\frac{\kappa^2}{(1+\kappa^2)^2} \quad (39)$$

$$\delta \gg \rho \Rightarrow \varphi = -2.\rho.\theta.\gamma.\beta\frac{\kappa}{(1+\kappa^2)} \quad (40)$$

In all the cases, the induced reaction-diffusion flux is proportional to θ, γ and β. When the chemical diffusion is slow compared to the chemical reaction (i.e. $\delta \ll \rho$), the diffusion is the limiting factor, and the flux is proportional to δ. Conversely, when the chemical reaction is slow compared to the chemical diffusion (i.e. $\delta \gg \rho$), the reaction is the limiting factor, and the flux is proportional to ρ.

E. Numerical Procedure

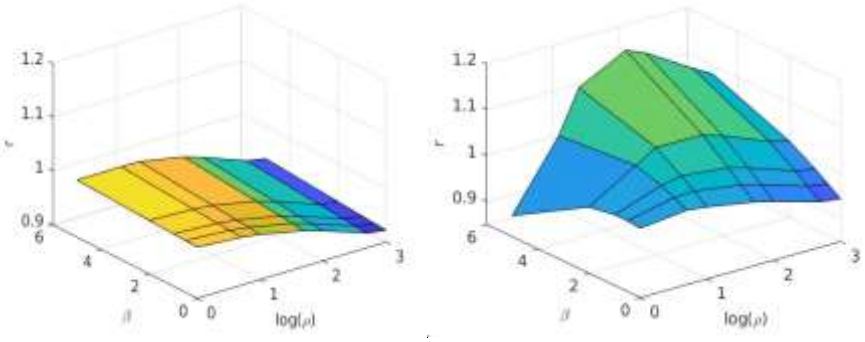
The numerical method adopted in this work is based on XMDS2 open source code [15,16]. XMDS2 is a code published under a GPL license, for the numerical integration of partial differential equations (PDE). The model to be solved is described in an XML-based script, which is converted by XMDS2 into a C++ code, greatly reducing the time required to create an efficient simulation program. The generated model can be adopted numerically to integrate initial value issues involving problems from single ordinary differential equations to systems of coupled stochastic partial differential equations. Furthermore, the use of XMDS2 makes it possible to express difficulties in a straight forward XML format rather than having to manually write hundreds of lines of code, which is possibly prone to errors. The spatial term of the equations is treated by applying the spectral method [17]. Then, the adaptive eighth-ninth order Runge–Kutta [18] issued for solving the system of equations governing the problem.

III. RESULTS AND DISCUSSION

This section is depicted into two separate subsections. The first subsection, is dedicated to the validation test by illustrating two figures associated with the ratio $\frac{\varphi_{Num}^{max}}{\varphi_{Theo}}$, that is related respectively to the maximum of theoretical and numerical chemical fluxes. Thereafter, the second subsection is devoted to investigate the behavior of chemical fluxes subjected to gradient of temperature as a form of Heaviside. The investigation will cover all dimensionless thermodynamic parameters associated with the problem.

A. Validation test

To ensure the validity of the numerical approach used in this work, we present in the Fig. 3 the profiles associated with the report $r = \frac{\varphi_{Num}^{max}}{\varphi_{Theo}}$ for different temperatures $\theta = 0.01$ and 0.15, as versus of ρ that are ranged from 10^0 to 10^3,and β from 0.01 to 5, and fixing γ at 0.2. φ_{Num}^{max} corresponds to the maximal value of the chemical flux obtained in the system computed by PDE integration, and φ_{Theo}corresponds to the value predicted by the compartment model, as defined in eq.(31). These plots were also computed for several values of γ ranging from 0.2 to 1, resulting in almost identical values.



(a) (b)

Fig. 3. Ratio r of the maximum numerical chemical fluxes over the value predicted in the compartment model, as a function of parameters β and ρ for γ=0.2 and θ=0.01 (a), or θ=0.15 (b).

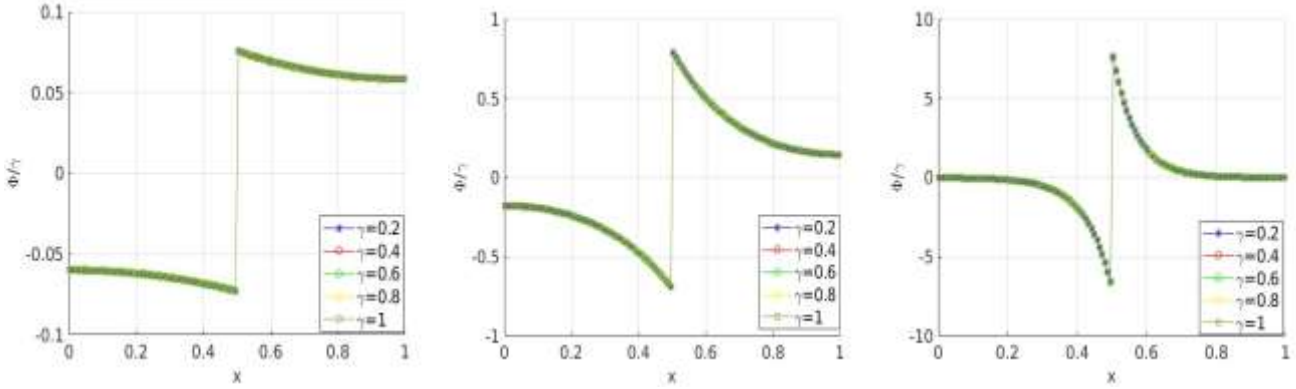
The results illustrated in the Fig. 3, for different temperatures $\theta = 0.01$ and 0.15 show an acceptable agreement between the analytical and the numerical approaches, which confirms the validity of the numerical method used in this research paper.

The ratio is very close to 1 in the case of $\theta = 0.01$ (which would correspond to variations of few degrees from an average temperature of 300K), in good agreement with equation (27). In that situation, the spatially distributed system thus behaves similarly to the two-compartment system. A small deviation can however be observed for larger values of ρ, where the chemical flux decreases moderately (about 5% decrease for $\rho = 10^3$).

When the temperature variation is more important, with $\theta = 0.15$ (e.g. corresponding to variations of temperature of $\pm 45\ K$ from an average temperature of 300K), the simple compartment model still holds, albeit more approximately. Interestingly, it can be seen that the spatially distributed system leads to higher chemical fluxes than in the compartment model, with increases of almost 20% for optimal values of β and ρ.

B. Spatial distribution of chemical fluxes

To describe clearly the phenomenology associated with the cycle of chemical reaction-diffusion, the spatial distribution of the chemical fluxes was plotted. They indicate where in the system the chemical transformations are steadily sustained at nonequilibrium. In the Fig. 4, chemical fluxes are presented for ρ ranging from 10^0 to 10^2,γ ranging from 0.2 to 1, and β being fixed at 1 for θ=0.15.



(a) (b) (c)

Fig. 4.Variation of the ratio $\frac{\varphi}{\gamma}$as a function of the spatial position λ, forγranging from 0.2 to 1, $\beta = 1$, and (a) $\rho = 10^0$, (b) 10^1 and (c) 10^2.

Once again, it is observed that the chemical fluxes are in all cases precisely proportional to γ. Furthermore, ρ is one of key parameters associated with the system spatial behavior. The results show that when$\rho$is small the reaction tends to happen in the full system, with homogeneous behavior in each temperature region. For large values, the chemical reaction only takes place in the interface between hot and cold regions, nothing happening far from it. This can be linked with the previous observation of a system behaving similarly as a two compartment system for low values of ρ, and deviating from it for larger values.

IV. CONCLUSION AND PERSPECTIVES

This research paper investigates a thermal reaction-diffusion system, where a first-order chemical is subjected to a gradient of temperature adopting a Heaviside shape. It could be shown that the system can be described in first approximation as a two-compartment model. The temperature gradient will generate two chemical transformations sustained in opposite directions in each hot and cold region, chemical exchanges between the two regions allowing the system to be maintained in a nonequilibrium steady-state.

When the thermal gradient is moderate, the induced chemical flux is essentially proportional to the kinetic and thermodynamic parameters. However, it could be observed a competition between the chemical diffusion (characterized by the parameter δ) and the chemical transformation rate (characterized by the parameter ρ), the chemical flux being limited by the slower of these two processes, while being optimal when they are of the same order of magnitude.

In a spatially extended system, a system limited by the chemical transformations (low values of ρ)is characterized by chemical transformations occurring in the whole system, and a system limited by the chemical diffusion (high values of ρ) is characterized by chemical transformations occurring only at the hot/cold interface. In all cases, the chemical fluxes are maximized close to the interface. This result is of important consequences for the application of these temperature gradient induced reaction-diffusion systems. Typically, for systems designed for chemical exchanges with the surrounding (e.g. in the case of CO_2 heat pump), the chemical diffusion would have to be optimized so that the chemical reactions can be processed at the extremity of the system.

As a perspective, the following step will be the study of nonlinear system, associated with higher order reactions, in order to get closer to real applications.

REFERENCES

[1] T.Saux, R.Plasson, andL. Jullien,"Energy propagation throughout chemical networks,"Chem. Commun, Vol. 57, 2021.

[2] Dupont. M, Macfarlane. D, Pringle. J,"Thermo-electrochemical cells for waste heat harvesting—Progress and perspectives," Chem. Commun, Vol. 53, pp. 6288–6302, 2017.

[3] Zhang. L, Kim. T, Li. N, Kang, T, Chen. J, Pringle, J, Zhang. M, Kazim. A.H, Fang. S, Haines, C, et al, "High Power Density Electrochemical Thermocells for Inexpensively Harvesting Low-Grade Thermal Energy," Adv. Mater, Vol. 29, 2017.

[4] J.Turner,"A Realizable Renewable Energy Future," Science, Vo. 285, pp, 687–689.1999

[5] A.A.M. Sayigh, Renewable Energy, Technology and the Environment, Elsevier Science, 2012

[6] K. Biswas, J. He, I.D. Blum, C.-I. Wu, T.P. Hogan, D.N. Seidman, V.P. Dravid, M.G. Kanatzidis,"High-performance bulk thermoelectrics with all-scale hierarchical architectures," Nature, Vol.489,pp.414–418,2012.

[7] T. Mancini,P.Heller, B. Butler, B.Osborn, W. Schiel, V.Goldberg, R.Buck, R.Diver, C. Andraka, J. Moreno, "Dish-Stirling Systems: An Overview of Development and Status," Journal of Solar Energy Engineering, Vol. 125, pp.135–151, 2003.

[8] T.J. Abraham, D.R. MacFarlane, J.M. Pringle,"HighSeebeck coefficient redox ionic liquid electrolytes for thermal energy harvesting," Energy & Environmental Science, Vol. 6, pp. 2639–2645.

[9] Züttel. A, Mauron. P, Kato. S, Callini, E. Holzer. M, Huang. J, "Storage of Renewable Energy by Reduction of CO2 with Hydrogen,"ChimiaVol. 69, pp. 264−268, 2015,

[10] Kato. S, Matam. S. K, Kerger. P, Bernard. L, Battaglia; C, Vogel. D, Rohwerder. M, Züttel, A, "The Origin of the Catalytic Activity of a Metal Hydride in CO_2 Reduction," Angew. Chem., Int. Ed, Vol. 55, pp. 6028−6032, 2016.

[11] T. Kim, B. E. Logan, and C. A. Gorski, " A pH-Gradient Flow Cell for Converting Waste CO_2 into Electricity," Environ. Sci. Technol. Lett. Vo. 4, pp.49−53, 2017.

[12] M. Emond, T. LeSaux, J.-F. Allemand, P. Pelupessy, R. Plasson, and L. Jullien,"Energy Propagation Through a Protometabolism Leading to the Local Emergence of Singular Stationary Concentration Profiles," Chem. Eur. J., Vol. 18, pp. 14375-14383, 2012.

[13] A. Sorrenti, J. Leira-Iglesias,A. Markvoort, T. Greef, T.Hermans,"Non-equilibrium supramolecular polymerization", Chem. Soc. Rev, Vol. 46,pp. 5476-5490, 2017.

[14] C.B. Mast, S. Schink, U. Gerland, D. Braun, "Escalation of polymerization in a thermal gradient", ProcNatlAcadSci USA, Vol 110(2) pp 8030-8035,2013.

[15] G. Dennis, J. Hope, M.Johnsson, "XMDS2: Fast, scalable simulation of coupled stochastic partial differential equations," Computer Physics Communications, Vol. 184, pp. 201-208,2013.

[16] J.J. Hope, G.R. Dennis, M.T. Johnsson,"XMDS website and documentation",http://www.xmds.org

[17] J.P. Boyd,"Chebyshev and Fourier Spectral Methods", second ed., Dover, 2000.

[18] C. Tsitouras," Optimized explicit Runge–Kutta pair of orders 9 (8)", Appl. Numer. Math, Vol. 38, pp.123–134, 2001.

A Comparative Study of Multiple Regression, ANN and Response Surface Method for Machining Force

Chahrazed Hiba Mimoun
Research Laboratory of Industrial Technologies, Faculty of Applied Sciences
University of Tiaret
Tiaret, Algeria
hiba.mimoun@univ-tiaret.dz

Kamel Haddouche
Research Laboratory of Industrial Technologies, Faculty of Applied Sciences
University of Tiaret
Tiaret, Algeria
haddouchekam@gmail.com

Souâd Makhfi
Research Laboratory of Industrial Technologies, Faculty of Applied Sciences
University of Tiaret
Tiaret, Algeria
souad.makhfi@univ-tiaret.dz

***Abstract*—In this work, a comparative study of predictive models for machining force in turning has been investigated. The developed models use the Multiple Regression, Artificial Neural Networks, and Response Surface Method. Simulations, based on our experimental tests, were performed under Statgraphics and Matlab. The prediction of cutting force by Response Surface Method is most powerful because it gives a great determination coefficient, minimal MSE and minimal MAPE.**

Keywords—machining force, multiple regression, artificial neural network, response surface method, modeling, simulation.

I. INTRODUCTION

Cutting forces are dependable variables in the machining process; they have been used for this in a variety of applications, including adaptive control, monitoring, and on-line tool wear observation. One of the main issues with the cutting theory is the modeling of machining force. The development of a theoretical model to accurately represent the cutting process is fairly challenging since many factors have a significant impact on the machining forces. Many researchers [1-12] have investigated the problem of modeling or estimating cutting forces. In this study, we developed different models to predict machining force in turning of steel by using carbide tool. Cutting parameters such as feed (f), cutting speed (V_c), and depth-of-cut (a_p) are the inputs of the models while machining force (F_R) is the output. To show the effectiveness of the developed models, we confronted the machining force predictions with experimental data.

II. MODELING

A. Multiple Regression

By fitting an equation to the observed data, Multiple Regression makes an attempt to predict the relationship between two or more explanatory variables and a response variable. We investigated in this study, two types of Multiple Regression: the first one is the Multiple Linear Regression without Intercept, and the second one is the Nonlinear Regression.

The following equations give respectively the mathematical formulations by:

$$F_R = a_1 f + a_2 V_c + a_3 a_p \tag{1}$$

$$F_R = k f^{\alpha} V_c^{\beta} a_p^{\gamma} \tag{2}$$

B. ANN Approach

The elaborated ANN uses a multilayer feed-forward structure: input, hidden, and output layers. The network design consists of an input layer that receives input data, an output layer that transmits final data to users, and a hidden layer that is not in direct contact with the environment.

In Fig.1, the input layer represents the cutting parameters (feed, cutting speed, depth-of-cut), and the machining force is the output of the ANN. The hidden layer consists of simple processors called neurons interconnected.

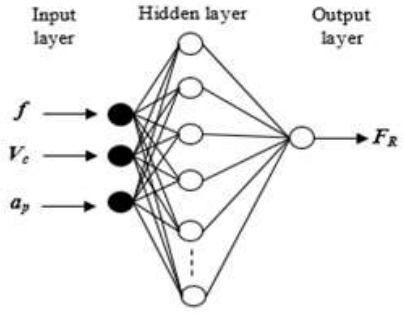


Fig. 1. Architecture of the develpped ANN.

Fig. 2 illustrated the model of an artificial neuron (model of McCulloch and Pitts [13]). The neuron output (s_j) is given by the following equation:

$$s_j = g\left(\sum_{i=1}^{m} w_{ij} \cdot e_i - b_j\right) \tag{3}$$

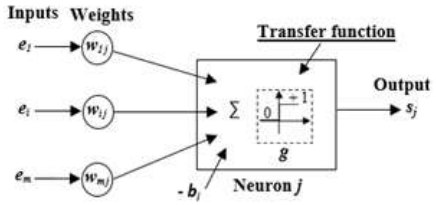


Fig. 2. Model of an artificial neuron.

The argument of the transfer or activation function (g), applied to the total of the inputs, becomes positive (equal to +1) when the activation level is equal or greater than the bias (b_j); otherwise, the argument is zero. The bias is similarly to a weight with a constant input of (-1).

The weights (w_{ij}) of neuron inputs (e_i) and (b_j) are both modifiable coefficients. Usually, the transfer function is chosen by the user. The parameters (w_{ij}) and (b_j) will then be modified by a learning or training algorithm to achieve the desired neuron input-output relationship. For training, we obtained the best results when the output layer's transfer function is linear and the hidden layer's is hyperbolic tangent sigmoid [12].

C. Response Surface Method

Response Surface design intended to select the optimal settings of a set of experimental factors. The designs involve at least three levels of the experimental factors, which must be quantitative. We used the Box-Behnken design in this study. The model based on two factors interaction, containing terms that represent main effects and second order interactions, is given by the following equation:

$$F_R = \beta_0 + \beta_1 f + \beta_2 V_c + \beta_3 a_p + \beta_{12} f \cdot V_c + \beta_{13} f \cdot a_p + \beta_{23} V_c \cdot a_p \quad (4)$$

III. EXPERIMENTAL DATA

A. Machining Process and Cutting Conditions

The machining process and the cutting conditions are defined as follows:

- The machining process is the turning of steel parts A60 (E335) with a diameter of 24 mm.
- Cutting conditions:
  - Cutting speed: V_c = 50; 70 and 90 m/min (three levels);
  - Feed: f = 0.02; 0.06 and 0.1 mm/rev (three levels);
  - Depth-of-cut: a_p = 0.25; 0.5 and 0.75 mm (three levels);
  - Nature of tool: carbide DCMT070204EN Seco;
  - The cutting is carried out dry.
- The machine is CNC lathe EMCO Compact 5 with a power of 300 W illustrated in Fig. 3.

Fig. 3. The used CNC lathe.

B. Experimental Data

We divided the experimental dataset into two databases (training and testing). On a total of 27 samples, 70% will be used for training and 30% for testing. As indicated in table I, we used 18 pairs of input-target data to train the elaborated ANN.

TABLE I. TRAINING DATA SET

Test No	Cutting Parameters			F_R (N)
	f (mm/rev)	V_c (m/min)	a_p (mm)	
2	0.02	50	0.5	102
3	0.02	50	0.75	114
4	0.06	50	0.25	102
6	0.06	50	0.75	180
7	0.1	50	0.25	120
8	0.1	50	0.5	198
11	0.02	70	0.5	77.14
12	0.02	70	0.75	81.42
13	0.06	70	0.25	72.85
15	0.06	70	0.75	141.42
16	0.1	70	0.25	90
17	0.1	70	0.5	154.28
20	0.02	90	0.5	60
21	0.02	90	0.75	66.66
22	0.06	90	0.25	60
24	0.06	90	0.75	120
25	0.1	90	0.25	76.66
26	0.1	90	0.5	133.33

On 9 additional pairs of input-target data (see table II) that weren't used in the training phase, the generalization capability will be assessed.

TABLE II. TESTING DATA SET

Test n°	Cutting Parameters			F_R (N)
	f (mm/rev)	V_c (m/min)	a_p (mm)	
1	0.02	50	0.25	78
5	0.06	50	0.5	150
9	0.1	50	0.75	258
10	0.02	70	0.25	55.71
14	0.06	70	0.5	111.42
18	0.1	70	0.75	222.85
19	0.02	90	0.25	46.66
23	0.06	90	0.5	93.33
27	0.1	90	0.75	170

Notice that for Multiple Regression models and Response Surface Method, the 27 examples are used.

IV. RESULTS

For modeling by Multiple Regression, the software used is Statgraphics, and for ANN we have used Matlab. Now, we report the mathematic formulation and the ANOVA analysis for each model.

A. Multiple Regression Models

After the introduction of experimental data into the Statgraphics software, the models provided by Multiple Regression are expressed by the following equations:

- For the Multiple Linear Regression without Intercept:

$$F_R = 1175.23\, f - 0.64225 V_c + 175.997\, a_p \quad (5)$$

- For the Nonlinear Regression:

$$F_R = 16918.1\, f^{0.496066}\, V_c^{-0.7351}\, a_p^{0.630825} \quad (6)$$

The tables III and IV summarize the ANOVA analysis.

TABLE III. ANOVA - MULTIPLE LINEAR REGRESSION

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	425749	3	141916	249.33	0.0000
Residual	13660.6	24	569.191		
Total	439410	27			

TABLE IV. ANOVA - NONLINEAR REGRESSION

Source	*Sum of Squares*	*Df*	*Mean Square*
Model	436887	4	109222
Residual	2523.04	23	109.697
Total	439410	27	
Total (Corr.)	75232	26	

B. Results of the ANN Approach

The following figure shows the elaborated ANN under Matlab Neural Network Toolbox.

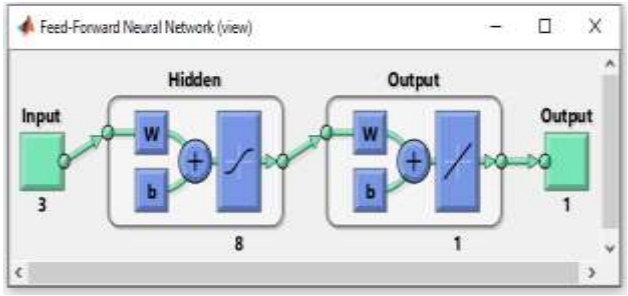


Fig. 4. Developed ANN structure under Matlab.

The values of the set of input and target vectors are normalized in the range of (-1 to 1) before training and testing the network to ensure efficient processing. The optimal design of the elaborated ANN is a multilayer feed-forward with 3-8-1 structure (Fig. 4). The Bayesian Regularization was been employed for training. We have found through numerous simulations that the choice of a single hidden layer is best. We have chosen the number of hidden neurons after various simulations as reported in table A.I. Notice that the optimal number of hidden neurons is chosen in order to obtain minimum Mean Square Error (*MSE*) and great linear regression coefficient (*R*) in the training phase.

C. Response Surface model

The Response Surface model obtained by Box-Behnken design was obtained as follows:

$$F_R = 148.783 - 260.969\, f - 1.30419 V_c - 12.43\, a_p + 2678.5\, f \cdot a_p \quad (7)$$

This model is based on 15 pairs of input-target data (see Table A.II) chosen randomly under Statgraphics software. Table V summarizes the ANOVA analysis as follows:

TABLE V. ANOVA – RESPONSE SURFACE METHOD

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
A:f	14882.4	1	14882.4	248.76	0.0000
B:Vc	5442.9	1	5442.9	90.98	0.0000
C:ap	10993.5	1	10993.5	183.75	0.0000
AC	2869.74	1	2869.74	47.97	0.0000
Total error	598.271	10	59.8271		
Total (corr.)	34786.8	14			

Note that we retained only (AC) interaction because the other interactions are insignificants as shown in Pareto chart. As shown in Fig. 5, the feed has a dominant effect on machining force; then, the depth-of-cut is taking the second position. The machining force increase when the feed and the depth-of-cut increases. Finally, the cutting speed has a lesser effect and the machining force decrease with its increase.

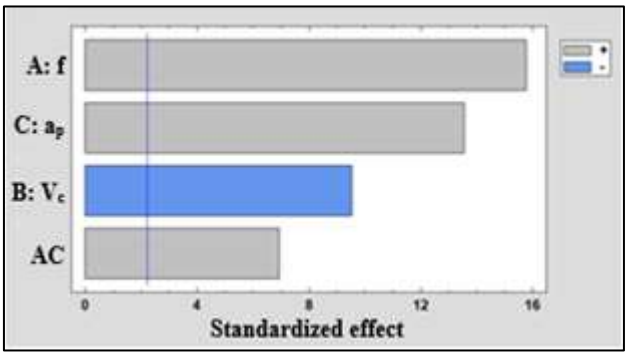


Fig. 1. Pareto chart for RSM.

V. COMPARATIVE STUDY OF THE PREDICTIVE MODELS

To evaluate the statistical performances of the predictive models, R-Squared statistic, Mean Square Error (*MSE*) and Mean Absolute Percentage Error (*MAPE*) [12] between prediction and experimental values are used.

- R-Squared statistic is given by the following equation:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{k=1}^{n}\left(c(k) - s(k)\right)^2}{\sum_{k=1}^{n}\left(c(k) - \bar{c}\right)^2}, \qquad (8)$$

where (c) and (s) are respectively the observed and predictive values of (F_R). ($\bar{c}$) and (n) are respectively the mean and the number of the observed values of (F_R). For the MLR without Intercept model, the SST is expressed by:

$$SST = \sum_{k=1}^{n}\left(c(k)\right)^2 \qquad (9)$$

- The MSE is defined by:

$$MSE = \frac{\sum_{k=1}^{n}\left(c(k) - s(k)\right)^2}{Df} \qquad (10)$$

- The MAPE formulation is given as follows:

$$MAPE(\%) = \frac{|c(k) - s(k)|}{c(k)} \qquad (11)$$

We reported in table A.III the simulation results obtained with the developed models under Statgraphics and Matlab softwares.

Table VI summarizes the performances comparison of the predictive models obtained by Multiple Regression, ANN and Response Surface Method.

TABLE VI. PERFORMANCES COMPARISON

	MLR	*NR*	*ANN*	*RSM*
R-Squared	96.89	96.64	92.44	97.29
MSE	569.191	109.697	247.25	92.64
MAPE	20.71	8.00	7.99	7.25

We can see from the last table that the Response Surface Method gives the best performances prediction of the machining force.

VI. CONCLUSIONS

The objective of this work is to develop a powerful model for machining force prediction in dry turning of A60 steel with carbide tool. We performed a comparative study of Multiple Regression, ANN and Response Surface Method for machining force. The predictive models use an experimental machining dataset of 27 examples where the parameters (feed, cutting speed and depth-of-cut) are the inputs. To show the effectiveness of the developed models, we confronted the results of machining force prediction with experimental data. The simulation, under Statgraphics and Matlab softwares, reveals that the Response Surface Method gives a great determination coefficient (or R-Squared statistic), minimal MSE and minimal MAPE. In addition, the RSM and NR models are globally very significant, but the RSM model is the most powerful.

ACKNOWLEDGMENT

We would like to thank Mr. A. Cherouik and Mr. S. Khodjet Kesba Engineers from the Department of Mechanical Engineering of the University of Tiaret for carrying out the tests.

REFERENCES

[1] LHS. Luong and TA. Spedding, "A neural network system for predicting machining behavior," Journal of Materials Processing Technology, (52), pp. 585–591, 1995.

[2] T. Szecsi, "Cutting force modeling using artificial neural networks," Journal of Materials Processing Technology, (92-93), pp. 344-349, 1999

[3] EO. Ezugwu, DA. Fadare, J. Bonney, RB. Da Silva, and WF. Sales, "Modeling the correlation between cutting and process parameters in high-speed machining of Inconel 718 alloy using an artificial neural network," International Journal of Machine Tools and Manufacture, (45), pp.1375–1385, 2005

[4] W. Hao, X. Zhu, X. Li, and G. Turyagyenda, "Prediction of cutting force for self-propelled rotary tool using artificial neural networks," Journal of Materials Processing Technology, (180), pp. 23–29, 2006

[5] V. S. Sharma, S. Dhiman, R. Sehgal, and S. K. Sharma, "Estimation of cutting forces and surface roughness for hard turning using neural networks," Journal of Intelligent Manufacturing, 19 (4), pp. 473–483, 2008

[6] T. Özel, A. Esteves Correia, and J. Paulo Davim, "Neural network process modeling for turning of steel parts using conventional and wiper inserts," International Journal of Materials and Product Technology, 35 (Nos. ½.), pp. 246-258, 2009

[7] M. Madić and M. Radovavić, "Methodology of developing optimal BP-ANN model for the prediction of cutting force in turning using Early Stopping method," Facta Universitatis. Series: Mechanical Engineering, 9 (1), pp. 21-32, 2011

[8] V. Upadhyay, P. K. Jain and N. K. Mehta, "Artificial neural network modeling of cutting force in turning of ti-6al-4v alloy and its comparison with response surface methodology," Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS), December 20-22, Advances in Intelligent and Soft Computing, (131), pp. 761-768, 2011

[9] S. Makhfi, R. Velasco, M. Habak, K. Haddouche, and P. Vantomme, "An optimized ann approach for cutting forces prediction in aisi 52100 bearing steel hard turning," Science and Technology, 3(1), pp. 24-32, 2013, dx.doi.org/10.5923/j.scit.20130301.03

[10] F. Kara, K. Aslantas, and A. Çiçek, "ANN and multiple regression method-based modeling of cutting forces in orthogonal machining of AISI 316L stainless steel," Neural Computing and Applications, 26 (Issue 1), pp. 237–250, 2015

[11] R. K. Sharma, S. Maurya, M. S. Ranganath, and Vipin, "Artificial neural network modeling for surface roughness and cutting force during conventional turning process," International Journal of Advanced Production and Industrial Engineering, 1(3), pp. 01-05, 2016

[12] S. Makhfi, K. Haddouche, A. Bourdim, and M. Habak, "Modeling of machining force in hard turning process," Mechanika, 24(3), pp. 367-375, 2018

[13] J. N. Baleo and al., "Méthodologie expérimentale : Méthodes et outils pour les expérimentations scientifiques," TEC & DOC, 2003

ANNEXES

TABLE A.I. CHOICE OF HIDDEN NEURONS NUMBER

ANN structure	*MSE*	*R*
3-2-1	2.8397e-04	0.99957
3-3-1	2.7347e-04	0.99958
3-4-1	2.0840e-04	0.99968
3-5-1	2.0978e-04	0.99968
3-6-1	2.0939e-04	0.99968
3-7-1	2.0922e-04	0.99968
3-8-1	2.0126e-04	0.99969
3-9-1	2.8016e-04	0.99957
3-10-1	2.8009e-04	0.99957

TABLE A.II. BOX-BEHNKEN DESIGN

Block	**Cutting Parameters**			***Observed***
	f **(mm/rev)**	V_c **(m/min)**	a_p ***(mm)***	F_R ***(N)***
1	0.06	70	0.5	111.42
1	0.1	70	0.25	90
1	0.02	50	0.5	102
1	0.02	70	0.25	55.71
1	0.02	90	0.5	60
1	0.06	50	0.25	102
1	0.1	70	0.75	222.85
1	0.06	70	0.5	111.42
1	0.1	90	0.5	133.33
1	0.02	70	0.75	81.42
1	0.06	50	0.75	180
1	0.1	50	0.5	198
1	0.06	90	0.25	60
1	0.06	90	0.75	120
1	0.06	70	0.5	111.42

TABLE A.III. SIMULATION RESULTS

Test n°	*Observed* F_R (N)	*Predictive* F_R *(N)*			
		MLR	***NR***	***ANN***	***RSM***
1	78	35.39	57.13	105.13	88.64
2	102	79.39	88.46	103.29	98.92
3	114	123.38	114.24	112.90	109.21
4	102	82.40	98.52	100.64	104.99
5	150	126.4	152.55	115.71	142.06
6	180	170.4	197.02	181.31	179.13
7	120	129.4	126.93	121.45	121.33
8	198	173.41	196.55	196.14	185.19
9	258	217.41	253.84	258.47	249.04
10	55.71	22.54	44.61	79.25	62.56
11	77.14	66.54	69.07	76.13	72.84
12	81.42	110.54	89.21	82.27	83.13
13	72.85	69.55	76.93	73.73	78.90
14	111.42	113.55	119.13	84.68	115.97
15	141.42	157.55	153.85	140.81	153.04
16	90	116.56	99.12	89.39	95.25
17	154.28	160.56	153.5	155.31	159.10
18	222.85	204.56	198.21	226.48	222.96
19	46.66	9.70	37.08	59.09	36.47
20	60	53.7	57.42	59.47	46.76
21	66.66	97.7	74.16	67.02	57.04
22	60	56.71	63.95	60.34	52.82
23	93.33	100.71	99.03	71.26	89.89
24	120	144.71	127.9	119.57	126.96
25	76.66	103.72	82.4	76.31	69.16
26	133.33	147.72	127.59	133.61	133.02
27	170	191.72	164.79	212.82	196.87

A Harmonic-based Fault Detection Algorithm for Microgrids

Wael Al Hanaineh
Electric Engineering Department
Polytechnic University of Catalonia
Barcelona, Spain
wael.hasan.ahmad.al.hanaineh@upc.edu

Jose Matas
Electric Engineering Department
Polytechnic University of Catalonia
Barcelona, Spain
jose.matas@upc.edu

Jorge. Elmariachet
Electric Engineering Department
Polytechnic University of Catalonia
Barcelona, Spain
jorge.el.mariachet@upc.edu

Josep.M. Guerrero
Department of Energy Technology
Aalborg University
Aalborg, Denmark
joz@et.aau.dk

***Abstract*— The trend toward Microgrids (MGs) is significantly increasing by employing Distributed Generators (DGs) which leads to new challenges, especially in the fault detection. This paper proposes an algorithm based on the Total Harmonic Distortion (THD) of the grid voltages to detect the events of faults in MGs. The algorithm uses the THD together with the estimate amplitude voltages and the zero-sequence component for the detection and identification of the faults. The performance is evaluated by using MATLAB/Simulink simulations to validate the capability for detecting different fault types in the least possible time.**

Keywords—power system faults, fault detection, total harmonics distortion, microgrids.

I. INTRODUCTION

A microgrid (MG) is a low-voltage power network with some distributed generations (DGs) and a cluster of loads that can run in parallel with the utility grid or independently. MGs are a means to increase the distributed penetration of renewable energy, such as photovoltaic and wind systems into the electrical grid. MGs have a positive impact on the environment and the economy by supplying power locally, eliminating losses in the lines, and offering continuous energy supply with improved reliability and efficiency [1].

MGs have several technical challenges related to its operational modes and characteristics, being the protection system a major issue, since should be protected from different kind of faults [2]. The magnitude and direction of fault currents in a microgrid change depending on the system configuration, due to the bi-directional power flow from the loads and the generators passing through the protective devices (PDs) [3]. The grid is the source of the majority of faults during grid-connected mode of operation, which results in very large fault currents. However, for islanding mode operation, the faulted currents are much smaller, due to the power limitations of semiconductor devices, which could be not enough to trigger a breaker [4]. Under such circumstances, the traditional PDs are unable to acquire enough information about the fault to detect the problems caused, which could lead to equipment instability and damage [5].

In recent years, several fault detection methods have been proposed for protecting microgrids, which can be grouped into differential, voltage, adaptive, and harmonics methods, each of them having advantages and weak points. Differential based methods are relatively simple [6], giving a high speed and sensitive fault detection response, and being unaffected by changes in the current's flow direction and magnitude. However, they have problems due to the rely on communication channels and because of imbalances and transients. Voltage based methods [7] have a good ability for preventing blackout in the system. But, they have inadequate sensitivity in grid-connected mode, and the voltage drops induced by faults can create errors. Adaptive based methods [8-10] are those in which the relay settings are automatically readjusted to be compatible with the power system conditions. However, they require a fault analysis and computation for determining the relay settings, as well as prior knowledge for network upgrades. Harmonic based methods [11-13] use the harmonic content of voltage and current for the fault protection. In [11], the Fourier transform (FFT) is used for achieving the required harmonics and, therefore, compute the THD and define the protection algorithm. However, the FFT implementation supposes a high computational burden for a digital processor. In [12], a cost-effective solution is introduced for microgrid protection using a new relay to detect and isolate the fault by injecting harmonic signals. Therefore, it acts like a directional relay with no need for any voltage transformer. Another interesting approach is presented in [13], which involves introducing a certain amount of a fifth harmonic to the fault current so that the protection device can identify the fault based on the low harmonics extracted using a digital relay with a FFT.

This paper presents a microgrid fault detection algorithm based on the measured THD levels of the grid voltages, the estimated amplitude voltages, and the zero-sequence components to detect, and identify, the faults that could happen in different locations of a MG [14]. This paper is submitted as a part of the Ph.D. of Electrical Engineering work at Polytechnic University of Catalonia. The rest of the paper is structured as follows: Section II presents the proposed detection algorithm in detail. Section III presents a MG as a case study to test the approach. Simulations are carried out to validate the performance of the proposed algorithm in section IV. Finally, the conclusion is provided in section V.

II. FAULT DETECTION ALGORITHM

Faults might occur in one or more phases of the grid to the ground, or between phases only. Then, in this paper, an algorithm is defined using three stages for fault detection. Fig. 1 depicts the block diagram of the algorithm.

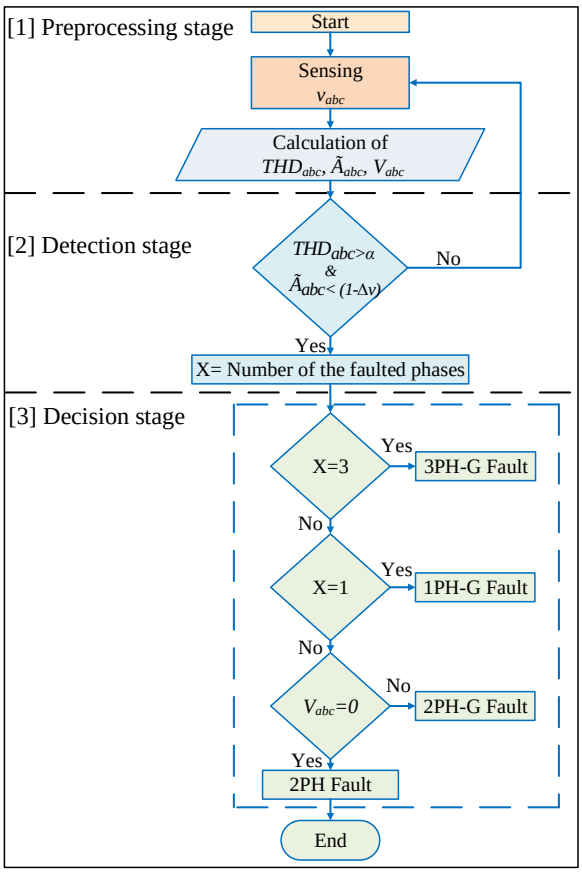


Fig. 1. Flowchart of the fault detection algorithm

A. *Pre-processing stage*

In this stage, see Fig. 1, the three-phase voltage signals, v_{abc}, are sensed in time domain at the MG to measure the THD, named as THD_{abc}, obtain an estimate of the voltage amplitudes, defined as $\tilde{A}_{abc}$, and the zero-sequence components which defined as V_{abc0}, for fault detection and identification.

The THD is computed for each phase of the grid, THD_{abc}, according to the method reported in [15] which is obtained according the standard definition of [16, 17] that uses the square root of the sum of the squared harmonic components of a given signal, divided by the fundamental component:

$$\mathrm{THD} = \frac{\sqrt{\sum_{h} |A_h|^2}}{A_1} \cdot 100\,, \qquad (1)$$

where h is the harmonic order and A_h is the amplitude of the h-th harmonic component, with $h \neq 1$, and A_1 is the amplitude of the fundamental component.

Fig. 2. Depicts the block diagram of the THD method, composed by few blocks: second order generalized integrator (SOGI) grid monitoring system [18], a LPF and few math operations. The SOGI is used to provide an estimate of A_1 and the rest of harmonic components contained in the voltage signal, named as $e(t)$.

The zero-sequence voltages are used to identify between 2PH and 2PH-G, as both have the same conditions at the fault starting. The zero-sequence voltages are calculated as follows:

$$V_{abc0} = \frac{1}{3}(v_a + v_b + v_c) \qquad (2)$$

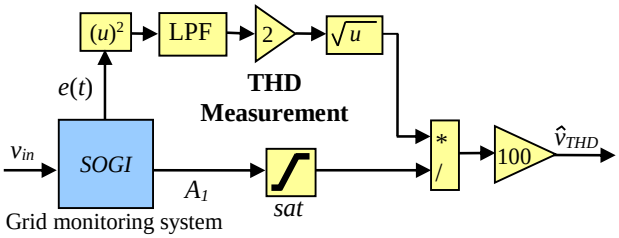


Fig. 2. THD measurement method block diagram

B. *Detection Stage*

The fault detection is made , at the middle of Fig. 1, by using a threshold comparison with the THD_{abc} voltages that had been measured in the previous stage. For the fault identification $\tilde{A}_{abc}$ and V_{abc0} are also used. The algorithm is tested in a 0.42kV MG. The IEEE standard 519-2014 provides recommended values to limit the voltage harmonic distortion to 5% [19]. And according to the technical requirements in the Spanish grid code for reliable energy integration, the acceptable grid voltage drop range at the same level is set to 7.5 % [20]. Therefore, in this paper two thresholds are defined to help in fault detection and identification:

- α to detect the fault when the THD_{abc} surpass a 5%.
- Δv to detect the fault when the estimate of the voltage amplitude $\tilde{A}_{abc}$ drops more than 7.5%.

C. *Decision Stage*

In this stage, at the lower part of Fig. 1, the detection is done based on the behavior of $\tilde{A}_{abc}$, THD_{abc}, and V_{abc0} as follows, and depending on the fault case. The faults had been classified into eleven categories numbered from 0 to 10 as shown in Table I.

TABLE I. FAULT TYPES CLASSIFICATION

Fault Case		Digital output of the algorithm
No fault		0
Single phase-to-ground fault	AG	1
	BG	2
	CG	3
Phase to phase fault	AB	4
	BC	5
	CA	6
Phase-to-phase-to-ground fault	ABG	7
	BCG	8
	CAG	9
Three phase fault	ABG	10

1) *Symmetrical faults*

These faults affect the three phases equally. The faults can happen between the three phases to ground (3PH-G) or between the three phases (3PH). In both cases, there are an abrupt

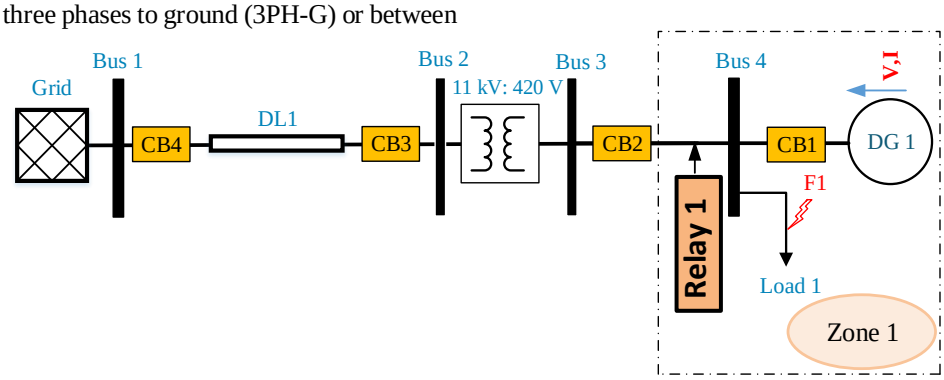


Fig. 3. Single line diagram of the studied electrical Network

increase in THD_{abc} and, at the same time, a sudden decrease in$\tilde{A}_{abc}$ to 0 pu.

2) *Unsymmetrical faults*

Unsymmetrical faults cause an imbalance between the phases, which regarding the ground can be classified into:

a) Phase-to-ground faults: These faults can occur in two of the phases to ground (2PH-G), or in only one of the phases to ground (1PH-G). In this case, there is an abrupt increase in the measured THDs and, at the same time, a sudden drop in the estimated amplitudes to 0 pu that happen at the phases affected by a fault.

b) Phase-to-phase faults: These faults can happen between two of the phases (2PH), between a-to-b, a-to-c, or b-to-c phases. Then, an abrupt increase in the THD and a sudden drop in the estimated amplitudes of two of the grid phases is produced. However, in this cases, unlike the other grounded fault cases (i.e 2PH-G), and due to the absence of the ground connection and the low impendence between the faulted phases the estimated amplitude voltages go down just to 0.5 pu. As there are no zero-sequence sources in the 2PH faults [21], then if $V_{abc0} = 0$ the fault consists in a 2PH and to 2PH-G if $V_{abc0} \neq 0$.

III. CASE STUDY

The electric system used for testing the behavior of the algorithm is shown in Fig. 3 and the parameters are listed in Table II. The system is composed by a 11kV grid with a Distribution Line (DL1), passing through a step-down transformer connected to a MG. The DL has the breakers (CB3 and CB4) that allows the disconnection in a fault event. A Distributed Generator (DG1) and local load (Load1) are

TABLE II. SYSTEM PARAMETERS

Main Grid	**MV/LV Transformer (Dyn11)**	**Distribution line DL1**		**DG1 Rating**	**Load1 Rating**
Rated voltage 11kV	Rated power 400kVA	R	0.16Ω/km	77kVA	480kW
	Rated Voltage 11/0.42kV	L	0.109H/km		
		C	0.31μF/km		

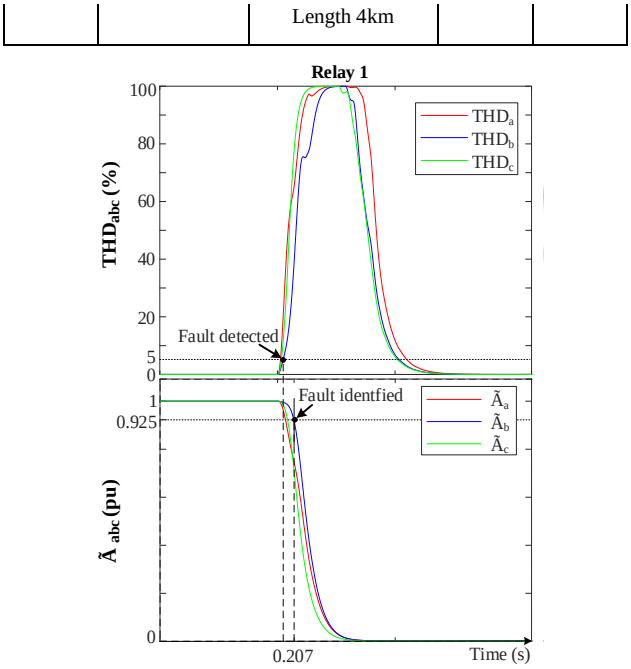


Fig. 4. Detection algorithm behavior during the 3PH-G fault at F1. Upper: THDs. Lower : estimated amplitudes $\tilde{A}$.

connected to form a MG in Zone 1, that has its own relay and breakers (Relay1, CB1 and CB2). The algorithm is defined inside the relay for fault detection.

IV. SIMULATION RESULTS

Simulations had been performed using MATLAB/Simulink at steady state to validate the performance of the detection algorithm under fault events. In this section, the fault cases are carried out in F1 location of Fig. 3 at 0.2s.

A. Three phase fault (3PH-G)

Fig. 4 shows the detection algorithm behavior during the fault. In the healthy condition, when the system operates normally before the faults, the voltages have no harmonics, so the waveforms are sinusoidal. The measured THD is 0 and the estimated amplitude voltages are 1 pu. In this condition, the algorithm is waiting for an event of fault, therefore no detection action is performed.

At 0.2s, the THD_{abc} increase abruptly, and when the condition $THD_{abc} > 5\%$ is meet the fault is detected. At the same time, $\tilde{A}_{abc}$ drops towards 0 pu and when $\tilde{A}_{abc} < 0.925$ pu, at this moment and due to the achievement of the two conditions, the fault is identified. Notice that THD_{abc} have a behaviour that

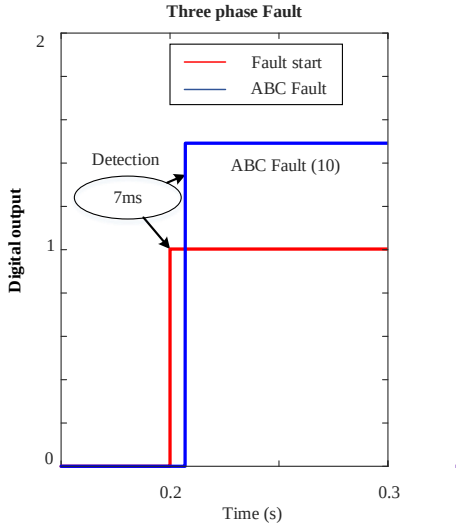


Fig. 5. Digital outputs of the detection algorithm in case of 3PH fault.

creates a peak and after a short time exponentially decays to zero. The detection process has been measured and it takes 7ms as shown in Fig. 5.

B. Phase-to-phase fault (2PH)

A fault between phases *b* and *c*, BC-fault, is considered in this case. Fig. 6 depicts the algorithm behavior during the fault. As in the previous case, at 0.2s THD_{bc} increase abruptly, which makes the fault to be detected when $THD_{bc} > 5\%$, while $THD_a < 5\%$. Meanwhile, $\tilde{A}_{bc}$ drops towards 0.5 pu, due to the absence of the ground connection and to the impedance between the phases ($Z_f = 1m\Omega$), while $\tilde{A}_a$ remains unaffected (1 pu). Then, when $\tilde{A}_{bc} < 0.925$ pu and V_{abc0} is checked to be zero, the fault is identified.

Similar to the previous case, the THD_{bc} peak behaviour exponentially decays to zero after a short time. The detection process has been measured and it consists in 7.5ms as shown in Fig. 7.

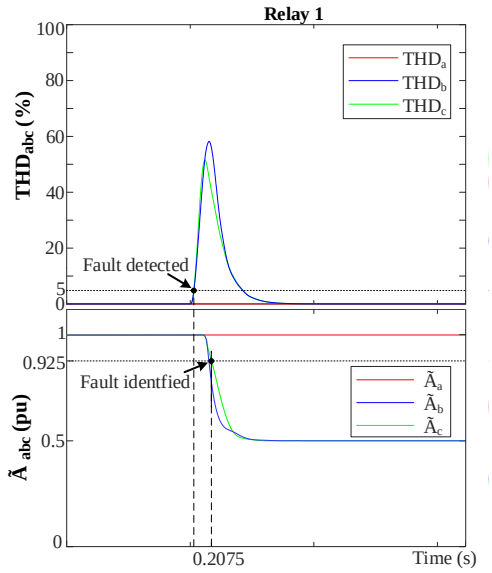


Fig. 6. Detection algorithm behavior during the 2PH fault at F1. Upper: THDs. Lower : estimated amplitudes $\tilde{A}$.

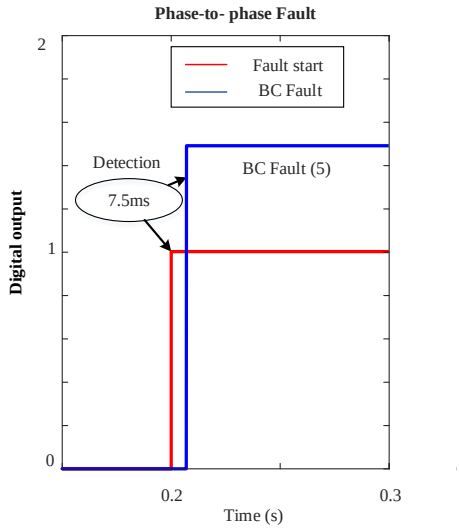


Fig. 7. Digital outputs of the detection algorithm in case of 3PH fault.

V. CONCLUSION

This paper presents a fault detection algorithm for MGs that is based on the total harmonic behavior of the grid voltages. The THD levels of the grid voltages, the estimated amplitude voltages, and the zero-sequence components have been used to design the algorithm. Each phase in the system has its own measurement block to provide the necessary data to be mentored in the algorithm.

The MATLAB/Simulink simulations results show that the algorithm has the capability to detect and identify different types of faults that might occur in the electric system in the least possible time. The detection time in all the fault cases is less than 10 ms. The method used for obtaining the THD supposes a low computational burden for being implemented in a digital signal processor.

REFERENCES

[1] R. H. Lasseter, "MicroGrids," *2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.02CH37309)*, 2002, pp. 305-308 vol.1, doi: 10.1109/PESW.2002.985003.

[2] O. S. Saadeh, W. Al-Hanaineh and Z. Dalala, "Islanding mode operation of a PV supplied network in the presence of G59 protection," *2022 IEEE 13th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, Kiel, Germany, 2022, pp. 1-5, doi: 10.1109/PEDG54999.2022.9923097.

[3] S. Mirsaeidi, D. Said, M. Mustafa, M. Habibuddin, K. Ghaffari. "An analytical literature review of the available techniques for the protection of micro-grids,". International Journal of Electrical Power & Energy Systems. 2014 Jun 1;58:300-6.

[4] X. Pei, Z. Chen, S. Wang and Y. Kang, "Overcurrent protection for inverter-based distributed generation system," *2015 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2015, pp. 2328-2332, doi: 10.1109/ECCE.2015.7309987.

[5] C. M. Colson and M. H. Nehrir, "A review of challenges to real-time power management of microgrids,"*2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1-8, doi: 10.1109/PES.2009.5275343.

[6] E. Casagrande, W. L. Woon, H. H. Zeineldin and D. Svetinovic, "A Differential Sequence Component Protection Scheme for Microgrids With Inverter-Based Distributed Generators," in *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 29-37, Jan. 2014, doi: 10.1109/TSG.2013.2251017.

[7] H. Al-Nasseri, M. A. Redfern and F. Li, "A voltage based protection for micro-grids containing power electronic converters,"*2006 IEEE Power Engineering Society General Meeting*, 2006, pp. 7 pp.-, doi: 10.1109/PES.2006.1709423.

[8] H. Wan, K. K. Li and K. P. Wong, "An Adaptive Multiagent Approach to Protection Relay Coordination With Distributed Generators in Industrial Power Distribution System," in *IEEE Transactions on Industry Applications*, vol. 46, no. 5, pp. 2118-2124, Sept.-Oct. 2010, doi: 10.1109/TIA.2010.2059492.

[9] S. M. Brahma and A. A. Girgis, "Development of adaptive protection scheme for distribution systems with high penetration of distributed generation," *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No.03CH37491)*, 2003, pp. 2083-2083, doi: 10.1109/PES.2003.1270934.

[10] V. A. Papaspiliotopoulos, G. N. Korres, V. A. Kleftakis and N. D. Hatziargyriou, "Hardware-In-the-Loop Design and Optimal Setting of Adaptive Protection Schemes for Distribution Systems With Distributed Generation," in *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 393-400, Feb. 2017, doi: 10.1109/TPWRD.2015.2509784.

[11] H. Al-Nasseri and M. A. Redfern, "Harmonics content based protection scheme for Micro-grids dominated by solid state converters," *2008 12th International Middle-East Power System Conference*, 2008, pp. 50-56, doi: 10.1109/MEPCON.2008.4562361.

[12] S. Beheshtaein, R. Cuzner, M. Savaghebi and J. M. Guerrero, "A New Harmonic-based Protection structure for Meshed Microgrids,"*2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1-6, doi: 10.1109/PESGM.2018.8585807.

[13] Z. Chen, X. Pei and L. Peng, "Harmonic components based protection strategy for inverter-interfaced AC microgrid," 2016 IEEE Energy Conversion Congress and Exposition (ECCE), 2016, pp. 1-6, doi: 10.1109/ECCE.2016.7855138.

[14] W. Al Hanaineh, J. Matas, J. Elmariachet, P. Xie, M. Bakkar, and J.M Guerrero. "A THD-Based Fault Protection Method Using MSOGI-FLL Grid Voltage Estimator," in *Sensors*, vol. 23, no. 2, pp. 980-1000, Jan. 2023. https://doi.org/ 10.3390/s23020980.

[15] J. Matas, H. Martín, J. de la Hoz, A. Abusorrah, Y. Al-Turki and H. Alshaeikh, "A New THD Measurement Method With Small Computational Burden Using a SOGI-FLL Grid Monitoring System," in IEEE Transactions on Power Electronics, vol. 35, no. 6, pp. 5797-5811, June 2020, doi: 10.1109/TPEL.2019.2953926.

[16] S. Santoso, M. McGranaghan, R. Dugan, H. Beaty, Electrical power systems quality. McGraw-Hill Education; 2012.

[17] A. Kusko, Power quality in electrical systems. McGraw-Hill Education, 2007.

[18] J. Matas, H. Martin, J. de la Hoz. A. Abusorrah, Y. Al-Turki, and M. Al-Hindawi, " A Family of Gradient Descent Grid Frequency Estimators for the SOGI Filter," *IEEE Trans. Power Electron.*, vol. pp, no. 99, 2017.

[19] "IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems," in *IEEE Std 519-2014 (Revision of IEEE Std 519-1992)*, vol., no., pp.1-29, 11 June 2014, doi: 10.1109/IEEESTD.2014. 6826459.

[20] Electrica, Red. "Technical Requirements for Wind Power and Photovoltaic Installations and Any Generating Facilities whose Technology Does Not Consist of a Synchronous Generator Directly Connected to the Grid." *Red Electrica, Madrid, Spain, Report* (2008).

[21] M. Bollen, "Understanding power quality problems." *Voltage sags and Interruptions*. Piscataway, NJ, USA: IEEE press, 2000.

.

Analysis of Structural Stiffness Reduction in Terms of Electromechanical Impedance Response for a Metallic Beam Structure Using PZT Transducer

Umakanta Meher
Department of Aerospace Engineering
Indian Institute of Technology Kharagpur
Kharagpur, India
umakanta.meher11@gmail.com

Mohammed Rabius Sunny
Department of Aerospace Engineering
Indian Institute of Technology Kharagpur
Kharagpur, India
sunny@aero.iitkgp.ac.in

***Abstract*—The present study analyzes the reduction of structural stiffness using Electro-mechanical impedance (EMI) responses. The EMI response is utilized to sense the progression of damage inside a metallic structure. At first, the EMI response of a pristine state lab-sized aluminum beam connected with a PZT transducer is obtained using E4990A impedance analyzer. The specimen is then modelled in FEM based software ANSYS and the obtained EMI response is validated with the experimental response. Various degrees of damages are incorporated in the FEM model by reduction of structural stiffness (more precisely Young's modulus) inside the structure. The damage progression status is analyzed by formation of efficient damage features like root mean square deviation (RMSD) using the EMI data for various damage scenarios. An artificial neural network (ANN) has been trained for the localization and quantification of damage status inside the structure.**

Keywords—Electro-mechanical impedance (EMI); E4990A Impedance Analyzer; Lead Zirconate Titanate (PZT); Artificial Neural Network (ANN); Root mean square deviation (RMSD); Young's Modulus (E).

I. INTRODUCTION

The EMI approach has gained wide acceptance as a potential structural health monitoring (SHM) method in many engineering applications since last two decades. Typically, a higher frequency range is applied in the EMI technique (up to 400KHz) as compared to the global dynamic technique for structural damage detection [1]. Typically, a PZT transducer is used along with a host structure which forms a feasible sensing/actuation system in EMI method. Some of the preliminary EMI analytical models of engineering structures derived by various researchers can be found in [1-5]. Meher et al. [6] proposed a data-driven frame work for multiple structural damage detection in metallic structures using both drive-point and cross EMI measurements. Quio et al. [7] combined the transfer matrix method (TMM) with the measured EMI responses to solve the inverse analysis of damage identification for a steel rod structure. Du and Wang [8] proposed a high precision probabilistic imaging method to monitor debonding in Solid Rocket Motor (SRM) case-insulator interface using EMI responses. Malinowski et al. [9] employed the principal component analysis (PCA) to assess the presence of damage inside GFRP composites applying EMI method. The feasibility of using PCA-based approach over traditional RMSD index-based scheme is shown in the study. ANN-based damage detection scheme using EMI measurements can be found in [10-12].

II. EMI METHOD

A. Experimental Measurements

The EMI technique analyzes a PZT-host structure interaction system by the application of a time- dependent harmonic voltage to the surface bonded PZT transducers within a specified frequency range. The mechanical strain generated inside the host structure due to the applied harmonic voltage is stored in the form of an electrical output signal (or the EM impedance) at the PZT transducer which can be measured by an impedance analyzer (or LCR meter). The ratio of the applied harmonic voltage to the current across the thickness of the bonded PZT transducer is known as impedance (Z) and the reciprocal of impedance (Z) is termed as admittance (Y). The real and imaginary part of the measured EM admittance are called Conductance (G) and Susceptance (B), respectively.

A lab-sized aluminum beam specimen of dimension $350mm \times 30mm \times 1.65mm$ is considered as the example structure for the present analysis. A square PZT 5H patch of dimension $15mm \times 15mm \times 0.4mm$ is fitted to the structure using epoxy adhesive. At first, the impedance measurements (refer Fig. 1) were taken using E4990A impedance analyzer within the frequency range 0-50KHz for the pristine state structure by the application of 1V harmonic voltage.

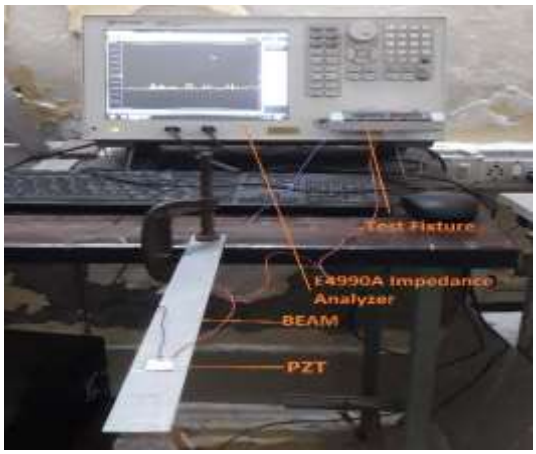


Fig. 1. Experimental Impedance Measurement for a PZT-host structure Connection

B. Numerical FEM Simulations

The lab-sized aluminum beam specimen is then modelled in FEM-based software ANSYS to obtain the EM admittance response of the structure for various damage scenarios. Reduction of structural stiffness is considered as the damage present inside the aluminum beam. At first the numerical model of the specimen was divided into six equal regions (R1-R6) along the length (refer Fig. 2). Damage inside the numerical model was incorporated by reducing the value of the Young's modulus (i.e. structural stiffness) in different regions of the structure. More precisely, the range of structural stiffness reduction is from 10% to 50% of the pristine state stiffness (refer Table I) for every region (i.e. R1-R6). So, a total of 30 damage cases along with pristine state ($S0$) was considered for the present analysis. It is to be noted that only single damage present at a time was considered for the present study. Fig. 3 shows the pristine state ANSYS numerical model of the structure. Table II consists the details of element type, mesh size etc. of the FEM analysis. The specimen was provided with a Cantilever boundary condition.The PZT material properties [6] provided by the supplier were used for numerical analysis of the PZT-structure interaction. Some of the key physical parameters of the used PZT 5H patch are listed in Table III.

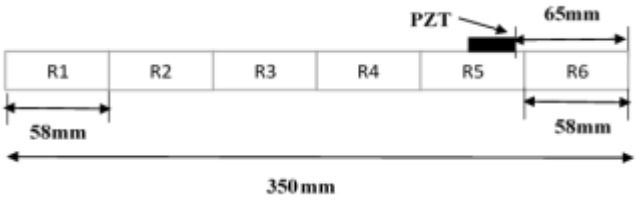


Fig. 2. Schematic of the FEM model of the structure.

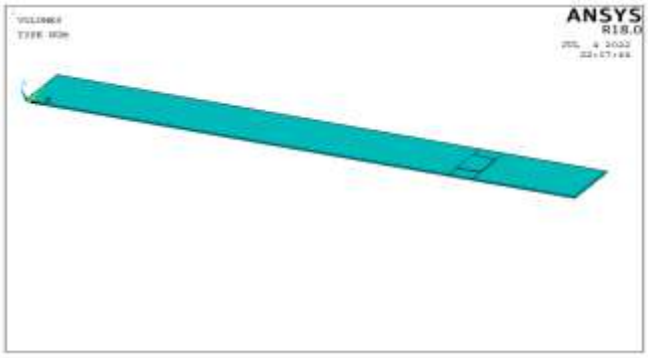


Fig. 3. Pristine State ANSYS Numerical model of the specimen

C. Damage Featutre Extraction

The EM admittance can be written in mathematical form as follows:

$$Y = G + jB \qquad (1)$$

where 'G' and 'B' denotes Conductance and susceptance, respectively.

Root mean square deviation (RMSD) of damaged state conductance w.r.t the pristine state conductance is taken as the damage indicator in the present study and is given as follows:

$$RMSD(\%) = \sqrt{\frac{\sum_{i=1}^{i=N}(G_D[i]-G_P[i])^2}{\sum_{i=1}^{i=N}(G_P[i])^2}} \times 100 \qquad (2)$$

where 'G_P' refers to pristine state conductance and 'G_D'is the damaged state conductance.

D. Damage Quantification using ANN

Input-output curve fitting method which employs Bayesian Regularization and also Lavenberg–Marquardt algorithm in 'MATLAB' environment was used to quantify the damage. Fig. 4 shows the ANN network used for the solution of the inverse problem of obtaining the damage status from EMI data.

As shown in Fig. 4, there are 20 RMSD based input and 6 target (degree of damage) values for each damage scenarios listed in Table I. The performance of the single hidden-layer feedforward back propagation ANN was measured in terms of linear regression (R) of targets relative to outputs. Out of 31 number of data sets, 27 data were used for training and 4 data were tested in the network.

TABLE I. LIST OF DAMAGE CASES CONSIDERED

Numerical Specimen Number	**Single Damage Case (Reduction in Young's Modulus (E))**					
	Region #1 (R1)	***Region #2 (R2)***	***Region #3 (R3)***	***Region #4 (R4)***	***Region #5 (R5)***	***Region #6 (R6)***
S0 (Pristine State)	E(69Gpa)	E(69Gpa)	E(69Gpa)	E(69Gpa)	E(69Gpa)	E(69Gpa)
S1	0.9E (10% reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S2	0.8E (20% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S3	0.7E (30% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S4	0.6E (40% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S5	0.5E (50% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S6	E (No reduction)	0.9E (10% reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S7	E (No reduction)	0.8E (20% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S8	E (No reduction)	0.7E (30% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S9	E (No reduction)	0.6E (40% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S10	E (No reduction)	0.5E (50% reduction	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S11	E (No reduction)	E (No reduction)	0.9E (10% reduction)	E (No reduction)	E (No reduction)	E (No reduction)
S12	E (No reduction)	E (No reduction)	0.8E (20% reduction	E (No reduction)	E (No reduction)	E (No reduction)
S13	E (No reduction)	E (No reduction)	0.7E (30% reduction	E (No reduction)	E (No reduction)	E (No reduction)
S14	E (No reduction)	E (No reduction)	0.6E (40% reduction	E (No reduction)	E (No reduction)	E (No reduction)
S15	E (No reduction)	E (No reduction)	0.5E (50% reduction	E (No reduction)	E (No reduction)	E (No reduction)
S16	E (No reduction)	E (No reduction)	E (No reduction)	0.9E (10% reduction)	E (No reduction)	E (No reduction)
S17	E (No reduction)	E (No reduction)	E (No reduction)	0.8E (20% reduction	E (No reduction)	E (No reduction)
S18	E (No reduction)	E (No reduction)	E (No reduction)	0.7E (30% reduction	E (No reduction)	E (No reduction)
S19	E (No reduction)	E (No reduction)	E (No reduction)	0.6E (40% reduction	E (No reduction)	E (No reduction)
S20	E (No reduction)	E (No reduction)	E (No reduction)	0.5E (50% reduction	E (No reduction)	E (No reduction)
S21	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.9E (10% reduction)	E (No reduction)
S22	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.8E (20% reduction	E (No reduction)
S23	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.7E (30% reduction	E (No reduction)
S24	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.6E (40% reduction	E (No reduction)
S25	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.5E (50% reduction	E (No reduction
S26	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.9E (10% reduction)
S27	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.8E (20% reduction
S28	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.7E (30% reduction
S29	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.6E (40% reduction
S30	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	E (No reduction)	0.5E (50% reduction

TABLE II. DETAILS OF ANSYS SIMULATION

Material	ELEMENT (ANSYS)	Mesh Size
Beam	SOLID 185	1mm
PZT	SOLID 5	1mm

TABLE III. Key Physical Parameters of PZT-5H patch

Sl. No	Parameter		Value
1	Piezo-electric Strain Coefficient	d_{31}	-265×10^{-12} C/N
		d_{33}	593×10^{-12} C/N
2	Relative Permitivitty at Constant Stress ($\varepsilon_{33}{}^{T}$)		3400
3	Density		7800 kg/m^3
4	Compliance	$\overline{S_{11}{}^{E}}$	16.5×10^{-12} m^2/N
		$\overline{S_{33}{}^{E}}$	20.7×10^{-12} m^2/N
5	Length of PZT patch (l)		15 $\times 10\text{^}(-3)m$
6	Width of PZT patch (w)		15 $\times 10\text{^}(-3)m$
7	Thickness of PZT patch (h)		0.4 $\times 10\text{^}(-3)m$

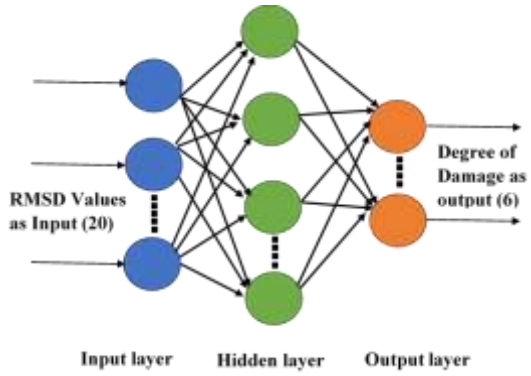


Fig. 4. ANN Architecture.

III. RESULTS AND DISCUSSIONS

The impedance measurements for the pristine state beam specimen was obtained by application of 1V harmonic voltage to the attached PZT transducer using the E4990A impedance analyzer. The FEM based impedance measurements for the pristine state test specimen as well as various damaged specimens were obtained using ANSYS Mechanical APDL. Fig. 5 shows a comparison between the experimental and FEM impedance responses for pristine state. Fig. 5 describes a good agreement between the obtained experimental and FEM admittance responses. The peaks inside the admittance spectra denotes the resonance frequency of the structure. All the major resonance peaks are successfully obtained by the FEM model (refer Fig. 5) when compared with the experiment, thus validating the FEM simulation. The peak locations and peak heights can be achieved by hit and trial of the structural stiffness and damping, respectively [1]. A need arises to update the structural properties in FEM simulations due to the unknown exact structural properties of the test specimen [1,6]. Fig. 6 gives a comparison between the obtained numerical impedance responses for different damage scenarios in region #5 (R5). Any changes inside the impedance spectra in terms of peak size, peak location or slope of the curve indicates the presence of damage/fault inside the structure. These changes are utilized to form feasible damage features for the solution of the inverse problem of finding damage status from impedance responses.

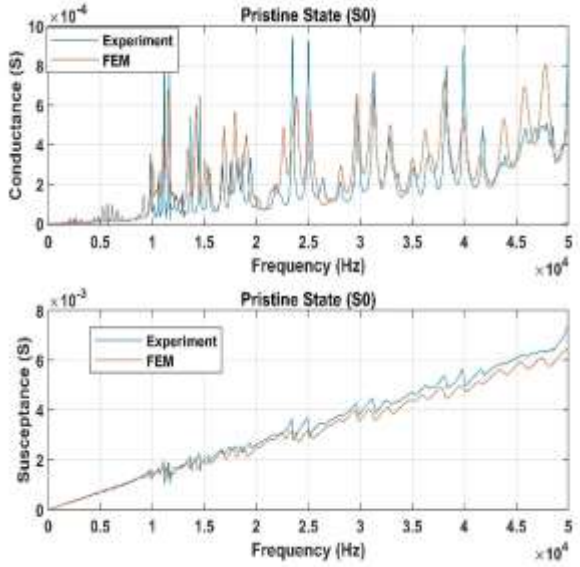
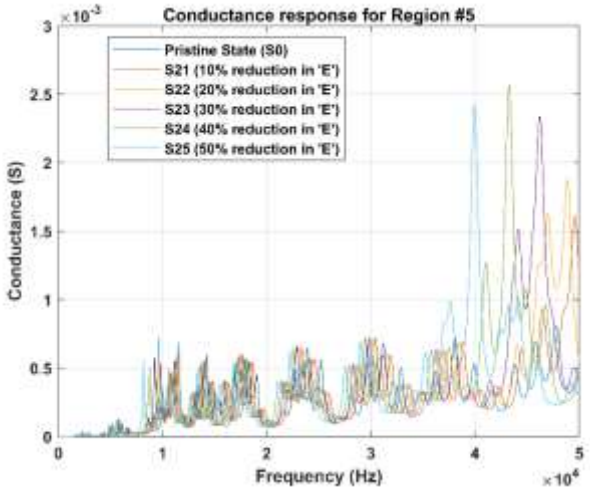
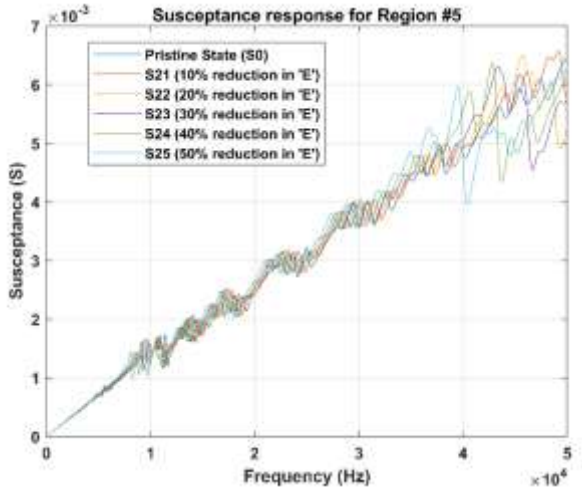


Fig. 5. Frequency Vs. Admittance Response for the Pristine State Specimen



(a)



(b)

Fig. 6. Comparison of numerical EM (a) Conductance, (b) Susceptance responses for various damage cases in region #5.

From Fig. 6 it can be seen that a clear change in impedance spectra in terms of resonance peak shifting occurs due to the presence of damage when compared with the pristine state spectra. Resonance peak shifting gives a clear sign of damage inside the structure [1]. It has been previously reported that conductance (G) is more susceptible to structural damages [1] than susceptance (B). Hence, the 20 RMSD value extracted for each and every damage case using Conductance (G) data was given as input to the ANN for the solution of the inverse problem in the present study. Fig.7 consists of the RMSD input for the damage case S21.

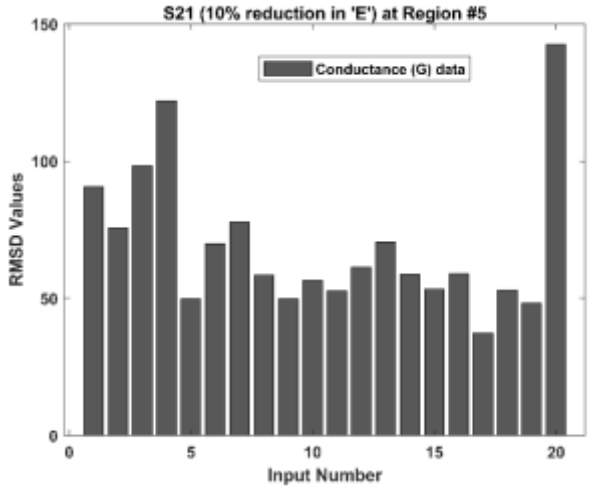


Fig. 7. RMSD values obtained from Conductance (G) for Damage Case S21.

Fig. 8 shows the efficiency of the trained ANN in terms of Regression (R). Fig. 9 and Fig. 10 gives a comparison between the actual damage present inside the structure and the damage predicted by ANN for some cases of training and testing of the network, respectively. Few false damage status (both positive and negative value) has been predicted by the employed network for both training and test (refer Fig. 9 and Fig. 10 respectively). However, these false values can be neglected as these are very small as compared to the actual damage status.

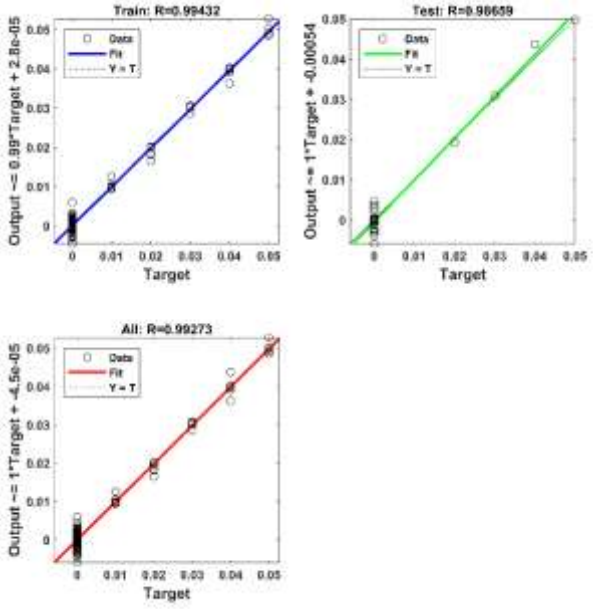


Fig. 8. Regression for actual damage and predicted damage by ANNs using RMSD based indices.

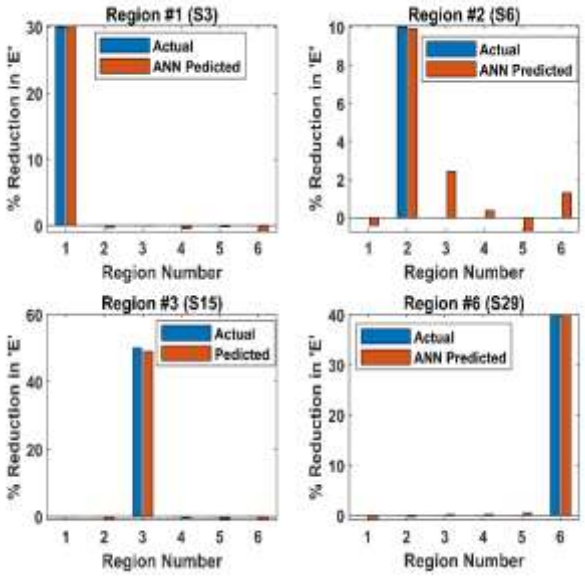


Fig. 9. Comparison of actual damage status vs. ANN predicted damage status for some cases in the training set.

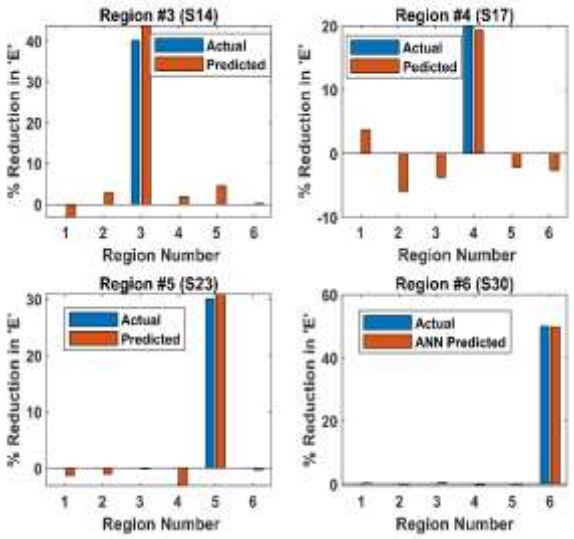


Fig. 10. Comparison of actual damage status vs. ANN predicted damage status for some cases in the test set

IV. CONCLUSIONS

The main purpose of the present work was to quantify and localize the reduction of structural stiffness inside a metallic beam using EMI responses. The successful implementation of the method (both localization and quantification) could be justified in terms of regression (R) coefficient achieved by the network during the train and test. The false damage values predicted by the network could be ignored due to their small magnitudes. The present work can be extended to detect multiple structural damages present at multiple locations at a time along the length of the beam.

ACKNOWLEDGEMENT

The authors are grateful for the support received through Grant No: IIT/SRIC/AERO/HAM/2015-16/74 from Indian Institute of Technology Kharagpur.

REFERENCES

[1] S. Bhalla, and C.K. Soh, Electro-mechanical impedance technique. In: Prof. Chee-Kiong Soh; Prof. Suresh Bhalla; Prof. Yaowen Yang (editors), ''Smart Materials in Structural Health Monitoring, Control and Biomechanics'', Zhejiang University Press, Hangzhou and Springer-Verlag Berlin Heidelberg (2012).

[2] C. Liang, F.P. Sun, and C.A. Rogers, "Coupled Electro-Mechanical Analysis of Adaptive Material Systems—Determination of the Actuator Power Consumption and System Energy Transfer", Journal of Intelligent Material Systems and Structures, 5: 12-20 (1994).

[3] S.W. Zhou, C. Liang, and C.A. Rogers, "An Impedance-Based System Modeling Approach for Induced Strain Actuator-Driven Structures", Journal of Vibrations and Acoustics, 118(3): 323-332 (1996).

[4] V.G.M. Annamdas and C.K. Soh, "Three Dimensional Electromechanical Impedance Model I: Formulation of Directional Sum Impedance", Journal of Aerospace Engineering, 20(1): 53-62 (2007).

[5] V. Giurgiutiu , C. A. Rogers , ''Modeling of electromechanical (E/M) impedance response of a damaged composite beam'', In Adaptive Structures and Material Systems Symposium, Nashville (1999).

[6] U. Meher, and M.R. Sunny, ''Detection of multiple structural damages from drive point and cross electro-mechanical impedance signatures'', '' Mechanics of Advanced Materials and Structures'', 1-21 (2021).

[7] Liang Qiao, Wei Yan and Shuyao Cao), ''Inverse analysis for damage detection in a rod using EMI method, Mechanics of Advanced Materials and Structures''(2021) DOI:10.1080/15376494.2021.2010845

[8] Du, F., Wang, G., Weng, J., Fan, H., & Xu, C. (2022). High-Precision Probabilistic Imaging for Interface Debonding Monitoring Based on Electromechanical Impedance. AIAA Journal, 1–11. https://doi.org/10.2514/1.j061577

[9] Malinowski, P. H., Wandowski, T., & Singh, S. K. (2021). Employing principal component analysis for assessment of damage in GFRP composites using electromechanical impedance. Composite Structures, 266(September 2020). https://doi.org/10.1016/j.compstruct.2021.113820

[10] J. Min, S. Park, C.B. Yun, C.G. Lee, and C. Lee, Impedancebased structural health monitoring incorporating neural network technique for identification of damage type and severity, Eng. Struct., vol. 39, pp. 210–220, 2012. DOI: 10.1016/j.engstruct.2012.01.012.

[11] Sepehry, N., Shamshirsaz, M., and Abdollahi, F., "Temperature variation effect compensation in impedance-based structural health monitoring using neural networks," Journal of Intelligent Material Systems and Structures, vol. 22, 2011, pp. 1975–1982

[12] V.Lopes, G. Park, H.H. Cudney, and D.J. Inman, Impedancebased structural health monitoring with artificial neural networks, J. Intell. Mater. Syst. Struct., vol. 11, no. 3, pp. 206–214,2000. DOI: 10.1106/H0EV-7PWM-QYHW-E7VF.

An Optimized Model Predictive Control of a Hybrid Standalone Microgrid System

Joy N. Eneh
Department of Electronic Engineering
University of Nigeria, Nsukka Enugu State.
Nsukka, Enugu State, Nigeria.
nnenna.eneh@unn.edu.ng

Solomon C. Nwafor
Department of Mechatronic Engineering
University of Nigeria, Nsukka, Enugu State.
Nsukka, Enugu State, Nigeria.
solomon.nwafor@unn.edu.ng

Oluchi C. Ugbe
Department of Electrical Engineering
University of Nigeria, Nsukka, Enugu State.
Nsukka, Enugu State, Nigeria.
oluchi.ugbe@unn.edu.ng

Timothy O. Araoye
Department of Mechatronic Engineering
University of Nigeria, Nsukka, Enugu State.
Nsukka, Enugu State, Nigeria.
timothy.araoye@unn.edu.ng

Henry I. Ugwu
Department of Electronic Engineering
University of Nigeria, Nsukka, Enugu State.
Nsukka, Enugu State, Nigeria.
henry.ugwu@unn.edu.ng

Sochima V. Egoigwe
Department of Mechatronic Engineering
University of Nigeria, Nsukka, Enugu State.
Nsukka, Enugu State, Nigeria.
sochima.egoigwe@unn.edu.ng

***Abstract*—Conventional power generation and usage are undergoing significant changes due to the integration of renewable energy sources into the power distribution network of a microgrid system. This system integrates renewable and conventional distributed generation, storage systems, and loads into a single entity that operates in isolated and grid-connected modes. First, however, energy management strategies are required. In this paper, a robust model predictive control (MPC) technique is proposed to optimize the energy management system of a standalone hybrid microgrid consisting of an energy storage system (ESS) and a diesel generator (DG) structure for tracking a reference load. The weights of the MPC are optimized using a genetic algorithm, and the modes of the ESS are optimized using convex logic. The simulation results demonstrated a reliable tracking of the reference load by the hybrid power generators. At the same time, the MPC maximized the output power of the energy storage system and minimized the use of diesel generator to eliminate emissions associated with diesel engine generators.**

Keywords— Microgrid, Renewable energy, Photovoltaic panel (PV), Diesel Generator, MPC, Genetic Algorithm

I. INTRODUCTION

Renewable energy, also known as clean energy, is derived from naturally replenished sources or processes. Renewable energy is becoming more competitive due to its environmental friendliness, according to [1]. The need for a dependable and robust system that integrates alternative energy resources is growing, particularly in isolated and remote areas where access to the national grid is limited due to technical and economic constraints [2], [3]. As a result, renewable energy systems are viewed as an appealing alternative and are thus preferred in many regions and countries. Due to the abundance of solar irradiance and wind speed, solar PV and wind are currently the two most commonly used renewable energy sources in the world today. Hybrid renewable energy systems (HRES) combine multiple renewable energy sources. According to [4], which offered an intriguing description of HRES, most hybrid systems use two or more energy methods. The system can be used off-grid or on the grid, and the energy sources can be conventional, sustainable, or a combination of both [5]. Such a system may include backup components such as a diesel generator and a battery bank to meet demand during peak hours.

It is possible to choose from a variety of hybrid renewable energy generation unit modules using commercial software [6]. [7] established Multiple Design Options (MDO) for a single-objective optimization and presented a novel method for scaling standalone hybrid energy sources using the particle swarm optimization technique. RETScreen technology is used to size a solar photovoltaic and wind turbine hybrid system, along with optimization methods, cost analysis tools, and plans for building an efficient storage system [8]. For feeding a rural village in southwest Algeria, [9] developed a hybrid system using storage, diesel generators, wind turbines, and solar photovoltaic technology. Using the HOMER PRO program for simulation and optimization, the optimal hybrid system consists of solar PV, two wind turbines, diesel engines, and lithium-ion batteries. The consumption requirements of the rural community are satisfied by this hybrid system. The four key axes of a hybrid renewable energy system—sizing, optimization, controlling the hybrid renewable energy, and managing its energy systems—are summarized globally in Ref. [10]. Ref. [11] created a supervisory power controller that divides power between an energy storage system and diesel to follow a constant and load-varying power demand profile. In

addition, MPC is utilized to regulate the power distribution for the energy storage system and diesel generator based on load demand.

The major aim of this work is to optimize the output power of the energy storage system while minimizing the use of diesel generators to reduce emissions connected with diesel engine generators. Furthermore, the MPC technique is used through simulation to optimize the error in the tracking of the reference load by the hybrid generated power and to maximize the efficient use of renewable generated power.

The following is the structure of the paper: Section II describes the microgrid (MG) system dynamics. Then, section III describes the optimized model predictive control of the MG. Finally, section IV presents the simulation results, and section V summarizes the research outcomes.

II. Microgrid Dynamic Model

Fig. 1 shows a model of a residential microgrid consisting of an energy storage system (ESS), solar generator (PV), wind turbines (WT), diesel generator (DG), and loads from homeowners. PV and WT supply energy to ESS through energy management system (EMS). While EMS schedules energy supply to the residents. The ESS consists of a battery bank, and the system operates in grid-connected mode.

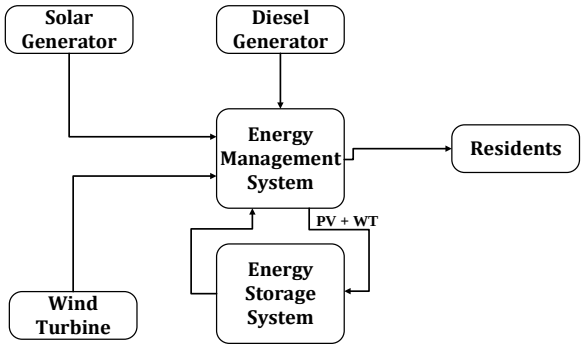


Fig. 1. Hybrid microgrid model with a standalone energy storage system and a diesel generator

A. Solar Power Model

The output power of a solar farm is determined by standard test conditions (STC), which include nominal irradiance, G_n of $1000\,W/m^2$ and ambient temperature, T_n of 25℃ [12]. The maximum power point tracking (MPPT) PV output power, P_{pv}, is defined in (1).

$$P_{pv} = \left(P_{pvs}\ \ G_t/G_n\ \ T_{pv}\right)\ N_{pv} \tag{1}$$

Where P_{pvs}, G_t , N_{pv} , and T_{pv} represent the STC-rated power, the actual irradiance, the number of solar panels, and the actual temperature, respectively. Definition of T_{pv} by [13] is specified as:

$$T_{pv} = 1 - \gamma\left(T_{pvc} - T_n\right) \tag{2}$$

γ and T_{pvc} represent the temperature coefficient of power at MPPT and the solar cell temperature, respectively. The temperature of the solar cell is defined in (3).

$$T_{pvc} = T_n + \left(\frac{NOCT - T_{pv,N}}{G_{t,N}}\right) x\ G_n \tag{3}$$

Where $NOCT$, $T_{pv,N}$, and $G_{t,N}$ are the nominal operating cell temperature, the ambient temperature of NOCT, and irradiance of NOCT, respectively.

B. Wind Turbine Power

Wind turbines (WTs) convert wind energy into electric energy. Through converters, electric energy is stored in ESS. [14] derived wind power, P_{wt} , from a relationship between aerodynamic turbine power and wind speed as follows:

$$P_{wt} = 1.571\ \rho\ R_b^2\ v_w^3\ C_p(\theta, \lambda) \tag{4}$$

ρ, R_b, v_w, C_p, θ, and λ are the air density, turbine radius, wind speed, turbine power coefficient, rotor pitch angle, and speed ratio at the rotor tip.

C. Energy Storage System

The energy storage system (ESS) is critical to the stability and reliability of the MG system. ESS also compensates for mismatches between load and power generated by the two renewable energy sources. An energy management system (EMS) is used to observe the stage of charge (SOC) and set SOC limits. The ESS capacity is constrained by an upper bound on energy, $E_{b,max}$, and a lower bound on energy, $E_{b,}$ min. Also, with a maximum charging power, $P_{b,max}$, and maximum discharging power, $-P_{b,max}$; and a charging coefficient, η_c, and a discharging coefficient, η_d, respectively.

$$0 \leq\ E_{b,min} \leq E_b \leq E_{b,max} \tag{5}$$

$$-P_{b,max} \leq P_b \leq P_{b,max} \tag{6}$$

The energy balance equation, on which the ESS dynamic model is based, evaluates an increase in ESS based on the charging/discharging efficiency of the MG as defined in (7).

$$\dot{E}_b = E_b - \frac{P_b}{E_{b,max}}\ x\ \eta\ x\ t_s \tag{7}$$

Where η, E_b, and P_b are the SOC efficiency, energy stored in the battery, and storage power, respectively. The sampling time step, t_s, assumes constant energy to-power ratio at each interval. To manage the different charging efficiency behaviors of the battery, (8) is used, and (9) represents the discharging efficiency.

$$\eta = \eta_c \tag{8}$$

$$\eta = 1/\eta_d \tag{9}$$

D. Storage System Power

The model dynamics of the storage system power, P_b, are based on the first order lag equation [15]. The difference between the input power fed to EMS, U_b and P_b describes the power dynamics of system storage power as defined in (10).

$$\dot{P}_b = \frac{1}{\tau_b}(U_b - P_b) \tag{10}$$

τ_b is the average delay incurred between U_b and P_b.

E. Diesel Generator Power

The dynamics of DG power are based on the first-order lag equation [15]. The DG is assumed to be dependent on the bounded power dynamics of the diesel engine ($P_{d,max}, P_{d,min}$) as defined in (11). As such, in accordance with [11], the difference between the input power fed to EMS, U_d and the power output of DG, P_d at a supervisory level is defined in (12).

$$-P_{d,max} \leq P_d \leq P_{d,max} \tag{11}$$

$$\dot{P}_d = \frac{1}{\tau_d}(U_d - P_d) \tag{12}$$

τ_d is the average delay incurred between U_d and P_d. The derived power equations necessitate power balance to equalize the load from the consumers, P_l, as defined in (13).

$$P_l = P_{pv} + P_{wt} + P_b + P_d \tag{13}$$

F. Load Power Reference

According to [16], the power reference for twenty-four hours displays a load reference power profile with a peak demand of 180kW [11]. The full reference power that the electric grid (EG) is required to deliver is defined in (14).

$$P_{ref} = P_l - P_r \tag{14}$$

where P_r represents renewable energy power (WT and PV).

III. MPC-BASED MICROGRID CONTROL

As shown in Fig. 2, an MPC scheme is developed to provide control input commands to the various components on EG. The MPC does not directly control the operation of renewable energies (REs), but it maximizes their utilization by tracking a reduced power reference, P_{ref}. The controller solves the MG model as a continuous linear time-invariant model and predicts future control input states with a sample time horizon period. MPC selects the optimal input control and transmits it to the EMS. The optimal MPC weights are tuned using genetic algorithm (GA).

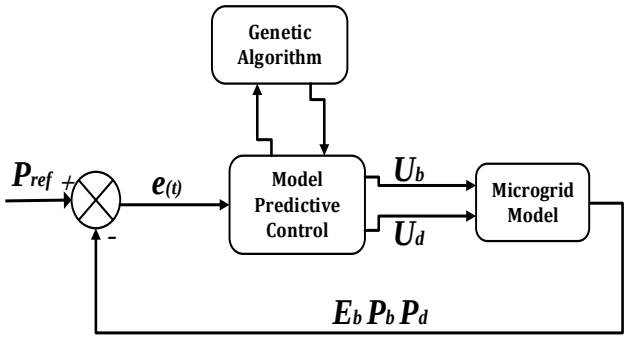


Fig. 2. Genetic algorithm optimized MPC technique for the hybrid microgrid model

A. State Space Model

As defined in (15) and (16), the power dynamics derived in (7), (10), and (12) are reformulated as a continuous linear time-invariant system with three state variables and two control inputs.

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{15}$$

$$y = Cx \tag{16}$$

C is an identity matrix. To achieve the objective of the MPC, the state space model is expanded into vector-matrix form as defined in (17) and augmented in (18).

$$\begin{bmatrix} \dot{E}_b \\ \dot{P}_b \\ \dot{P}_d \end{bmatrix} = \begin{bmatrix} 1 & -\eta/E_{b,max} & 0 \\ 0 & -1/\tau_b & 0 \\ 0 & 0 & -1/\tau_d \end{bmatrix} \begin{bmatrix} E_b \\ P_b \\ P_d \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1/\tau_b & 0 \\ 0 & 1/\tau_d \end{bmatrix} \begin{bmatrix} U_b \\ U_d \end{bmatrix} \tag{17}$$

$$\tilde{x}(t) = \tilde{A}x(t) + B\tilde{u}(t) \tag{18}$$

$\tilde{A} = v\ x\ A$ and $\tilde{u} = v\ x\ u$. Where v is the current mode of operation of the system, that is $v \in [0, 1]$.

B. MPC Performance Index

The main objective of the MPC is to maximize the use of REs while minimizing the use of DG power. Consequently, it is necessary to solve an optimal power reference tracking problem in which energy consumption from the DG is minimized. At the same time, the efficiency of the REs is maximized.

1) *Assumptions:* For the stated objective to be achieved, it is assumed that the DG, ESS charge and ESS discharge efficiencies are constant. In addition, transmission line losses are assumed to be included in the reference load.

2) *Performance Index:* The optimization problem is a quadratic cost function [11] in (19).

$$J = \sum_{k=1}^{N_c}\left(w_1 P_d^2 + w_2\left(P_d + P_b - P_{ref,l}\right)^2\right) + \sum_{k=1}^{N_p} w_3\left(E_b - E_{b,n}\right)^2 + \varphi(x) \quad (19)$$

Where $E_{b,n}$ is the initial state of E_b and $\varphi(x)$ is defined in (20).

$$\varphi(x) = (x < x_l).(x < x_l)^2 + (x < x_u).(x < x_u)^2 \quad (20)$$

x_l and x_u represent the minimum and maximum state variable values, respectively.

3) *Mode Operation:* The MPC depend on the v described in (18) to obtain an optimal control sequence using two optimization algorithms [11]. The first algorithm generates v as a continuous value between zero and one. Once a value is assigned, the mode is either 0 or 1, matching the following associated criteria:

$$if\ v \geq 0.5; \quad v = 1 \quad (21)$$

$$else\ v = 0$$

The second algorithm states that U_b sign uses the following logic to identify whether the system enters charging or discharging mode:

$$if\ U_b > 0; \quad v = 0 \quad (22)$$

$$else\ v = 1$$

C. MPC Weight Tuning

The weights of the MPC in (19) are optimized using an effective non-deterministic iterative search heuristic algorithm known as the genetic algorithm (GA). The GA employs the process of (19) to seek out the control moves that satisfy the process constraints and provide the optimal weights for the MPC [17]. Fig. 3 is a flowchart showing the execution of the GA.

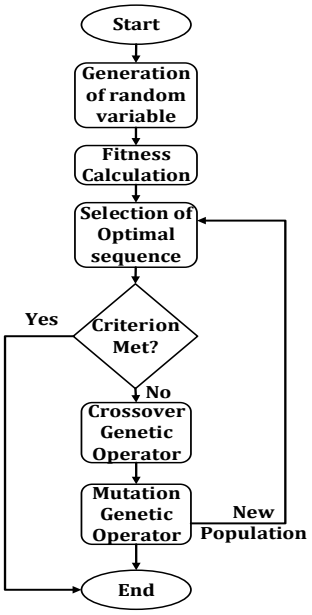


Fig. 3. Genetic algorithm flowchart for the MPC weights tuning.

IV. Results

The parameters initialized to track the power reference profile for the MG and GA models are shown in Table I and II, respectively.

TABLE III. Microgrid Model Initialized Parameters

Parameters		Values
ESS data	η_c	0.92
	η_d	0.92
	τ_b, average delay	0.2 sec
	τ_d, average delay	0.3sec
	Minimum E_b	400 kWh
	Maximum E_b	800 kWh
	Minimum P_b	-200 kW
	Maximum P_b	200 kW
DG data	Minimum P_d	0 kW
	Maximum P_d	150 kW
MPC data	MPC horizon period, N_p	12
	MPC horizon period, N_c	5

TABLE IV. Genetic Algorithm Initialized Parameters

Parameters		Values
GA data	Number of generations	75
	Population size	150
	Mutation probability	0.05
	Crossover probability	0.8
MPC weights	Tuned weights [w_1, w_2, w_3]	[4, 15, 10]

The OMPC tracks a load reference profile from [16] to evaluate the robustness of the optimized MPC (OMPC), as shown in Fig. 4. The off-peak period of the reference load ranges from 0 to 5 hours. It attains a peak load of 180 kW around the 20th hour before dropping to a stable load of 100 kW. The simulation result demonstrates that the OMPC achieved robust tracking performance by minimizing the error between the reference load and MG power states.

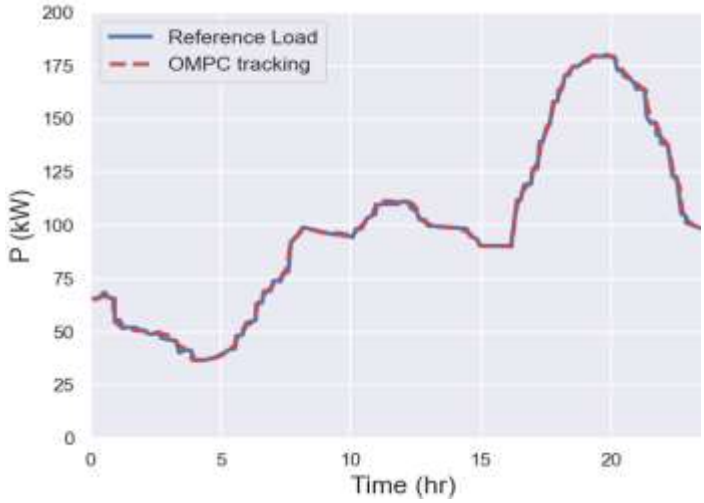


Fig. 4. GA optimized MPC load reference tracking as a function of time

Fig. 5 shows the control inputs (U_b, U_d) and mode (v) performance of the MG model. The power necessary to deliver the reference load profile is a mix of the ESS input power, shown by the blue dotted line U_b, and the DG input power, denoted by the green dotted line U_d. The mode profile, shown by the purple dashed line v, controls the ESS charging and discharging power switching.

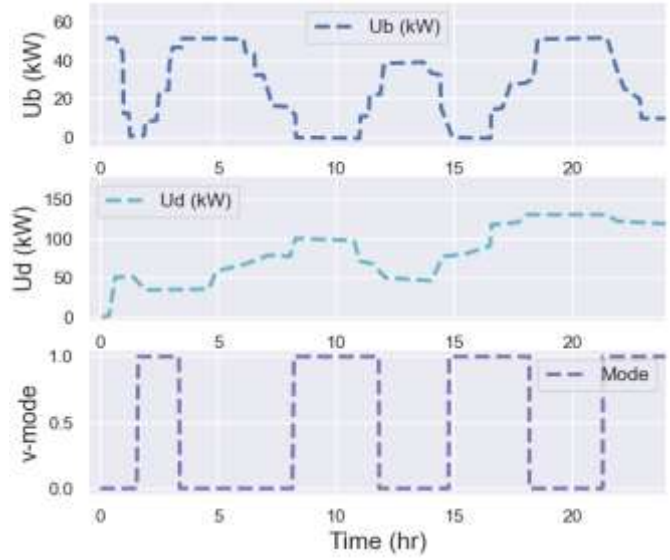


Fig. 5. GA optimized MPC outputs and convex logic mode as a function of time

Fig. 6 shows the performance of the MG output state variables, which include the energy storage Eb (purple line), the energy storage power Pb (blue line), and the DG power (green line). According to the simulation analysis, Eb maintained an average efficiency of 0.90, and the OMPC maximized the ESS power and minimized the DG power, thereby achieving the purpose of this study.

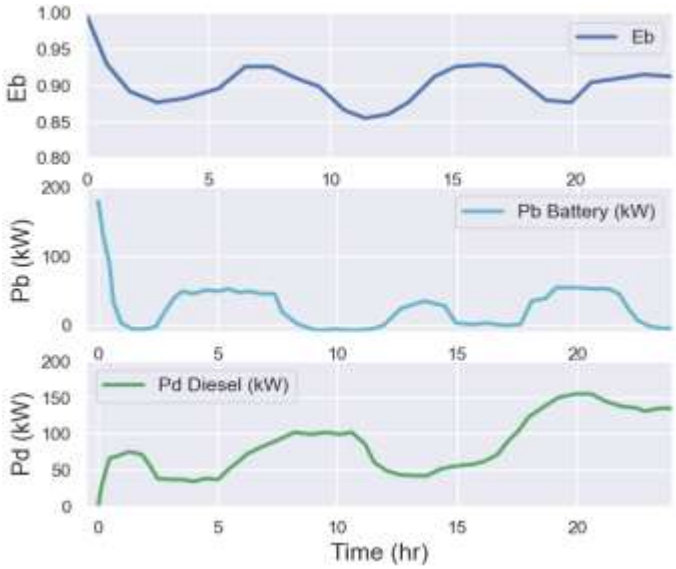


Fig. 6. Hybrid microgrid output states performance

V. Conclusion

In this paper, a microgrid dynamic model composed of energy storage, energy storage power, and diesel generator power is derived and reformulated into state space form in order to meet a required residential usage reference load. Through the use of the OMPC algorithm, this is accomplished. Furthermore, the OMPC is modeled and tuned with a genetic algorithm to optimize the amount of power transmitted to an electric grid by the two power sources.

A mode charging and discharging switch efficiency is modeled for the energy storage power source using convex logic that maximizes battery storage power and minimizes power consumption from the diesel generator. For enhanced visualization, the dynamic model and OMPC algorithm are implemented in Python and simulated with MATPLOTLIB and SEABORN. The simulation results show a robust tracking of the reference load, and the storage power is maximized. At the same time, the diesel generator power supply is minimized to eliminate the emissions associated with diesel engine generators.

References

[1] IRENA, "Renewable Power Generation Costs in 2019" https://www.irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019 (accessed: Feb. 13, 2022).

[2] NRDC, "Renewable energy, the clean fact" https://www.nrdc.org/stories/renewable-energy-clean-facts (accessed: Feb. 15, 2022)

[3] **M.** Jos´e, B. Abdul, J. Sicilia and A. Luis "Optimizing a MINLP problem for the grid-connected PV renewable energy consumption under Spanish regulations" Computers & Industrial Engineering, 2022, vol. 168 pp. 108109. https://doi.org/10.1016/j.cie.2022.108109

[4] A. Oladeji, M. Akorede, S. Aliyu, A. Mohammeda and A.Salami. "Simulation-Based Optimization of Hybrid RenewableEnergy System for Off-grid Rural Electrification" **2021**Int. Journal of Renewable Energy Development, 2021, vol. 10 no. 4 pp. 667-686. https://doi.org/10.14710/ijred.2021.31316

[5] D Belatrache, N Saifi, A Harrouz, S. Bentouba, Modelling and numerical investigation of the thermal properties effect on the soil temperature in Adrar region, Algerian J. Renew. Energy Sustain. Dev. 2020, vol. 2 no. 2 pp. 165–174. https://doi.org/10.46657/ajresd.2020.2.2.9

[6] C. Ammari, D. Belatrache, B. Touhami,and S. Makhloufi "Sizing, optimization, control and energy management of hybrid renewable energy system—A review" Energy and Built Environment. 2021 Available: https://doi.org/10.1016/j.enbenv.2021.04.002

[7] D. Fioriti, D. Poli, P. Duenas, and A. Micangeli "Multiple design options for sizing off-grid microgrids: A novel single-objective approach to support multi-criteria decision making" Sustainable Energy, Grids and Networks, 2022, vol. 30. https://doi.org/10.1016/j.segan.2022.100644

[8] F.A. Khan, N. Pal, S.H. Saeed "Review of solar photovoltaic and wind hybrid energy systems for sizing strategies optimization techniques and cost
analysis methodologies" Renewable and Sustainable Energy Reviews, 2018 vol. 92 pp. 937–947. https://doi.org/10.1016/j.rser.2018.04.107

[9] C. Ammari, H. Messaoud, and M. Salim, "Sizing and Optimization for Hybrid Central in South Algeria Based on Three Different Generators," *International Journal of Renewable Energy Development*, vol. 6, no. 3, pp. 263-272, Nov. 2017. https://doi.org/10.14710/ijred.6.3.263-272

[10] S. Ishaq, I. Khan, S. Rahman, T. Hussain, and A. Iqbal "A review on recent developments in control and optimization of micro grids" energy report 2022, vol. 8, pp. 4085-4103. https://doi.org/10.1016/j.egyr.2022.01.080

[11] N. M. Jali "Control Strategies and Hybrid Optimization Scheme for an Electric Power Grid" https://github.com/NikhilMJ/Micro-Grid-Power-Management/blob/master/Micro%20Grid%20Energy%20Optimization%20using%20MPC.pdf (accessed: March 2, 2022)

[12] IEC 60904-3, "Photovoltaic Devices - Part 3: Measurement Principles for Terrestrial Photovoltaic (PV) Solar Devices with Reference Spectral Irradiance Data" 2008.

[13] V. Q. Ngo, K. Al-Haddad and K. K. Nguyen, "Particle Swarm Optimization – Model Predictive Control for Microgrid Energy Management," *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*, 2020, pp. 264-269. http://dx.doi.org/10.1109/ZINC50678.2020.9161790

[14] C. Bordons, F. Garcia-Torres and M. A. Ridao, "Microgrid Control Issues" in Model Predictive Control of Microgrids, Springer, pp. 1-23, 2020. https://doi.org/10.1007/978-3-030-24570-2

[15] K. Uthaichana, R. DeCarlo, S. Bengea, M. Zefran and S. Pekarek, Hybrid Optimal Theory and Predictive Control for Power Management in Hybrid Electrc Vehicle. Journal of Nonlinear Systems and Applications 2011, vol. 2, no. 2, pp. 96-110. **https://doi.org/10.48550/arXiv.1804.00757**

[16] H. Wu, X. Liu, M. Ding, "Dynamic economic dispatch of a microgrid: Mathematical models and solution algorithm" Elec. Power and Energy Sys. 2014, vol. 63, pp336-46. https://doi.org/10.1016/J.IJEPES.2014.06.002

[17] N. G. Semenova, L. A. Vlatskaya and A. M. Semenov, "Application of genetic algorithms in problems of compensating devices placement optimization," *2020 International Conference on Electrotechnical Complexes and Systems (ICOECS)*, 2020, pp. 1-6, https://doi.org/10.1109/ICOECS50468.2020.9278406

Kinematic Analysis of Rat Motion After Spinal Cord Injury

Maxim E. Baltin
Research Laboratory Mechanobiology
Kazan Federal University
Kazan, Russia
ORCID: 0000-0001-5005-1699

Viktoriya V. Smirnova
Shell Mechanics Laboratory
Kazan Federal University
Kazan, Russia
ORCID: 0000-0002-1107-2152

Oskar A. Sachenkov
Department of Theoretical Mechanics
Kazan Federal University
Kazan, Russia
ORCID: 0000-0002-8554-2938

Adel E. Khairullin
Department of Biochemistry
Kazan State Medical University,
Research Laboratory Mechanobiology
Kazan Federal University
Kazan, Russia
khajrulli@ya.ru

Tatyana V. Baltina
Research Laboratory Mechanobiology
Kazan Federal University
Kazan, Russia
ORCID: 0000-0003-3798-7665

***Abstract*—To evaluate and compare the gait of rats in the control group with spinal cord injury, when treated with methylprednisolone and when using methylprednisolone in a polymer composition, a video motion analysis method was used. To assess the mobility of the limb after SCI, foot movement trajectories were constructed to determine the volume of limb movement and the maximum point of foot elevation. To calculate the volume of movement, the coordinates of the sacrum at the beginning of movement and the coordinates of the foot at the beginning of movement were determined. Our results showed the ability of rats after spinal cord injury during treatment with methylprednisolone to maintain lateral stability during locomotion. The article presents a mathematical formulation of the definition of kinematic parameters, which makes it possible to accelerate the diagnosis of the disease and individualize treatment. The presented methods are as general as possible, which allows them to be used for various experimental schemes.**

Keywords— kinematic, motion, injury

I. INTRODUCTION

The informative significance of biomechanical gait parameters has long attracted attention, since their definition is of applied importance [1,2]. The kinematic characteristics of motion allow us to describe the spatial movements of the body and its individual links in space. Moreover, using the kinematics of motion, it is possible to solve the inverse dynamic problem of finding the forces causing these movements. Motion analysis has become popular in in vivo studies [3,4]. In experimental animal models, kinematics can be studied together with other biological parameters. One of the main problems in preclinical research is the reliable evaluation of therapeutic strategies in appropriate animal models to achieve good translational effectiveness. In this regard, reading systems have been developed that track the movements of animals. The estimated parameters include walking speed, spatial and temporal indicators of step cycles (distances between paw prints, standing time, turning time), pressure and area of paw prints and rotation in the corresponding limb [5,6]. However, conventional gait and balance assessment tests may not be accurate enough to detect subtle motor disorders [7]. Kinematic gait analysis provides high-resolution data on the nature of movements and is useful for objective monitoring of various interventions. We have developed a system for assessing locomotion in a rat using video analysis of movement by parameters: the angle of flexion in the joint, the trajectory of the foot, the volume of movement of the limb, the maximum point of elevation of the foot and lateralization of the foot. The aim of the study was to assess the recovery of motor activity of rats after spinal cord injury using local delivery of methylprednisolone in combination with a copolymer.he formatter will need to create these components, incorporating the applicable criteria that follow.

II. METHODS

To evaluate and compare the gait of rats in the control group, with spinal cord injury, with treatment with methylprednisolone and when applying methylprednisolone in the polymer composition, a video motion analysis method was used. A detailed description of the injury, the characteristics of the polymer, and the application of the polymer are presented in our article [8].

Motion is recorded with six Vicon MX cameras (Vicon Motion Systems, Oxford, UK). An Active Wand calibration marker (Vicon Motion Systems, Oxford, UK) was used to calibrate and synchronize the cameras. A Sony camcorder was used to obtain a standard video image. 10 passive reflective markers were placed on the muscles of the back, sacrum, knee joints, and ankle joints (Fig. 1A). The data obtained from the cameras are used to calculate the investigated kinematic parameters. To process the obtained data, the Vicon Nexus 2.5

The work was funded by subsidy for the state assignment № 0671-2020-0059 in the sphere of scientific activities.

software was used to manually complete the 3D motion model and remove artifacts from the recording. The data received by Vicon Nexus 2.5 was converted into text format using the ASCII module and then processed using MATLAB software.

The input data is presented in the following structure (1):

$$Data_N = \{t_i, \vec{m}_i^1(x_i^1, y_i^1, z_i^1), \ldots, \vec{m}_i^N(x_i^N, y_i^N, z_i^N)\}, i\overline{(1,M)}, \quad (1)$$

where t_i- time data, $\mathbf{m}_i$ - marker data, and x_i, y_i, z_i - coordinates of marker $\mathbf{m}_i$, M - number of frames, and n - number of markers.

The X and Y axes were placed in the horizontal plane, the Z-axis was normal to the plane. The data were used to assess the articular angles [2].

The next step is apportioning the entire record into steps performed by the subject. The criterion of complete step is the contact of the toe with the support surface. It can be found by local minima by Z axis of the corresponding marker. The initial data array can be divided according to these data. So, a new dataset can be calculated (2):

$$D_j = \{i_j^{start}, i_j^{end}, Data_K\} K \in i = \overline{(1,M)}, j = \overline{(1, N_{step})} \quad (2)$$

where i^{start}, i^{end}, - indices of the beginning and end of the step, $Data_K$ - the data corresponding to the step, N_{step} - number of whole steps.

For the analysis of the step, such values as the step length, the maximum lifting height, and swing were estimated. Let's denote swing as a deviation in a plane perpendicular to the direction of movement of the subject. The mentioned parameters can be calculated for the corresponding marker. Let's denote index of the specific marker by the *X*, and the marker as m^X.

The direction of movement in a step can be calculated be Eq. (3):

$$\vec{L}_j = \left(\vec{m}_{i_j^{end}}^X - \vec{m}_{i_j^{start}}^X\right)\Big|_{z=0} \quad (3)$$

So, the length of each step can be calculated by the equation:

$$L_j = \|\vec{L}_j\| = \left\|\left(\vec{m}_{i_j^{end}}^X - \vec{m}_{i_j^{start}}^X\right)\Big|_{z=0}\right\| \quad (4)$$

To estimate the range of limb movement, the following triangle was constructed: The sacrum at the beginning of movement, the coordinates of the foot at the beginning of movement, the global maximum of the foot marker (Fig. 1(b)).

In Fig.1B the example of positions of a limb during and range of motion triangle in a step phase is shown. To quantify the range of motion of the limb the area was used. The area of a triangle can be calculated by Eq. (5):

$$S = \frac{1}{2}\left\|\left(max(\vec{m}_i^x \cdot \vec{k}_i) - \vec{m}_{i_j^{start}}^x\right)\left(\vec{m}_{i_j^{start}}^x - \vec{m}_{i_j^{start}}^y\right)\right\| \quad (5)$$

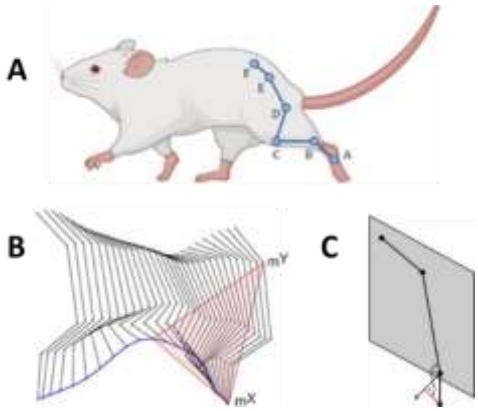


Fig. 1. Image of the position of the hind limb of rats during the step phase (a), the triangle to estimate the range of limb movement (b), vectors for calculating the lateral deviation (c).

Lateral deviation can be estimated as a deviation of the foot from the direction of movement to the side (Fig. 1(c)). To calculate the deviation, a plane through the point of the hip, knee, and ankle should be built, the plane can be determined by a normal line (Fig. 1(c)). So, the lateral deviation of the foot is the projection of the foot vector onto the normal line. For this purpose, vector connecting hip and knee (Fig. 1(a)), and the vector connecting knee and ankle (Fig. 1(a)) should be calculated and normalized (6,7):

$$\overrightarrow{BC} = \frac{\vec{C} - \vec{B}}{\|\vec{C} - \vec{B}\|} \quad (6)$$

$$\overrightarrow{CD} = \frac{\vec{D} - \vec{C}}{\|\vec{D} - \vec{C}\|} \quad (7)$$

Then the lateral deviation can be calculated by the following equation:

$$T_j = \left(\overrightarrow{BC_j} \times \overrightarrow{CD_j}\right) \cdot \vec{L}_j \quad (8)$$

III. RESULTS AND DISCUSSION

According to the representative histograms of the movement of the hind limbs, examples of which are shown in Fig. 2 (a, b, c), the volume of movement of the hind limb was determined (Fig. 2D). As can be seen from Fig. 2, the volume of movement of the hind limb in control animals was 250 ± 25 mm2.

In the group with spinal cord injury (SCI) and in the group with polymer (SCI+pol), as can be seen in Fig. 3, the volume of movement of the hind limb was 110±20 mm^2 ($p < 0.05$), the rats could not consistently reproduce the step cycle with a constant frequency, walking was interrupted, and the leg dragged along support, but there were separate two-phase steps.

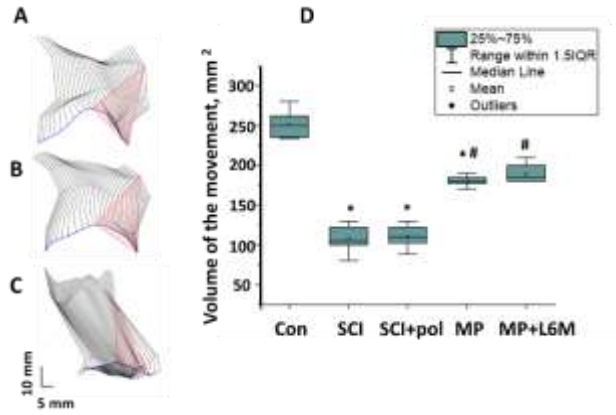


Fig. 2. Changes in the volume of movements of the hindlimb of a rat: (A; B; C) Representative image of the position of the hindlimb of rats during the step phase: the blue line represents the trajectory of the foot, the red triangle shows the volume of movements of the hindlimb; A – in the control group (without any intervention); B - after spinal cord injury; C - the group receiving methylprednisolone; (D) The volume of movement of the hind limb in experimental animals: in control (Con), with spinal cord injury (SCI), with spinal cord injury and gel application (SCI+pol), with spinal cord injury and intravenous infusion of methylprednisolone (MP), with spinal cord injury and local delivery of methylprednisolone in combinations with a copolymer (MP+L6M); * - $p < 0.05$ - the reliability of differences relative to the control group. # - $p < 0.05$ - the reliability of differences relative to the group with injury.

In the group receiving methylprednisolone intravenously during the first 48 hours after SCI (MP) and with local administration of methylprednisolone with polymer (MP+L6M) into the spinal cord after SCI, there was an improvement in the locomotion of rats, which was manifested in an increase in the volume of movement of the hind limb. The volume of movement of the hind limb was 180 ±15 mm2 ($p <0.05$) and 195 ±25 mm2 ($p <0.05$), respectively, which indicates the presence of a better therapeutic effect of treatment with methylprednisolone in combination with a copolymer after SCI, which could not be detected using the BBB test [9].

The lateral deviation of the foot is also an important factor for assessing gait. The smallest lateral deviation of the foot of the hind limbs was recorded in animals of the control group. The rat can confidently lean on the surface of the foot and maintain its weight. The lateral deviation was 1 ± 1 mm (Fig. 3).

In SCI, the rat almost did not move the hind limbs, bent the inside of the limb to the support, and thus dragged the hind limbs, the main locomotor actions were performed by the front limbs. Lateral foot deviation was 9 ± 2.7 mm. Rats with the polymer were not able to reproduce any movement of the hind limbs, the hind limb remained bent throughout the locomotor. Lateral deviation was 9 ± 4 mm ($p < 0.05$). In the MP group, the lateral deviation was 4.7 ± 2.5 mm. In the MP + L6M group, the lateral value was 2.1 ± 2 mm ($p < 0.05$).

Thus, in terms of lateral foot abnormality, the group treated with methylprednisolone and polymer showed better recovery of function, which again confirms a positive effect on the restoration of rat hind limb movements.

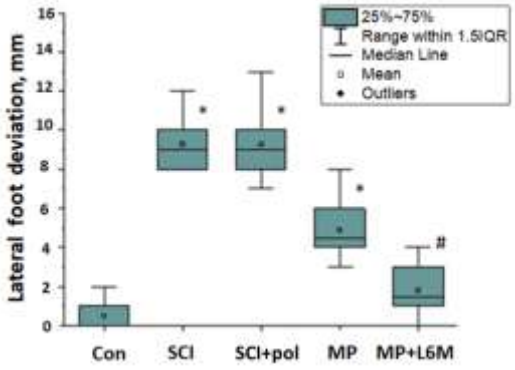


Fig. 3. Lateral foot deviation in experimental groups of animals. Designations as in Fig. 2; * - $p \leq 0.05$ - reliability relative to the control group. # - ($p \leq 0.05$) - reliability of results relative to the group with concussion injury

Both bipeds and quadrupeds require a smooth transition from standing to walking for the beginning and end of motor activity [10]. Such a transition is possible only if the neural mechanisms underlying postural and locomotor control are closely integrated [11]. SCI-induced changes in muscle coordination consist of negative (maladaptive) changes, which are characterized primarily by loss of movement in the foot after injury [12]. The rat is an adequate model for assessing changes in locomotion after injury, since it has been shown that in rats many of the secondary injuries are noted after SCI, as in humans [13]. The process of restoring the motor activity of the hindlimbs of rats after SCI is usually described using the BBB scale [9]. It is recognized that although the 21-point BBB locomotor assessment scale covers a wide range of functional recovery, it may not be sensitive to determining the exact coordination of limbs. To solve this problem, we used video analysis motion with an assessment of additional parameters to obtain information about the dynamic function of the hind limb. Only a few studies of the specific kinematics of the hindlimbs in rats have been conducted [14,15]. Our results showed that local delivery of methylprednisolone succinate in combination with a polymer in the chronic period after spinal cord injury in rats promotes the restoration of locomotion with support of body weight, control of walking direction and balance. It was shown that the rat after treatment was able to take consecutive steps in a straight line, like the animals of the control group and maintain lateral stability when moving. We can talk about the neuroprotective effect of methylprednisolone in combination with a copolymer. The proposed assessment method provides valuable information about gait in SCI. The proposed method can be successfully used to determine the quality of therapeutic interventions in treatment and rehabilitation. The neural mechanisms involved in such complex physiological phenomena of fine integration of motor and postural systems carried out at the level of the spinal cord and trunk require further study.

IV. CONCLUSION

An important goal of this study was to objectively and quantitatively analyze the effect of local delivery of methylprednisolone in combination with a copolymer on functional recovery of movement after SCI in rats. Kinematic gait analysis allows you to more accurately assess deviations

than visual assessment. Using motion capture video, we have developed a method for quantifying visually observed changes in gait biomechanics and identifying the main impact of trauma on joint kinematics and adaptation or restoration of foot movement. This evaluation method allowed us to detect significant small differences in the restoration of locomotion with the support of body weight, control of walking direction, the ability to maintain balance, as well as the configuration of the body posture when walking in paralyzed rats. These results are of great importance for quantifying the effectiveness of both regenerative and rehabilitation therapy in the treatment of neurotrauma.

ACKNOWLEDGMENT

This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

REFERENCES

[1] H. Cheng, S. Almström, L. Giménez-Llort, et al, "Gait analysis of adult paraplegic rats after spinal cord repair," Exp Neurol., vol. 148, pp. 544–557, 1997.

[2] S. Chen, J. Lach, B. Lo and G. -Z. Yang, "Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 6, pp. 1521-1537, Nov. 2016, doi: 10.1109/JBHI.2016.2608720.

[3] M. Parks, J.H. Chien, K.C. Siu, "Development of a mobile motion capture (MO2CA) system for future military application," Mil. Med., vol. 184, pp. 65–71, Mar. 2019.

[4] S. Wang, X. Zeng, L. Huangfu, et al, "Validation of a portable marker-based motion analysis system," J. Orthop. Surg. Res., vol.16, 425, 2021.

[5] D.H. Vrinten, F. F.T. Hamers, "'CatWalk' automated quantitative gait analysis as a novel method to assess mechanical allodynia in the rat; a comparison with von Frey testing," Pain, vol. 102, pp. 203-209, Mar. 2003, doi: 10.1016/s0304-3959(02)00382-2.

[6] K. Matsuda, K. Orito, Y. Amagai, H. Jang, H. Matsuda, A. Tanaka, "Swing time ratio, a new parameter of gait disturbance, for the evaluation of the severity of neuropathic pain in a rat model of partial sciatic nerve ligation," J. Pharmacol. Toxicol. Methods, vol.79, pp. 7–14, May–Jun. 2016, doi: 10.1016/j.vascn.2015.12.004.

[7] S. Gillain, E. Warzee, F. Lekeu, et al., "The value of instrumental gait analysis in elderly healthy, mci or alzheimer's disease subjects and a comparison with other clinical tests used in single and dual-task conditions," Ann. Phys. Rehabil. Med., vol.52, pp.453–474, Jul. 2009.

[8] M. E. Baltin, D. E. Sabirova, E. I. Kiseleva, et al., "Comparison of systemic and localized carrier-mediated delivery of methylprednisolone succinate for treatment of acute spinal cord injury," Experimental Brain Research, vol. 239, pp. 627–638, Feb. 2021.

[9] D.M. Basso, M.S. Beattie, J.C. Bresnahan, "A sensitive and reliable locomotor rating scale for open field testing in rats," J. Neurotrauma, vol.12, pp.1–21, Feb.1995.

[10] S. Mori, "Integration of posture and locomotion in acute decerebrate cats and awake, freely moving cats," Progress in Neurobiology, vol. 28, pp.161-195, 1987.

[11] S. Mori and K. Takakusaki, "Integration of posture and locomotion," in: Posture and Gait; Development and Modulation, A. Amblard, F. Berthoz, F. Clarac, Eds. Amsterdam: Excepta Medica, pp. 341–354, 1989.

[12] B.K. Hillen, D.L. Jindrich, J.J. Abbas, G.T. Yamaguchi, R. Jung, "Effects of spinal cord injury-induced changes in muscle activation on foot drag in a computational rat ankle model," J. Neurophysiol., vol.113, no. 7, pp. 2666-2675, Apr. 2015.

[13] S.M. Onifer, A.G. Rabchevsky, S.W.Scheff, "Rat models of traumatic spinal cord injury to assess motor recovery," ILAR J., vol.48, pp. 385–395, Oct. 2007.

[14] P.A. Couto, V.M. Filipe, L.G. Magalhaes, et al., "A comparison of two-dimensional and three-dimensional techniques for the determination of hindlimb kinematics during treadmill locomotion in rats following spinal cord injury," J. Neurosci. Methods, vol.173, pp.193–200, 2008.

[15] C.C. Diogo, L.M. da Costa, J.E. Pereira, et al., "Kinematic and kinetic gait analysis to evaluate functional recovery in thoracic spinal cord injured rats," Neurosci. Biobehav. Rev., vol.98, pp.18–28, Mar. 2019.

2-D Modeling of Transverse-Type Rectangular Piezoelectric Transformers with Common Ground Electrodes Utilizing the Legendre Polynomial Approach

Joli Randrianarivelo
LAboratoire de Physique Appliquée de l'Université de Fianarantsoa (LAPAUF)
Université de Fianarantsoa
Fianarantsoa, Madagascar
khicksjah@gmail.com

Faniry Emilson Ratolojanahary
LAboratoire de Physique Appliquée de l'Université de Fianarantsoa (LAPAUF)
Université de Fianarantsoa
Fianarantsoa, Madagascar
fratolo.rakotozafy@gmail.com

Mohamed Rguiti
Laboratoire des Matériaux Céramiques et Procédés Associés (LMCPA), Z.I. du Champ de l'Abesse Université Polytechnique Hauts-de-France (UPHF)
F-59600, Maubeuge, France
mohamed.rguiti@uph.fr

Derandraibe Jeannot Falimiaramanana
LAboratoire de Physique Appliquée de l'Université de Fianarantsoa (LAPAUF)
Université de Fianarantsoa
Fianarantsoa, Madagascar
falimiaramanana@gmail.com

Lahoucine Elmaimouni
Univ. Ibn Zohr, LSIE-ERMAM,
Polydisciplinary Faculty of Ouarzazate, BP.638, 45000 Ouarzazate, Morocco
la_elmaimouni@yahoo.fr

Ismael Naciri
Univ. Ibn Zohr, LSIE-ERMAM,
Polydisciplinary Faculty of Ouarzazate, BP.638, 45000 Ouarzazate, Morocco
nacirismail@gmail.com

***Abstract*—In this paper, a two-dimensional modeling of the transverse-type piezoelectric transformer with common ground electrodes located on its whole bottom surface is presented. Using a thin film PZT5A (lead zirconate titanate) ceramic material, a Legendre polynomial approach (PA) by applying the plane stress hypothesis is proposed taking into account the gap (length between the primary and secondary electrodes). Analytical formulations, obtained from the detailed calculation of the PA by the automatic incorporation of boundary conditions into the equations of motion, are simulated numerically. Then, series and parallel resonance frequencies, mechanical displacements, electrical potentials are obtained. In addition, electrical behaviors of the transformer are provided. Through a comparison of our results with the three-dimensional Finite Element ones (FEM), an excellent agreement is found. Furthermore, the PA permits to optimize the rate of metallization in order to improve the performance of the piezoelectric transformer.**

Keywords—Piezoelectric transformer, polynomial approach, plane-stress hypothesis, finite element method.

I. INTRODUCTION

Until now, demand of miniaturized technology equipment in electronics such as digital cameras, Smartphone and notebook computers are increasing. In micro-fabrication domain (in micro-robotic for example), requirements are conducted through the level of power and energy density, actuation forces, low mass and actuators sizes as well as the operation frequencies. Work reported in [1] has proposed that piezoelectric micro actuators integrated with flexible amplification mechanisms could balance higher step frequencies with larger ranges of motion. Preferred solution is the use of miniaturized transformers. Piezoelectric transformer (PT) is entirely dedicated to the voltage boosting [2,5-6] and has been demonstrated successfully in MOSFETs/IGBT's gate drive circuits for galvanic isolation [3]. Several advantages are provided through these devices. They exhibit a good efficiency, higher power density, higher voltage gain at resonance, low cost, simpler fabrication, no electromagnetic noise, and easier miniaturization over the conventional electromagnetic ones [4-6].

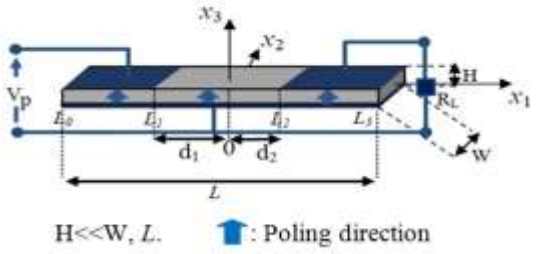
This work is motivated by ongoing challenges of PT with common ground electrodes (PTCGE) on its bottom surface commercialized in monolithic micro-fabrication as presented in figure 1. Recently, studies conducted on this type of PTs neglected the distance between the primary and secondary electrodes called "gap", using either the analytical method [5] or Mason's equivalent circuit model [6] or the Hamilton's principle [7]. These approaches experience difficulties finding the right parameters of PT for a correct operation. Moreover, the FEM is a purely numeric method, requires high storage capacity and leads to high computational time. The polynomial approach (PA) is a semi-analytical method that is at the same time analytic and numeric and has been successfully used for Rosen-type PTs modeling [8] and MEMS resonators [9-10]. Now, the PA model is applied in this paper to analyze and illustrate the free vibration modes and electrical behaviors of the transverse-type PT taking into account the gap.

The second section describes the PT structure with the boundary and continuity conditions. The third section summarizes the detailed developments of the mathematical solving equations of motion. The fourth part details the numerical solution. The fifth gives the analytical results and the sixth one, from there, the numerical simulation results. The model is validated through a comparison of the PA results with those obtained by the FEM using Comsol Multiphysics software. Many illustrations are given; the Legendre PA allows optimizing the rate of metallization to improve the performance of the studied converter. Finally, this paper will end by a brief conclusion and prospects.

II. STRUCTURE DESCRIPTION

A. PT description

Fig. 1 shows the studied PT, made of lead titanate zirconate (PZT5A) ceramic material and of class crystal hexagonal 6-mm, with *L* length, W width and H thickness. At the driving part (region 1), it is polarized along its thickness direction by an input voltage of amplitude V_p and connected to a load resistance R_L at the secondary electrodes with an output voltage V_s. The d_1+d_2 non metalized part defines the gap. PT's whole bottom surfaces are completely covered with ground electrodes and different patterns of electrodes are deposited on the upper surface to create the input and output parts [11] (separated by insulating spaces). The (x_1, x_2, x_3) coordinate system's origin is at the center of the structure where the x_3 direction coincides to the crystallographic Z-axis. We admit the temporal dependence as exp(jωt), where ω is the angular frequency and t the time, $j^2 = -1$.



re description.

Throughout this work, variables and parameters are normalized. Changes of variables are: $q_1 = 2x_1 / L$; $q_2 = 2x_2 / W$; $q_3 = 2x_3 / H$. In addition, we use the normalization systems $\bar{T}_{ij}^{(R)} = T_{ij}^{(R)} / \bar{C}_{11}^{E}$ for the stress tensor and $\bar{D}_i^{(R)} = D_i^{(R)} / \varepsilon_{33}$ for the electrical displacement, where *R* = 1, 2, 3 denotes the region number, $\bar{C}_{11}^{E} = C_{11} - C_{13}^2 / C_{33}$ and. $C_{11}, C_{13}, C_{33}, \varepsilon_{33}$ are given in table I. We assume also $E_1^{(1)} = E_1^{(3)} = E_2^{(1)} = E_2^{(3)} = 0$ for the electric field components and we normalize the driving frequency as $\Omega = \omega / \omega_p$, where $\omega_p = (\pi / 2L)\sqrt{\bar{C}_{11}^{E} / \rho}$ the one-dimensional thickness resonance angular frequency of the PT, ρ is the mass density.

B. Boundary and continuity conditions

The following assumptions are adopted: the thicknesses of electrodes are neglected; we suppose that the mechanical outer surfaces are all traction-free. That means:

$$\begin{cases} T_1^{(1)}(q_1=-1,q_2)=T_6^{(1)}(q_1=-1,q_2)=T_1^{(3)}(q_1=1,q_2)=T_6^{(3)}(q_1=1,q_2)=0 \\ T_2^{(R)}(q_1,q_2=\pm 1)=T_6^{(R)}(q_1,q_2=\pm 1)=0, \forall\, R=1,2,3. \end{cases}$$

The transformer is assumed as a thin plate structure (H⟨⟨W,*L*), then $T_3 \langle\langle T_1, T_2$; $T_3 \approx T_4 \approx T_5 \approx 0$. We adopt a 2-D modeling under the plane-stress hypothesis [8] taking into account the effect of the finite lateral dimension.

Mechanical stresses and displacements, electrical potentials and displacements are continuous at the $q_1 = q_p$ and $q_1 = q_s$ interfaces, that is, $\forall\, q_2,\ q_3$:

$$u_k^{(1)}(q_1=q_p,q_2)=u_k^{(2)}(q_1=q_p,q_2) \quad ;$$

$$u_k^{(2)}(q_1=q_s,q_2)=u_k^{(3)}(q_1=q_s,q_2);$$

$$\Phi^{(1)}(q_1=q_p,q_2,q_3)=\Phi^{(2)}(q_1=q_p,q_2,q_3) \quad ;$$

$$\Phi^{(2)}(q_1=q_s,q_2,q_3)=\Phi^{(3)}(q_1=q_s,q_2,q_3);$$

Where *u* and Φ designate the mechanical displacement and electrical potential respectively.

III. PIEZOELECTRICITY EQUATIONS

The application of an electrical field along the x_3 direction on the primary electrodes generates a mechanical vibration (acoustic) which can induce an electric energy recoverable to the secondary part. This can be translated by the propagation (1) and constitutive (2) equations in piezoelectric materials given by:

$$\begin{cases} \dfrac{\partial \bar{T}_{ij}^{S}}{\partial q_j} = -\rho\omega^2 \dfrac{L}{2\bar{C}_{11}^{E}} u_i^{S} \\ \dfrac{\partial \bar{D}_i^{S}}{\partial q_i} = 0 \end{cases} \tag{1}$$

$$\begin{cases} \bar{T}_{ij}^{(R)} = \dfrac{\bar{C}_{ijkl}'^{E}}{L}\left(\dfrac{\partial u_k^{(R)}}{\partial q_l} + \chi \dfrac{\partial u_l^{(R)}}{\partial q_k}\right) + f(\bar{e}'_{kij}/r)\dfrac{\partial \Phi^{(R)}}{\partial q_j} \\ \bar{D}_i^{(R)} = \dfrac{\bar{e}'_{ijk} r}{L}\left(\dfrac{\partial u_j^{(R)}}{\partial q_k} + \chi \dfrac{\partial u_k^{(R)}}{\partial q_j}\right) - f\bar{\varepsilon}'_{ij}\dfrac{\partial \Phi^{(R)}}{\partial q_j} \end{cases} \tag{2}$$

With $i,j,k,l = 1,2,3$; $u^S = u_i^{(1)} + u_i^{(2)} + u_i^{(3)}$. *S* defines the global structure.To automatically incorporate the boundary and continuity conditions into the equations of motion, the rectangular windows functions Π defined by:

$$\Pi^{-1,q_p}(q_1)=\begin{cases}1, if\ -1\le q_1\le q_p\\ 0,\ otherwise\end{cases};\ \Pi^{q_p,q_s}(q_1)=\begin{cases}1, if\ q_p\le q_1\le q_s\\ 0,\ otherwise\end{cases};$$
$$\Pi^{q_s,1}(q_1)=\begin{cases}1, if\ q_s\le q_1\le 1\\ 0,\ otherwise\end{cases};\ \Pi^{-1,1}(q_2)=\begin{cases}1, if\ -1\le q_2\le 1\\ 0,\ otherwise\end{cases} \quad (3)$$

are used. Then, the mechanical stress and the electrical displacement in the structure are respectively:

$$\bar{T}_{ij}^{S}=\left(\bar{T}_{ij}^{(1)}\Pi^{-1,q_p}(q_1)+\bar{T}_{ij}^{(2)}\Pi^{q_p,q_s}(q_1)+\bar{T}_{ij}^{(3)}\Pi^{q_s,1}(q_1)\right)\Pi^{-1,1}(q_2) \quad (4)$$

$$\bar{D}_{i}^{S}=\left(\bar{D}_{i}^{(1)}\Pi^{-1,q_p}(q_1)+\bar{D}_{i}^{(2)}\Pi^{q_p,q_s}(q_1)+\bar{D}_{i}^{(3)}\Pi^{q_s,1}(q_1)\right)\Pi^{-1,1}(q_2) \quad (5)$$

In our approach, the components of the mechanical displacement $u_k^{(R)}$ and the electrical potential $\Phi^{(R)}$ are developed on a basis of orthonormal polynomials adapted to the geometry given by the following expressions:

$$u_k^{(1)}(q_1,q_2)=Q_m^{(1)}(q_1)Q_n(q_2)p_{k,mn}^{(1)}; \quad (6.a)$$

$$u_k^{(2)}(q_1,q_2)=u_k^{(1)}(q_1=q_p,q_2)+(q_1-q_p)Q_m^{(2)}(q_1)Q_n(q_2)p_{k,mn}^{(2)}; \quad (6.b)$$

$$u_k^{(3)}(q_1,q_2)=u_k^{(2)}(q_1=q_s,q_2)+(q_1-q_s)Q_m^{(3)}(q_1)Q_n(q_2)p_{k,mn}^{(3)}; \quad (6.c)$$

$$\Phi^{(1)}(q_1,q_2,q_3)=\frac{V_p}{2}(q_3+1); \quad (7.a)$$

$$\Phi^{(2)}(q_1,q_2,q_3)=\frac{(q_3+1)}{2}(q_1-q_p)(q_1-q_s)Q_m^{(2)}(q_1)Q_n(q_2)r_{mn}^{(2)}$$
$$+\frac{-q_1+q_s}{q_s-q_p}\Phi^{(1)}(q_1=q_p,q_2,q_3)+\frac{q_1-q_p}{q_s-q_p}\Phi^{(3)}(q_1=q_s,q_2,q_3); \quad (7.b)$$

$$\Phi^{(3)}(q_1,q_2,q_3)=\frac{V_s}{2}(q_3+1); \quad (7.c)$$

$p_{k,mn}^{(R)}$ and $r_{mn}^{(2)}$ are the expansion coefficients and $k = 1,2$.

$$\cdot\ Q_m^{(R)}(q_1)=\sqrt{\frac{(2m+1)L}{2\left(L_R-L_{R-1}\right)}}P_m\left(\frac{q_1L}{\left(L_R-L_{R-1}\right)}-\frac{\left(L_R+L_{R-1}\right)}{\left(L_R-L_{R-1}\right)}\right);$$

$L_0=-L/2; L_3=L/2; L_1=-d_1;\ L_2=d_2; Q_n^{(R)}(q_2)=\sqrt{(2n+1)/2}\ P_n(q_2)$

P_m and P_n the Legendre polynomials of degree m and n [8].

IV. MATHEMATICAL RESOLUTION

Taking into account previous assumptions, (1) and (2) become:

$$\begin{cases}\left(\frac{2}{\chi}\delta_{i2}+\delta_{i1}\right)\sum_R\frac{\partial\bar{T}_{ij}^{(R)}}{\partial q_i}+\frac{2}{\chi}\delta_{2j}\sum_R\bar{T}_{ij}^{(R)}\left[\delta(q_j+1)-\delta(q_j-1)\right]-\\ \delta_{1j}\bar{T}_{ij}^{(3)}\left[\delta(q_j-1)-\delta(q_j+1)\right]+\delta_{i1}\left[\bar{T}_{ij}^{(2)}-\bar{T}_{ij}^{(1)}\right]\delta(q_i-q_p)+\\ \delta_{i1}\left[\bar{T}_{ij}^{(3)}-\bar{T}_{ij}^{(2)}\right]\delta(q_i-q_s)=-\left(\pi^2\Omega^2/4\right)\sum_R u_j^{(R)}\\ \frac{\partial\bar{D}_1^{(2)}}{\partial q_1}+\chi\frac{\partial\bar{D}_2^{(2)}}{\partial q_2}+\chi\bar{D}_2^{(2)}\left(\delta(q_2+1)-\delta(q_2-1)\right)-f\bar{D}_3^{(2)}\delta(q_3-1)=0\end{cases} \quad (8)$$

$$\bar{T}_1^{(R)}=\frac{2}{L}\left\{\bar{C}'_{11}\frac{\partial u_1^{(R)}}{\partial q_1}+\chi\bar{C}'_{12}\frac{\partial u_2^{(R)}}{\partial q_2}+f(\bar{e}'_{31}/r)\frac{\partial\Phi^{(R)}}{\partial q_3}\right\} \quad (9.a)$$

$$\bar{T}_2^{(R)}=\frac{2}{L}\left\{\bar{C}'_{12}\frac{\partial u_1^{(R)}}{\partial q_1}+\chi\bar{C}'_{11}\frac{\partial u_2^{(R)}}{\partial q_2}+f(\bar{e}'_{31}/r)\frac{\partial\Phi^{(R)}}{\partial q_3}\right\} \quad (9.b)$$

$$\bar{T}_6^{(R)}=\frac{2}{L}\bar{C}'_{66}\left\{\frac{\partial u_2^{(R)}}{\partial q_1}+\chi\frac{\partial u_1^{(R)}}{\partial q_2}\right\} \quad (9.c)$$

$$\bar{D}_1^{(R)}=-\frac{2}{L}\bar{\varepsilon}'_{11}\left\{\frac{\partial\Phi^{(R)}}{\partial q_1}\right\} \quad (10.a)$$

$$\bar{D}_2^{(R)}=-\frac{2}{L}\chi\bar{\varepsilon}'_{11}\left\{\frac{\partial\Phi^{(R)}}{\partial q_2}\right\} \quad (10.b)$$

$$\bar{D}_3^{(R)}=\frac{2}{L}\left\{\bar{e}'_{31}r\frac{\partial u_1^{(R)}}{\partial q_1}+\chi\bar{e}'_{31}r\frac{\partial u_2^{(R)}}{\partial q_2}-f\bar{\varepsilon}'_{33}\frac{\partial\Phi^{(R)}}{\partial q_3}\right\} \quad (10.b)$$

$\bar{C}'_{11}=(C_{11}-(C_{13}^2/C_{33}))/\bar{C}_{11}^E;\ \bar{C}'_{66}=C_{66}/\bar{C}_{11}^E;\ r=10^{-10}\sqrt{\bar{C}_{11}^E/\varepsilon_{33}};$
$\bar{C}'_{12}=(C_{12}-(C_{13}^2/C_{33}))/\bar{C}_{11}^E;\ \bar{\varepsilon}'_{11}=\varepsilon_{11}/\varepsilon_{33};$

$\bar{\varepsilon}'_{33}=1+e_{33}^2/(\varepsilon_{33}C_{33});\ \bar{e}'_{31}=(e_{31}-(C_{13}/C_{33})e_{33})/\sqrt{\varepsilon_{33}\bar{C}_{11}^E}$; $C_{12},C_{66},e_{31},e_{33},\varepsilon_{11}$ are given in table 1. $f=L/H$ and $\chi=W/H$ are the form factor and the width-thickness ratio.

The terms $\delta_{2j}\bar{T}_{ij}^{(R)}\left[\delta(q_j\pm1)\right]$; $\delta_{1j}\bar{T}_{ij}^{(3)}\left[\delta(q_j-1)\right]$; $\delta_{1j}\bar{T}_{ij}^{(1)}\left[\delta(q_j+1)\right]$, in (8) ensure that $\bar{T}_{i1}^{(1)}=\bar{T}_{i1}^{(3)}=\bar{T}_{2j}^{(R)}=0$ and $\bar{D}_2^{(2)}=0$ at the mechanically free surfaces. At the junctions $q_1=q_s$ and $q_1=q_p$, the continuity of stresses is introduced respectively by the terms $\delta_{i1}\left[\bar{T}_{ij}^{(3)}-\bar{T}_{ij}^{(2)}\right]\delta(q_i-q_s)$; $\delta_{i1}\left[\bar{T}_{ij}^{(2)}-\bar{T}_{ij}^{(1)}\right]\delta(q_i-q_p)$. In region 2, the electrical displacement component is constant along the x_3 direction. That is introduced by the $\bar{D}_3^{(2)}\delta(q_3-1)$ term. The electrical potentials are averaged all along this thickness direction.

In Eq. (8), substituting $u_k^{(R)}$ and $\Phi^{(R)}$ by their expressions given in (6) and (7) and after having multiplied by the conjugated $Q_j^*(q_1)$ and $Q_k^*(q_2)$, by integrating over q_1 from -1 to q_p, q_p to q_s, q_s to 1 and over q_2 from -1 to 1 in region R respectively, the matrix equation (11) is obtained.

$$AA*P+J*V_s=-(\pi^2\Omega^2/4)*CC*P-A*V_p \quad (11)$$

where $P=\left[p_{1,mn}^{(1)},p_{2,mn}^{(1)},p_{1,mn}^{(2)},p_{2,mn}^{(2)},p_{1,mn}^{(3)},p_{2,mn}^{(3)}\right]^T$ the unknown vector in region R. T denotes a transposed matrix. A, AA, CC, J are the matrices with m lines and n columns.

V. ANALYTICAL RESULTS

A. Harmonic analysis

Operating to its mechanical resonance, the PT undergoes different losses from mechanical, piezoelectric, and dielectric, origin due to the defects of the substrate which lead to an energy dissipation. In this manuscript, we have just considered the mechanical losses. For that, the elastic rigidities are assumed such as: $\tilde{C}^{E}_{ijkl}=\overline{C}'^{E}_{ijkl}(1+\mathrm{j}/\mathrm{Q_m})$, $\mathrm{Q_m}$ is the mechanical quality factor which produces the attenuation of the acoustic wave in the piezoelectric medium.

1. Electrical input admittance

The electrical input admittance $\mathrm{Y_p}$ is calculated from the Ampere's law [8]:

$$\mathrm{Y_p}=\frac{i_p}{\mathrm{V_p}}=\frac{LW\varepsilon_{33}}{4\mathrm{V_p}}\int_{-1}^{q_p}\int_{-1}^{1}\frac{\partial\overline{D}_3^{(1)}}{\partial t}dq_2dq_1$$

By replacing $\overline{D}_3^{(1)}$ by its expression, we have:

$$\mathrm{Y_p}=\mathrm{j}\omega C_{sp}\left[-1+by_{11mn}p_{1,mn}^{(1)}+b\chi\sqrt{(1+q_p)/2}\,y_{21mn}p_{2,mn}^{(1)}\right] \quad (12)$$

$$y_{11mn}=\left(Q_m^{(1)}(q_p)-Q_m^{(1)}(-1)\right)J1_{0n}^{-1,1,0};$$

$$y_{21mn}=J1_{0m}^{-1,q_p,0}(Q_n(1)-Q_n(-1));\ b=\chi Wr\bar{e}'_{31}\varepsilon_{33}/(\sqrt{2}C_{sp});$$

$$C_{sp}=(LW/2\mathrm{H})(1+q_p)\bar{\varepsilon}_{33};\ \bar{\varepsilon}_{33}=\varepsilon_{33}+(e_{33}{}^2/C_{33});$$

Where i_p and C_{sp} are respectively the input current and the static capacity of the primary side.

2. Output voltage

The output voltage $\mathrm{V_S}$ is given by the Ohm's law [8]:

$\mathrm{V_S}=\mathrm{R_L}i_s$; $i_s=\mathrm{j}\omega\frac{LW}{4}\varepsilon_{33}\int_{q_s}^{1}\int_{-1}^{1}\overline{D}_3^{(3)}dq_2dq_1$. Then:

$$\mathrm{V_S}=a\left\{v_{13mn}p_{1mn}^{(3)}+v_{21mn}p_{2mn}^{(1)}+v_{22mn}p_{2mn}^{(2)}+v_{23mn}p_{2mn}^{(3)}\right\} \quad (13)$$

with $a=\mathrm{W}\varepsilon_{33}/(2C_{ss}(1-\mathrm{j}Q/\Omega))$; $C_{ss}=(LW/2H)(1-q_s)\bar{\varepsilon}_{33}$ the static capacity in secondary side; $Q=1/(C_{ss}\omega_p\mathrm{R_L})$ the electrical quality factor;

$$v_{13mn}=\bar{e}'_{31}r\sqrt{2}(1-q_s)Q_m^{(3)}(1)J1_{0n}^{-1,1,0};$$

$$v_{21mn}=\bar{e}'_{31}r\chi(1-q_s)Q_m^{(1)}(q_p)(Q_n(1)-Q_n(-1));$$

$$v_{22mn}=\bar{e}'_{31}r\chi(1-q_s)(q_s-q_p)Q_m^{(2)}(q_s)(Q_n(1)-Q_n(-1));$$

$$v_{23mn}=\bar{e}'_{31}r\chi\sqrt{1-q_s}\left(J1_{0m}^{q_s,1,1}-q_sJ1_{0m}^{q_s,1,0}\right)(Q_n(1)-Q_n(-1)).$$

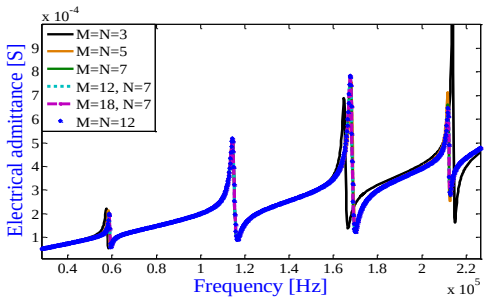


Fig. 2. Convergence of the method.

B. Free vibration analysis

By short circuiting the driving electrodes ($\mathrm{V_p}=0$), the series resonance frequencies are obtained by vanishing the load resistance ($\mathrm{R_L}$ = 0). Equation (11) then becomes:

$$\mathrm{CC}^{-1}*\mathrm{AA}*\mathrm{P}=-(\pi^2\Omega_r{}^2/4)*\mathrm{I_d}*\mathrm{P}. \quad (14)$$

Besides, we get the parallel resonance frequencies by opening the receiving electrodes ($i_s=0$) and letting the input short-circuited. From (11), we have:

$$\mathrm{CC}^{-1}*\mathrm{MM}*\mathrm{P}=-(\pi^2\Omega_a{}^2/4)*\mathrm{I_d}*\mathrm{P}. \quad (15)$$

Ω_r and Ω_a give respectively the resonance and anti-resonance frequencies. I_d is the identity matrix.

VI. NUMERICAL SIMULATIONS

First of all, material data used in the simulation are given in table I.

TABLE I. PIEZOELECTRIC CONSTANTS AND STRUCTURE DIMENSIONS

Parameters	Values
Length *L*, width W, thickness H (mm)	25 x 5 x 1.7
Length of the non-metalized part at primary d_1 and secondary d_2 (mm)	6x6
Input voltage amplitude V_p (V)	1
Mechanical quality factor Q_m	300
Mass density ρ (Kg/m^3)	7750
Vacuum dielectric permittivity ε_0 (F/m)	8.854*10^{-12}
Elastic stiffness (10^{10} N/m^2)	C_{11}=12.1; C_{12}=7.54; C_{13}=7.52 C_{33}=11.1; C_{66}=2.26
Piezoelectric constant (C/m^2)	e_{31}=–5.4; e_{33}=15.8; e_{15}=12.3
Relative permittivity (F/m)	ε_{11}=916; ε_{33}=830;

A. PA Convergence

In the numerical simulation, summation over *m* and *n* in (6) and (7) is truncated to the finite values M and N respectively. Fig. 2 shows the input admittance curves calculated in (12) for different truncation values M and N. The solutions to be accepted in the calculations are those for which the convergence is obtained when higher order terms become essentially negligible for M and N are increased [8]. Then, the convergence of the first 4 modes is obtained from the orders of truncation M = N = 7.

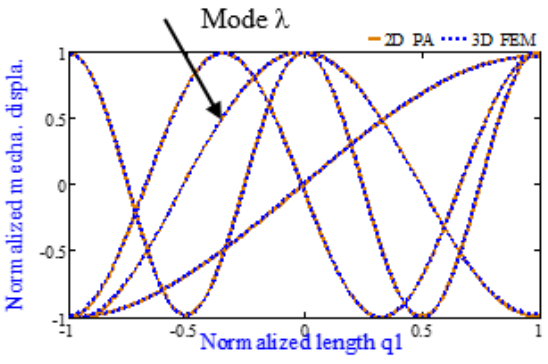


Fig. 3. Normalized mechanical displacements.

B. *Validation of the model and discussion*

The resonance f_r and anti-resonance f_a frequencies of the first four modes found with the 2D PA and 3D FEM are grouped respectively in tables II and III. The relative errors ε_f are calculated as follows:

$$\varepsilon_f = 100 * \left| \left(f_{FEM} - f_{PA} \right) / f_{FEM} \right| \quad (16)$$

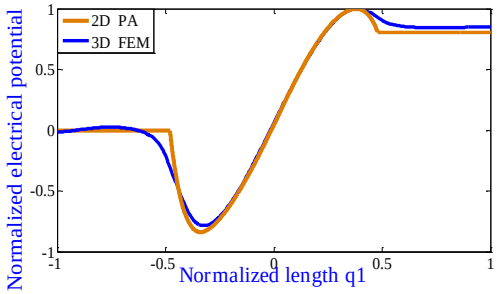


Fig. 4. Normalized electrical potentials at $q_2 = q_3 = 0$.

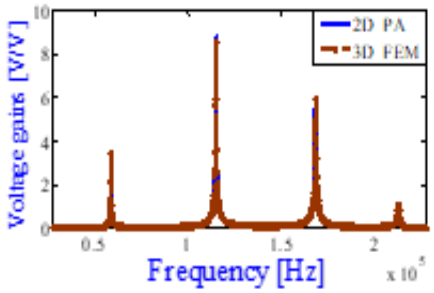


Fig. 5. Voltage gains vs frequency.

TABLE II. RESONANCE FREQUENCIES

2D $f_{r,\,PA}$ (*10^5 Hz)	3D $f_{r,\,FEM}$ (*10^5 Hz)	ε_{fr} (%)
0.58638	0.58657	0.0324
1.1340	1.1337	0.0265
1.6647	1.6622	0.1500
2.1174	2.1160	0.0661

TABLE III. ANTI-RESONANCE FREQUENCIES

2D $f_{a,\,PA}$ (*10^5 Hz)	3D $f_{a,FEM}$ (*10^5Hz)	ε_{fa} (%)
0.58999	0.59053	0.0915
1.1494	1.1498	0.0348
1.6812	1.6783	0.1700
2.1214	2.1197	0.0801

As shown in tables II and III, the agreement is good with modal relative errors well below 1%. The profiles of the mechanical displacements, electrical potentials and voltage gains are presented respectively in figures 3, 4 and 5 in opened-circuit case. Results obtained with the PA are strongly matched with those in 3D FEM.

The predicted boundary and continuity conditions are all verified through the mechanical displacement and electrical potential plots. The voltage gain as a function of the operating frequency for both methods is obtained with good accuracy below 5%. These results validate the proposed polynomial approach.

C. *Illustrations*

In order to maintain the PT's performance, it is essential to predict the expressions of the fields for locating the mechanical tethers at the nodal points [7-8] of the operating resonance mode. The expressions of the mechanical displacements are obtained from the coefficients of the vectors $p_{k,mn}^{(R)}$ given by the simulation. The mechanical tethers should apply at the points where vibrations are neglected [8]. In figure 3, these points are located at the junctions (interfaces) for the λ mode. This mode is then the adequate functioning mode so that all vibrations give some energy density for the PT.

Presented in figures 5 and 6, voltage gain, output power and efficiency depend on the frequency and R_L load located at the output electrodes. Compared to others, the second mode has higher peak gain. In addition, the PA results agree with the FEM ones (figure 6) with relative error less than 3.27%. R_L from 100 Ω to 1 MΩ gives maximum efficiency above 82 % and two maxima of power. Every maximum power corresponds to 50% efficiency for 12 mm of gap.

Although, the PT's driving frequencies are also influenced by the load parameter. As seen in figure 7, the frequency profiles of the PT for both PA/FEM models are in good agreement with a relative error of 0.34 %. They are limited by series and parallel resonance frequencies. At low load values, the behavior tends to the short-circuited output while it is to the opened circuit at high R_L values. There is an abrupt increasing of the driving frequency for a load resistance between 400 Ω to 40 MΩ. Outside this interval, there is no variation. Physically, the best performance of the PT in terms of power and efficiency only occurs within this load range.

In addition, the PT's performance depends on the metallization rate τ. Indeed, presented in figure 8, the voltage gain increases until to the 1 mm of gap and decreases again.

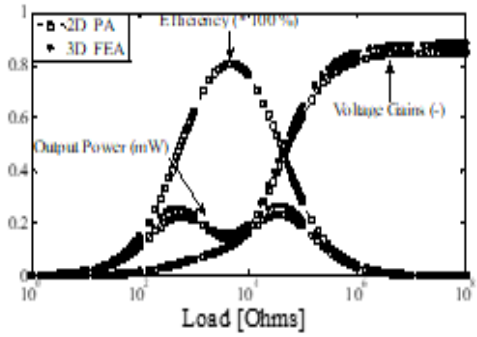


Fig. 6. Voltage gain, Output power and Efficiency vs load.

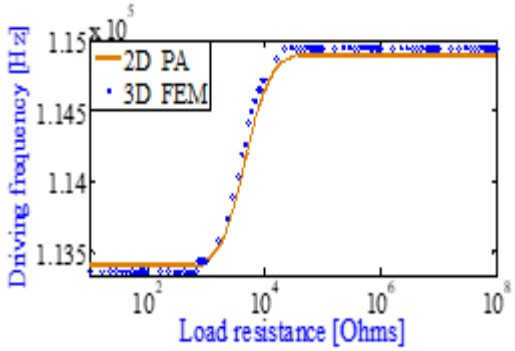


Fig. 7. Operating frequency as a function load resistance.

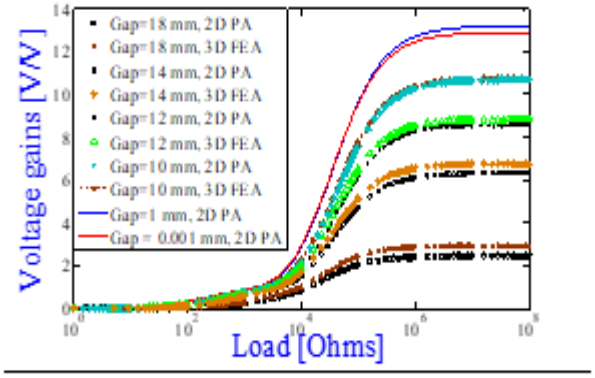


Fig. 8. Voltage gain vs load for different value of gap.

This means that there is an optimal value of gap around 1 mm to maximize the voltage gain as well as the efficiency.

VII. CONCLUSIONS

A model using the Legendre polynomial functions for modeling the transverse-type PTCGE has been reported taking into account the lateral dimension effects. Results found with the 2D PA are validated and agree very satisfactorily with the 3D FEM ones. Boundary and continuity conditions are verified. PA allows analyzing the effects of the characteristic parameters such as frequency and load connected on the output electrodes, permits to locate the mechanical tethers and to optimize the rate of metallization to improve the performance of the PT. In the future, an experimental validation will be conducted in order to confirm the validity of our method.

REFERENCES

[1] M. Shin, J. Choi, R. Q. Rudy, C. Kao, J. S. Pulskamp, R. G. Polcawich, Oldham, "Micro-Robotic actuation units based on thin-Film Piezoelectric and High-aspect ratio Polymer Structures," Proceedings of the ASME 2014 International Design Engineering Technical Conference & Computers and Information in Engineering Conference, Buffalo. New York. USA, August 2014.

[2] C. A. Rosen, "Ceramic transformers and filters," Proc. Electronic Comp. Symp., pp. 205-211, 1956.

[3] D. Vasic, F. Costa, E. Sarraute, "A new method to design Piezoelectric transformer used in MOSFET/IGBT Gate Drive Circuits, " unpublished.

[4] F. Boukazouha, F. Boubenider, Guylaine P. Vittrant, L. P. Tran-Huu-Hue, Marc L., M. Rguiti, "Parameter Determination of a Rosen Type Piezoelectric Transformer Operating in Second Mode," IEEE, 4th International Conference on Power Elgineering, Energy and Electrical Drives, Istanbul, Turkey, pp. 485-489, May 13-17, 2013.

[5] J. Yang, J. Liu, membre, IEEE and J. Li, "Analysis of a Rectangular Ceramic Plate in Electrically Forced Thickness-Twist Vibration as a Piezoelectric Transformer," IEEE transactions on ultrasonics, ferroelectrics, and frequency control, vol. 54, No. 4, pp. 830–835, April 2007.

[6] Y. Zhuang, Seyit O. Ural, R. Gosain, S. Tuncdemir, A. Amin, K. Uchino, "High Power Piezoelectric Transformers with $Pb(Mg_{1/3}Nb_{2/3})O_3$-$PbTiO_3$ Single Crystals," Appl. Phys. Express 2, vol. 121402, 2009.

[7] O. M. Barham, M. Mirzaeimoghri, and D. L. DeVoe, "Piezoelectric Disc Transformer Modeling Utilizing Extended Hamilton's Principle," IEEE, vol. 0885-8993, pp. 1-10.

[8] D. J. Falimiaramanana, F. E. Ratolojanahary, J. E. Lefebvre, L. Elmaimouni, M. Rguiti, "2D Modeling of Rosen-type piezoelectric transformer by means of a polynomial approach,"IEEE transactions on ultrasonics, ferroelectrics, and frequency control, vol. 67, no. 8, pp. 1701-1720, august 2020.

[9] A. Raherison, "Modélisation des résonateurs RF MEMS par l'approche polynomiale," thesis, p. 36, december 2009.

[10] P. M. Rabotovao, "Modélisation d'un résonateur MEMS à métallisation partielle par l'approche polynomiale". University of Fianarantsoa, Ecole Doctorale de Modélisation Informatique (EDMI), thesis, p. 41, December 2014.

[11] Petr Půlpán *, Jiří Erhart, "Transformation ratio of "ring-dot" planar piezoelectric transformer," Department of physics Center Piezoelectric Research, Technical University of Liberec, Hálkova 6, CZ-461 17 Liberec 1, Czech Republic, Sensors and Actuators, vol. A 140, pp. 215-224, June 2007.

Multimodal Calibration - a Simple Approach for Radar, 2D Camera and ToF Camera Calibration inside a Vehicle Compartment

Lap Yan Chan
Physics: Experimental sensors
Technische Universitat Chemnitz¨
Chemnitz, Germany
lap-yan.chan@physik.tu-chemnitz.de

Alessandro Zimmer
Research and Test Center CARISSMA
Technische Hochschule Ingolstadt
Ingolstadt, Germany
alessandro.Zimmer@thi.de

Luis Gustavo Tomal Ribas
Research and Test Center CARISSMA
Technische Hochschule Ingolstadt
Ingolstadt, Germany
luis.TomalRibas@carissma.eu

Ulrich Theodor Schwarz
Physics: Experimental sensors)
Technische Universitat Chemnitzt¨
Chemnitz, Germany
ulrich.schwarz@physik.tu-chemnitz.de

***Abstract*—Detection of vehicle occupants' body and health conditions such as postures, breathing rate and heartbeat rate can be used to trigger situation specific safety measures just before, during and after a car accident. The detection of occupants' conditions often involves using different sensors that are installed in different locations inside the vehicle to cover the surrounding. Especially when the calculation uses fused data from the sensors, to obtain accurate results, all sensors should be intrinsically and extrinsically calibrated. As the data structure of different sensors are usually different, and sensors are usually installed in different locations with different orientations inside a vehicle compartment, extrinsic calibration of all sensors is usually complex and challenging. In this paper, we propose a method to estimate the pose of a multimodal setup with radars, RGB cameras and Time-of-Flight (ToF) cameras inside a vehicle compartment using only position markers (so-called ArUco markers). This method is illustrated in a vehicle mock-up. At first, the markers are mapped using a high-resolution external camera and one marker is chosen to be the vehicle origin reference. The optic sensors - 2D and the 3D ToF cameras, estimate their pose over the ArUcos detected in a single shot automatically. As the radar sensors do not detect the ArUco markers as a form of image, an ArUco marker is placed on the cover of the radar instead, and the transformation from the vehicle origin to the radar is determined from this. With a calibrated sensor setup, it is possible to evaluate algorithms based on radar data and 2D images referred to the 3D groundtruth data - which is the point cloud obtained from the ToF sensor. The experiments shown a maximum error in ArUco centers position of 1.79 cm and the maximum error of radar positions is 1.9 cm.**

***Index Terms*—intrinsic calibration, extrinsic calibration, color camera, ToF camera, radar, ArUco markers**

I. INTRODUCTION

Accidents happen on the road every day but the risk can be reduced by the implementation of passive and active safety system in the vehicle. Safety systems can be designed to protect the vehicle occupants in many different ways, such as occupant monitoring for adaptive airbag deployment [1], driver's attention level detection [2] and non-contact vital signs detection [3] using radar. As the technology continue to advance, more sensors are being introduced into the vehicle compartment to collect data, and the data from different sensors are often fused together for joint analysis and robustness. In order to get precise and reliable measurement results, the sensors must be well calibrated before in order to produce meaningful information.

Calibration of sensors can be divided into intrinsic and extrinsic calibration. To achieve accurate calibration, sensors should be intrinsically calibrated, followed by extrinsic calibration. Different calibration methods for sensors have been proposed in the past [4]–[13]. These methods have their advantages and have reached a maturity stage in terms of 2D cameras, but they are less accurate when applied to calibrate Time-of-flight (ToF) cameras and radars, especially in a vehicle in-cabin environment.

The radar calibration takes a major concern in this situation: unlike calibration in an open space, calibration in a vehicle compartment has more limitations. First, there are many highly reflective objects to radar signal such as the seats, the vehicle chassis and steering wheel. This makes it more difficult to distinguish the calibration targets from the noise. Further, although the radar used in the experiment has an antenna array in the horizontal plane, there is an elevation angle (Fig. 1) between the horizontal plane and the line of sight. Meaning that a calibration target at any location within this elevation angle will be detected as a target on the horizontal plane, causing

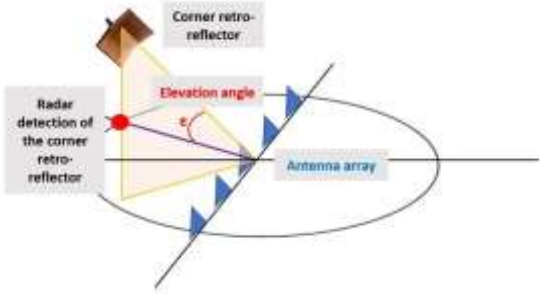


Fig. 1. Elevation angle of radar

Fig. 2. Vehicle mock-up - The sensors installed in the mock-up includes two Azure Kinect (green circles), two radars (red circles) and one fish-eye lens color camera (yellow circle)

the calibration to be less accurate. The problems of applying traditional calibration method inside vehicle compartment will be described in detail in Section II.

In this paper, we propose a calibration method for positioning a multimodal sensor setup, based on 2D cameras, 3D ToF cameras and radars inside a vehicle compartment. Our method does not require using corner retro-reflector for calibration of the radars, instead, fiducial markers as ArUco [14] were used to automatic positioning all sensors referred to a local reference in a single shot.

An ArUco [14] is a type of fiducial marker - a printed QR pattern designed for augmented reality [15], and have being used in a variety of computer vision applications [16], due to easy detection and high precision. The marker detection uses a local threshold, followed by contour detection and polygonal approximation to remove irrelevant contours. Finally, a perspective transformation is done and a Bit assignment for each cell is done, decoding a known pattern - e.g 4x4 and 40 mm of size.

Our experiment was done in a vehicle mock-up shown in Fig. 2, where ArUco markers placed around the interior of the mock-up and on the cover of the radar serves as spatial markers. The ArUcos on the interior were used to calculate the position and orientation of color camera and ToF camera with respect to the vehicle coordinate system, while the ArUco on the radars were used to calculate the position and orientation of the radars with

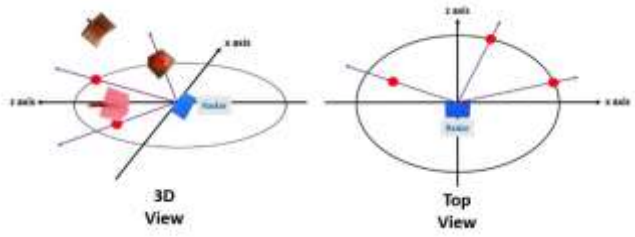


Fig. 3. three corner retro-reflectors in radar's POV (left) and their projections onto the radar's coordinate system.

respect to the vehicle coordinate system. To visualize the object detected by the radar, the CAPON beam-forming algorithm [17] was used to generate the radar detection point cloud. The calibration result was finally verified by projecting the radar detection point cloud onto a 3D space and comparing it to the result from the ToF point cloud.

II. PROBLEM

A. ELEVATION ANGLE OF RADAR ANTENNA

There are many different kinds of radars available in the market. Different radars can have different hardware configurations. Single channel radar, which has only one transmitting and receiving antenna, returns data from a single direction (1D). Multi-Input-Multi-Output (MIMO) radar has multiple transmitting and receiving antennas, depending on the arrangement of these antennas, MIMO radar can return data from two directions, range-azimuth (2D) or range-azimuth-elevation (3D).

The radar used in our experiment has a linear antenna array consist of 16 virtual antennas. Such an antenna arrangement only returns range-azimuth (2D) information for our calibration. To achieve accurate calibration for the radars, at least three common targets should be detected by both radars. However, because of the fact that elevation angle exist, a calibration target can appear in a radar's POV even if it is not located on the radar's horizontal plane (illustrated in Fig. 3).

B. DIFFICULTY TO LOCATE THREE COMMON CALIBRATION TARGETS FOR BOTH RADAR

In addition to the elevation angle of the radar, another problem with using traditional calibration method is that it is difficult to locate three common calibration targets for both radars in a vehicle compartment. Because the radars are pointing towards different directions, in most locations inside the vehicle mock-up, a retro-reflector cannot be clearly detected

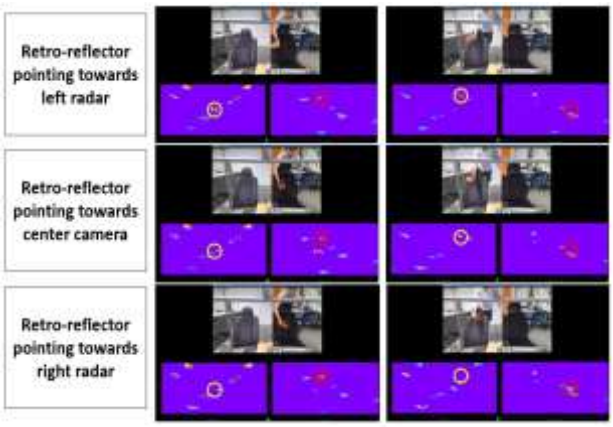


Fig. 4. CAPON detection when rector-reflector pointing at different direction. Yellow circles indicate detection of the retro-reflector by the left radar and red circles indicate detection of the retro-reflector by the right radar.

by both radar at the same time. Fig. 4 shows the CAPON detection results when the retro-reflector was pointing towards

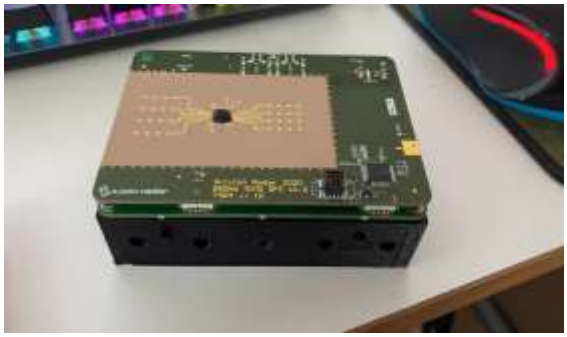


Fig. 5. 4Tx-4Rx MIMO FMCW Radar

the left radar, the center camera and the right radar. A clear detection of the target was obtained only when the retro-reflector was pointing towards the radar. Thus it is very difficult to detect three retro-reflector with both radars at the same time for calibration.

III. Related work

For camera calibration, one of the first methods for camera intrinsic and extrinsic calibration was proposed by Tsai [4] in 1987 using a planar calibration targets - checkerboards. In 2000, Zhang [5] proposed a technique for camera calibration which only requires the camera to observe a checkerboard shown at two or more different orientations. Later in 2014, Straß et al. [6] proposed an improved checkerboard calibration method by adding binary code patterns on the edge of the planar targets. Apart from rectilinear lens calibration methods, Beck [7] introduced a B-spline camera model that specifically works for fish-eye lens camera.

For radar calibration, Anderson et al. [8] designed a networked radar system that combines two radar system into a single network for tracking of UAVs across a wide open area. Their method requires the locations of the targets to be known from the radar position by using a GPS system. Olutomilayo et al. [9] proposed an extrinsic calibration method for 77-GHz automotive radar. His method was designed to calibrate radars sensors that were installed for monitoring the external environment of the vehicle. The calibration is done by registering the radar detections of corner retro-reflector targets placed in known locations in the vehicle coordinate space. These methods are only applicable for calibration between two or more radar sensors, but they do not work if RGB and ToF cameras are introduced into the system.

For multimodal sensor calibration, Domhof el al. [10] presented an extrinsic calibration tool for radar, camera and lidar. They designed a novel calibration board which is detectable for all sensor modals. The board has four separated circular holes, which can be accurately detected by a few lidar beams, and a corner retro-reflector at the center which can be detected by the radar. After getting the coordinates of the calibration boards in each sensors' coordinate system, the cameras were calibrated by minimizing the total distance error between all targets. In [11] and [12], they both use special calibration target which is a combination of triangular board and corner retroreflector for calibration of multimodal sensors. The experiment were all done for the external of a vehicle in an open field.

Recently, Kinzig et al. [13] posted an article about multimodal sensor calibration in an environment similar to our experimental set up. They proposed an approach to calibrate one lidar, one vertical stereo camera, two RGB cameras, two radars and one IMU inside an autonomous shuttle bus. They first determined all the intrinsic and extrinsic parameters of the four cameras and then calculated the transformation between the lidar and the cameras. Then the transformation between the lidar and the radars was estimated using a styrofoam sphere with a diameter of 500mm and a corner retro-reflector inside it. The corner retro-reflector appears as a strong single target to the radar in the two-dimensional range-azimuth map, while the styrofoam sphere can be detected by the lidar with high accuracy. The lidar and radar can thus be calibrated.

All the reviewed methods for multimodal sensors calibration require using large calibration targets in an open field for the calibration, which is not suitable for a small vehicle compartment. In addition to the problems described in Section II, we therefore propose a new calibration approach.

IV. Hardware system

The sensors used in the experiment hardware system is described in this section.

A. MIMO RADAR

Two radar sensors were used in the experiment, they are MIMO-FMCW radar produced by Silicon Radar GmbH (Fig. 5). Each has 4 transmitting antennas and 4 receiving antennas. The antennas are arranged to synthesize a 1D virtual antenna array that consists of 16 virtual antennas.

The configurations of the radars are listed in Table I.

TABLE I. Radar configurations

	Radar 1	Radar 2
Base Frequency	77 GHz	81 GHz
Band Width	3 GHz	3 GHz
Range resolution	5 cm	5 cm
No. samples/chirp	512	512
Ramp Duration	204.8 us	204.8 us
Ramp Interval	300 us	300 us
Radar frames/sec	30	30
chirp/Radar frame	8	8

B. MICROSOFT AZURE KINECT DK (DEVELOPER KIT)

The Microsoft Azure Kinect DK was released in 2019. It consists of an 1-Megapixel Amplitude Modulated Continuous Wave (AMCW) Time-of-Flight (ToF) camera and a OV12A10

12MP CMOS RGB camera. Additionally to the high standard hardware specification, Microsoft has also developed a body tracking SDK for the Kinect DK which estimates 3D human body key points using Convolutional Neural Networks based on ToF data. The data that were used from the Kinect in this experiment is the color image and the ToF point cloud. The color camera was set to 1920x1080 resolution. The ToF camera was set to NFOV unbinned mode with 640x576 resolution. Despite Kinect offers a high-quality RGB image, in this work only the ToF data have being used, in order to test the capability of such type of sensor as ground truth data.

C. VEHICLE MOCK-UP

The vehicle mock-up used in our experiment was built for data collection of occupant postures (Figure 2). There are two Azure Kinects, two MIMO radars and one fish-eye lens camera installed in the mock-up. One Kinect is installed in the front and one is installed on the top. The radars are installed on the left and right side of the windshield pointing towards the center console. The fish-eye lens camera is installed next to the radar on the co-driver's side. In this paper, the calibration method for the front Kinect (color image + ToF point cloud) and the radars will be illustrated. The rest of the cameras can be calibrated with the same approach.

V. METHODOLOGY

Our proposed method involves applying existing camera calibration method and radar object detection algorithm to project radar point cloud onto 3D space in global coordinate system using ToF data as reference. Due to its detection principle, the ArUco method's precision proposed by [18] is strictly related to the image quality - resolution and lens deformation. Cameras with higher resolution and good lens quality show better results since the reprojected pixel error during the calibration process is lower. In order to perform the localization of all ArUcos in the mapped environment, a external high resolution camera - an Iphone 11 Pro Max with 3024x4032 resolution was used.

The first step was to calibrate the intrinsic parameter for all 2D cameras. For the external high-resolution camera, a total of 20 images were taken with the camera configured in the "professional mode" - with the auto focus disabled in order to keep fx, and fy constant. The intrinsic calibration for the highresolution camera shown an average projective error less than 1.5 pixel, which is determined by computing the euclidean distance between the real and the re-projected point in the un-distorted image.

After the intrinsic parameters were obtained, a set of images were taken, where each one contains at least 3 markers. The ArUco method implemented in OpenCV detects the regular pattern of each item and decode the address of each one, localizing and identifying each target automatically. Using the known marker dimensions in mm, the corner detected and the "K" and "D" matrices from the camera, it is possible to estimate the 3D position (x,y,z,Rx,Ry,Rz) of each marker with respect to the center of the camera and thus, the camera pose relative to each markers.

With the markers being fixed in specific positions and previously measured, it is possible to recreate a 3D representation of the environment, considering one of the ArUcos as the "ZeroReference". With a reference, every optical sensor capable of detect this marker can derive its own position just performing an ArUco position estimation in its own manner, finding automatically its position in space. As stated before, only visual sensor such as 2D RGB , 2D IR and ToF cameras can make use of ArUcos pattern detection - posing a problem for detection it in a on-line fashion with the radar.

In this work we propose an offline calibration method for the radars using the ArUcos placed on top of the sensor cover - which allows us to obtain a reliable and precise position of the radars in the vehicle In-Cabin environment.

A. INTRINSIC CALIBRATION

In terms of 2D cameras, such as the RGB camera, the intrinsic parameters referred as "K" and "D" matrices, consists in sets of data containing the focal length (fx,fy) and the image centre (cx,cy) - "K" matrix, and coefficients related to the lens model deformation - "D" matrix. This information is obtained using the chessboard method [5] using a sequence of several images that were taken in different positions, distances and rotations/inclinations. It is a homography based principle - mapping a plane with a well known pattern with precise distance to "u,v" coordinates detected in the image. The problem can be solved by applying a DLT between two planes, solving the system using the pinhole model parameters (fx,fy,cx,cy) and the distortions caused by fish-eye projection. On the other hand, ToF and other types of 3D sensors are calibrated before-hand, in the production phase. An AMCW Time-of-Flight camera, for example, must be calibrated for several internal parameters of modulation, additionally to the lenses characteristics and distortions. That being said, no intrinsic calibration was needed for the Kinect ToF camera, using the one provided by Microsoft.

Fig. 6. Three images taken from different angles at the vehicle mock-up

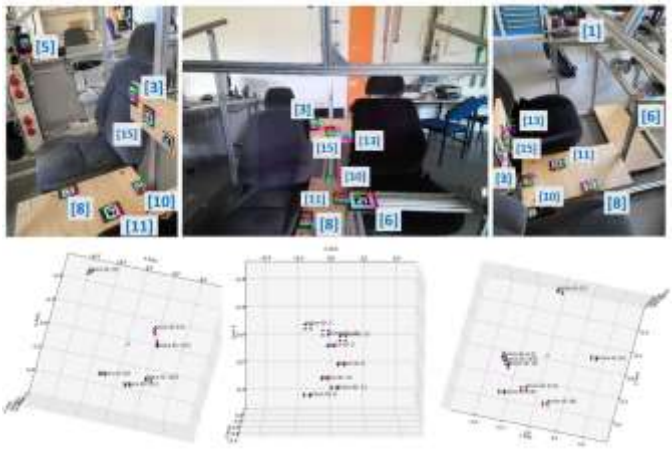


Fig. 7. ArUcos and their corners detected and mapped in 3D space

B. EXTRINSIC CALIBRATION OF RADARS

The matrix which contains the rotation *R* and translation *T* referred to a zero-reference point of a specific camera in space is called extrinsic matrix. As mentioned before, the ArUco markers were placed in different areas in the vehicle mock-up. One of the ArUcos was chosen to be the origin of the vehicle coordinate system. This ArUco must be included in the images that will be taken for the calibration later. In our experiment, three color images were taken from three different angles. One was taken by the center Azure Kinect. The other two were taken by the iPhone 11 Pro Max. The camera positions for taking these images are indicated in Fig. 6.

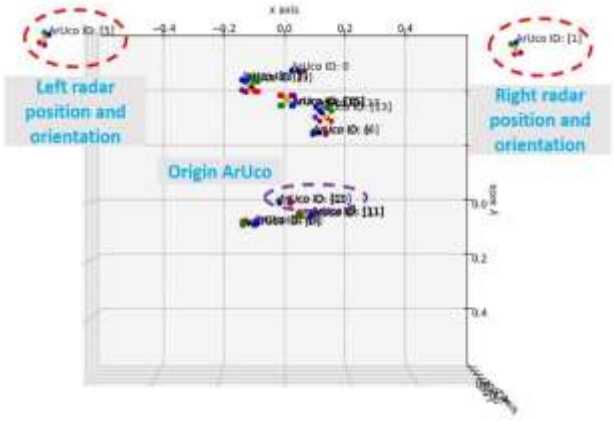


Fig. 8. All three frames calibrated into one coordinate system

The OpenCV ArUco library [18] was then used to find the position of all ArUcos in the images taken. With the library, the 3D coordinates of the center and four corners of the detected ArUcos were found in the camera's coordinate system (Fig. 7). The unique ArUco ID is drawn next to the ArUco. The frame where the radar on the co-driver's side was captured is named as the 'left radar frame', The frame where the radar on the driver's side was captured is named as the 'right radar frame' and the frame taken by the center Kinect is named as the 'center frame'.

At this point, the detected ArUcos are in three different coordinate systems. The next step is to transform all of them to the vehicle coordinate system. During the transformation, all ArUcos in each camera frame are treated as a rigid body such that the distance between all points are preserved after any transformation. The first step is to translate the center of the reference ArUco in all coordinate system to the vehicle origin (0, 0, 0). In our experiment, ArUco number 10 was chosen to be the reference ArUco. All ArUcos were translated as a rigid body such that the center of ArUco 10 is located at the vehicle origin. Then the rotational matrix to rotate the ArUcos in the radar frame to fit the ones in the center frame were calculated. The calculation is done by minimizing the distance error between corners of the commonly detected ArUcos in both frames. By applying the rotational matrix to the left and right radar frame, all three frames were calibrated into one coordinate system (Fig. 8), the vehicle coordinate system. In this coordinate system, the transformation matrix from the origin to the radars can be calculated. Finally, the radar detection point cloud can be projected into 3D space in the vehicle coordinate system.

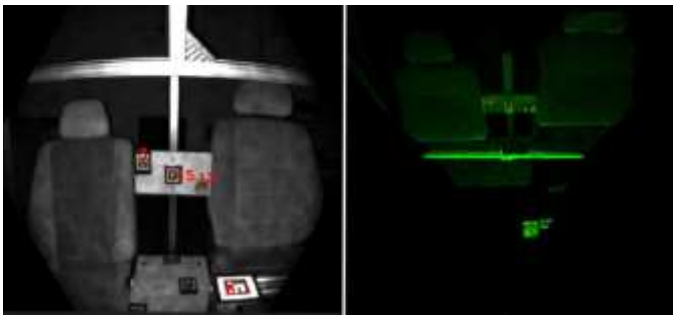


Fig. 9. Amplitude image from ToF point cloud (left) and ArUco markers detected in ToF camera's coordinate system in 3D space (Right)

C. EXTRINSIC CALIBRATION OF TOF CAMERA - ARUCO ON POINT CLOUD DATA

ArUco marker detection and decode was performed on the amplitude image generated from the ToF point cloud. Each pixel from the amplitude image is mapped to a 3D point from the point cloud, then detecting the marker into the amplitude image lead us directly to the respective 3D point (Fig. 9). In this way, it is possible to perform an on-line positioning using the Microsoft Kinect sensor - using the transformation matrix referred to the "Zero" marker. In terms of practical use, ToF sensors contains a random and systematic error related to its precision, which can lead to errors in the 3D positioning. In order to obtain a more reliable measurement, all the markers detected in the point cloud are used and a mean of all positions is taken to minimize position estimation errors.

D. RADAR DETECTION POINT CLOUD - CAPON ALGORITHM

To generate a point cloud to represents objects detected by the radar, the CAPON beam-forming algorithm [17] was used. It is a signal processing technique applied in array of sensors for determining the direction of arrival (DOA) of incoming signals. The algorithm was applied to the raw radar data after applying a fast Fourier transform (FFT).

The first step of the CAPON algorithm is to calculate the covariance matrix across the virtual antennas for all FFT range gates. The covariance matrix indicates the covariance between every variables, which represents how strong each variable is related to the others. The equation of the covariance matrix R is given by:

$$R = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \quad (1)$$

where N is the number of virtual antennas; x_i is the 1D vector of FFT data across a certain range gate; $\bar{x}$ is the mean of the 1D vector and T is the transpose. Then the next step is to define the steering matrix of the MIMO system ($Steer_{MIMO}$). In a MIMO system with 1D antenna arrangement, $Steer_{MIMO}$ is the Kronecker product between the transmitting antenna's steering matrix (S_{Tx}) and the receiving antenna's steering matrix (S_{Rx}). S_{Tx} and S_{Rx} can be define as:

$$S_{Tx/Rx}(k) = [e^{-2\pi j X(k) sin(\theta_0)}, e^{-2\pi j X(k) sin(\theta_1)}, ..., e^{-2\pi j X(k) sin(\theta_n)}] \quad (2)$$

where X is the antenna position vector, k is the transmitting/receiving antenna number and $[\theta_0,\theta_1,...,\theta_n]$ are the angles of arrival. The steering martix of the MIMO system can then be calculated:

$$S_{MIMO}(\theta) = S_{Tx}(\theta) \otimes S_{Rx}(\theta), \qquad \theta = [\theta_0,\theta_1,...,\theta_n] \quad (3)$$

The final step of the CAPON algorithm is to calculate the power of the uncorrelated signals $\bar{\sigma}_0^2$ impinging on the virtual antenna array:

$$\bar{\sigma}_0^2(\theta) = \frac{1}{S_{MIMO}(\theta)^T R^{-1} S_{MIMO}(\theta)} \quad (4)$$

After applying the CAPON algorithm to all FFT range gates, a range-azimuth spectrum that shows the position of strong reflective objects within the radar's Field of View (FoV) is generated.

E. PROJECTION OF ALL POINT CLOUDS ONTO VEHICLE COORDINATE SYSTEM

The point cloud generated by the sensors in the calibrated system can be projected onto the vehicle coordinate system if the transformation matrix from the vehicle origin to the sensor is known. In Subsection V-B, the transformation matrices for the radars were found; in Subsection V-C, the transformation matrix for the ToF camera was found. Therefore, the radar point clouds and ToF point cloud can now be projected onto the vehicle coordinate system for comparison.

VI. RESULTS

To measure the accuracy of our calibration method, the absolute distance error between the ArUcos from different frames were measured. Further, the detection of a retroreflector by the ToF camera was compared with the detection by the radars in the vehicle coordinate system.

A. ARUCO DISTANCE ERROR BETWEEN RADAR FRAME AND CENTER FRAME

The ArUco distance error between the radar frame and the center frame indicates how good the different frames fit together, smaller maximum and average error means better calibration result. The results are shown in Table II. The maximum distance error between commonly detected ArUcos does not exceed 2 cm and the average distance error is about 1 cm.

B. DISTANCE ERROR BETWEEN RADAR ARUCOS AND OTHER ARUCOS IN CENTER FRAME

To Further verify the accuracy of our calibration method, the maximum and average distance error between the radar ArUcos and other ArUcos in the center frame were also measured. The results were calculated by comparing the calculated distance and actual measured distance between the radar ArUcos and other ArUcos in the center frame. The results are shown in Table III. The maximum and average distance error are less than 2 cm.

TABLE II. DISTANCE ERROR OF ARUCOS BETWEEN RADAR AND CENTER FRAME

	Left radar frame vs center frame	Right radar frame vs center frame	Right radar frame vs left radar frame
Maximum distance error between commonly detected ArUco centers	1.051 cm	1.79 cm	1.642 cm
Average distance error between commonly detected ArUco centers	0.702 cm	1.05 cm	0.797 cm

TABLE III. DISTANCE ERROR BETWEEN RADARARUCOS AND OTHER ARUCUS IN CENTER FRAME

	Left radar ArUco vs other ArUcos in center frame	Left radar ArUco vs other ArUcos in center frame
Maximum distance error	1.9 cm	1.4 cm

Average distance error	1.46 cm	0.96 cm

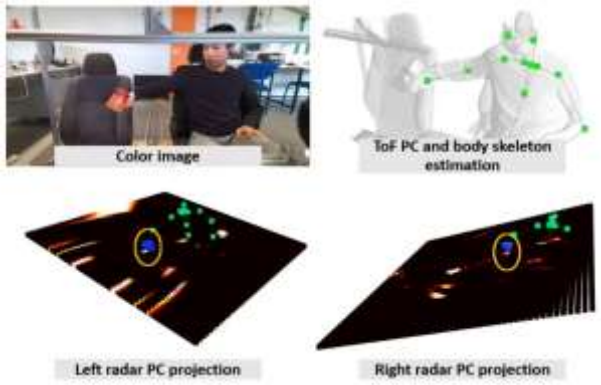


Fig. 10. Calibrated results between radar and ToF PC. The green dots represent body keypoints estimated by Kinect. The black grid represents the radar detection point cloud from CAPON algorithm. The bright spots on the black grid indicate objects detected by the radar. The blue cone represents the location of the retro-reflector.

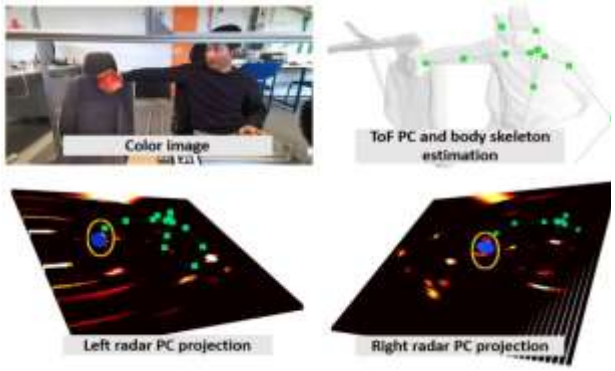


Fig. 11. Calibrated results between radar and ToF PC. The green dots represent body keypoints estimated by Kinect. The black grid represents the radar detection point cloud from CAPON algorithm. The bright spots on the black grid indicate objects detected by the rradar. The blue cone represents the location 0f the retro-reflector.

C. COMPARING RADAR DETECTION POINT CLOUD WITH TOF POINT CLOUD

Fig. 10 and Fig. 11 show the calibrated detection point cloud from the radars and ToF camera in the mock-up coordinate system. As the density of the ToF point cloud is much higher than the radar detection point cloud, for better visualization of the results, the ToF point cloud and radar point clouds are shown separately.

The top right image shows the ToF point cloud from the Kinect sensor, the body skeleton was also estimated by the sensor. In the lower left and lower right are the radar point cloud from left and right radar, the green dots connected by lines indicates the body skeleton estimated by the Kinect using the ToF point cloud. The blue cone marks the location of the corner retro-reflector. Inside the yellow circle, the radar detection of the corner retro-reflector is accurately projected onto the 3D mock-up coordinate system.

VII. CONCLUSIONS

In this paper, we presented a calibration method for calibrating color cameras, ToF cameras and radars in a vehicle. We combined traditional calibration method to calculate intrinsic target. A corner retro-reflector was only used to verify and visualize the calibration results at the end. The proposed method was tested in a vehicle mock-up and achieved high accuracy. The maximum and average distance error of this method is less than 2cm.

For future work, we plan to collect occupant postures and vital sign data using the calibrated sensor system in the vehicle mock-up. The data will be labelled and analysed for vehicle safety system development. With the radar point clouds accurately projected onto the vehicle coordinate system, it is possible to develop an algorithm for localizing occupants' body position and posture in 3D using the fused data from color camera and radar.

REFERENCES

[1] L. G. T. Ribas, M. P. Cocron, J. L. da Silva, A. Zimmer, and T. Brandmeier, "In-cabin vehicle synthetic data to test deep learning based human pose estimation models," in *IEEE Intelligent Vehicles Symposium, IV 2021, Nagoya, Japan, July 11-17, 2021*, pp. 610–615, IEEE, 2021.

[2] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2018.

[3] E. Cardillo and A. Caddemi, "A review on biomedical mimo radars for vital sign detection and human localization," *Electronics*, vol. 9, no. 9, p. 1497, 2020.

[4] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[5] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[6] T. Strauß, J. Ziegler, and J. Beck, "Calibrating multiple cameras with non-overlapping views using coded checkerboard targets," in *17th international IEEE conference on intelligent transportation systems (ITSC)*, pp. 2623–2628, IEEE, 2014.

[7] J. Beck and C. Stiller, "Generalized b-spline camera model," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2137–2142, IEEE, 2018.

[8] B. Anderson, J. Ellingson, M. Eyler, D. Buck, C. K. Peterson, T. McLain, and K. F. Warnick, "Networked radar systems for cooperative tracking of uavs," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 909–915, IEEE, 2019.

[9] K. T. Olutomilayo, M. Bahramgiri, S. Nooshabadi, and D. R. Fuhrmann, "Extrinsic calibration of radar mount position and orientation with multiple target configurations," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[10] J. Domhof, J. F. Kooij, and D. M. Gavrila, "An extrinsic calibration tool for radar, camera and lidar," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8107–8113, IEEE, 2019.

[11] J. Persi˘ c, I. Markovi´ c, and I. Petrovi´ c, "Extrinsic 6dof calibration of˘ a radar–lidar–camera system enhanced by radar cross section estimates evaluation," *Robotics and Autonomous Systems*, vol. 114, pp. 217–230, 2019.

[12] C.-L. Lee, Y.-H. Hsueh, C.-C. Wang, and W.-C. Lin, "Extrinsic and temporal calibration of automotive radar and 3d lidar," in *2020 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS), pp. 9976–9983, IEEE, 2020.

[13] C. Kinzig, M. Horn, M. Lauer, M. Buchholz, C. Stiller, and K. Dietmayer, "Automatic multimodal sensor calibration of the unicaragil vehicles," *tm-Technisches Messen*, vol. 89, no. 4, pp. 289–299, 2022.

[14] S. Garrido-Jurado, R. Munoz-Salinas, F. J. Madrid-Cuevas, and M. J.˜ Mar´ın-Jimenez, "Automatic generation and detection of highly reliable´ fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[15] D. Avola, L. Cinque, G. L. Foresti, C. Mercuri, and D. Pannone, "A practical framework for the development of augmented reality applications by using aruco markers," in *International Conference on Pattern Recognition Applications and Methods*, vol. 2, pp. 645–654, SciTePress, 2016.

[16] M. F. Sani and G. Karimian, "Automatic navigation and landing of an indoor ar. drone quadrotor using aruco marker and inertial sensors," in *2017 international conference on computer and drone applications (IConDA)*, pp. 102–107, IEEE, 2017.

[17] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[18] F. J. Romero-Ramirez, R. Munoz-Salinas, and R. Medina-Carnicer,˜ "Speeded up detection of squared fiducial markers," *Image and vision Computing*, vol. 76, pp. 38–47, 2018.

A Novel Hybrid Model for Prediction of Shear Modulus in Clayey Soil

Ehsan Mehryaar
Department of Civil and Environmental Engineering
New Jersey Institute of Technology
Newark, USA
em355@njit.edu

SeyedArmin MotahariTabari
Department of Civil Engineering
University of Akron
Akron, USA
sm577@uakron.edu

***Abstract*—In recent years, it was shown that machine learning can replace many statistical techniques in the prediction of the engineering properties of soil. However, still reliability of these solution is under investigation because of their dependence on small available databases. In this study, a new hybrid model based on M5p model tree and Support Vector Machine techniques are proposed to predict soil shear modulus at different strain levels. A data base with more than 1200 data points from laboratory experiments was used. To assure best performance of the model, a random search optimization technique was utilized to find best hyper-parameter combination for the model. The performance of the model is investigated using coefficient of determination and Root Mean Squared Error and 5-fold Cross-Validation and compared to available statistically derived equation from literature. A sensitivity analysis was performed to find importance of each input parameter for capturing complexity of the problem.**

Keywords— M5p model tree, Support Vector Machine, Shear Modulus, Clays, Machine Learning

I. INTRODUCTION

Many dynamic soil problems such as soil-structure interaction during earthquake, Industrial machines foundation vibration and etc. depends on accurate soil dynamic curves. For acquiring stiffness-strain curve of the soils, engineers rely on time consuming tests that are specific to a certain site [1]. Therefore, in recent years, researchers show interest in using data from various soils to find a general answer to this common problem.

Slope of shear stress (τ) and shear strain (γ) is expressed as secant shear modulus (G_s). Since the slope of the curve at each level of strain can changes, therefore, G_s can vary at different levels of strain [2]. This non-linearity adds to the complexity of the problem.

There are various techniques for acquiring soil stiffness curve. In general, they can be divided into two categories of in-situ tests and laboratory tests. In-situ testing includes surface wave testing, bore-hole testing using shear wave velocity (V_s) [3]. The well-known equation that relates maximum shear modulus to shear wave velocity and density of soil (ρ) is

$$G_{max} = \rho V_s^2 \quad (1)$$

In laboratory testing can be further divided into monotonic and cyclic or static or dynamic. Some of these tests are resonant column, cyclic simple shear, cyclic torsional shear, and cyclic triaxial tests [4].

Initial efforts of researchers in finding a more general solution to assessment of soil shear modulus at different strain levels were limited to statistical analysis. Zhang et al. [5] provided polynomial predictive equations for soil shear modulus and damping ratio of three different age categories of soils. They although provided a case study of their proposed equations for Charleston, S. C. to indicate applicability of their equations. Bayat and Ghalandarzadeh [6] acquired shear modulus based on resonant column, cyclic triaxial, and shear waves velocity measurements and proposed empirical equations for shear modulus and damping ratio of granular soil. Rahman et al. [7] proposed a novel equivalent granular state parameter technique for investigation of effect of stress state, density state, and fines content on dynamic response of granular soil with different fine content.

In recent years, with rapid development of machine learning and artificial intelligence techniques, researchers are turning toward these techniques for prediction of soil shear modulus. Sharma et al. [8] utilized Artificial Neural Network (ANN) and gene-programming (GP) for prediction of dynamic response of geogrid reinforced foundation beds. They found that GP is better suited for prediction of dynamic response and main affecting factor in their models was operating frequency of the machines installed on the foundation. Khajeh et al. [9] developed two models based on ANN and Support Vector Machine (SVM) fore prediction of damping ratio and shear modulus of soil mixed with geomaterials. They found that gravel fraction in their mixtures has high impact in dynamic response of the soil. They also found that an increase in gravel fraction and mean confining pressure increases shear modulus of the mixture. Diaz et al. Used ANN and decision trees to improve accuracy of seismic maps by combining geological and geophysical data.

In this study, a hybrid model based on M5p model tree (M5p), and Support Vector Machine (SVM) techniques is proposed. To ensure the model provides highest accuracy possible, a random search hyper-parameter tuning technique is

utilized. Performance of the developed model is compared to the existing empirical equation developed on the same data and simple SVM model. An uncertainty analysis is performed to find the importance of each input variable.

II. MACHINE LEARNING FRAMEWORK

A. M5p model tree

The M5p model tree is a modified technique based on the M5 algorithm first proposed by Quinlan [10]. The M5p algorithm can be divided into four steps: splitting the data space, training the models, pruning, smoothing [11]. Fig. 14 shows a schematic view of the M5p model.

Splitting criterion in M5p model is based on the assumption that reducing the standard deviation of dataset will lead to reduction in error. Therefore, algorithm performs a test on each input parameter to find the one parameter which provides most reduction in standard deviation and uses that for splitting. The Algorithm uses the following formula to calculate the standard deviation reduction (SDR).

$$SDR = \mathrm{sd(T)} - \sum \frac{|\mathrm{T_i}|}{|\mathrm{T}|} \mathrm{sd(T_i)} \quad (2)$$

Where sd is standard deviation, T is the subset of data at that node, and T_i is the subset of data from probable split at the current node. The splitting ends when the tree reaches a pre-defined maximum number of splits or the standard deviation at that node is less than a small value [13].

The next step is developing the model at each leaf (terminal nodes). Previous M5p models use linear regression at terminal nodes, however, in this study, the algorithm uses SVM for regression [14].

Sometimes this process can create a very big tree which might lead to high accuracy in training data and not enough accuracy in testing data which is called over-fitting. Pruning is a technique that is used to prevent this issue. The tree would be pruned back until the accuracy reduced to a certain threshold [15].After trees are built, a high level of discontinuity is usually observed at the boundaries of each terminal node. To prevent that, the models at each terminal node are updated in a way that their prediction for cases close to the boundaries of converge to a certain value [10].

B. Support Vector Machine

SVM is a supervised machine learning algorithm that can handle both classification and regression problems [16]. The main idea in SVM for classification consists of projecting data into a higher dimension of feature space and finding a hyperplane with biggest margin distance to the closest data point of different classes [17]. However, for regression, to use the concept of margins, Vapnik proposed the concept of an insensitive loss function (ε) [18]. In Support Vector Regression (SVR) the purpose is to find a hyperplane that all data points haave a distance less than ε. Therefore, for a sample of k training data $((x_1, y_1), \dots, (x_k, y_k))$, a linear model can be constructed as

$$g(x, w) = \sum_{k=1}^{m} w_k \phi_k + b \quad (3)$$

Where ϕ_k is nonlinear transformation to R^N new feature space, w_k is adjustable parameters and indicate flatness and $w_k \in R^N$, b is bias and $b \in R$. To achieve a smaller space and reduce the complexity of the model, SVR proposes minimization of Euclidean distance of $\|w\|^2$. Thus, an optimization problem can be defined in a way to minimize following function

$$\tfrac{1}{2}\|w\|^2 \quad (4)$$

If insensitive loss function conditions are satisfied

$$\begin{Bmatrix} 0 & ,if\ |y_i - g(x,w)| \le \varepsilon \\ |y_i - g(x,w)| - \varepsilon, & otherwise \end{Bmatrix} \quad (5)$$

To make the SVR process easier, concepts of kernels are introduced to map the existing data into new feature space. The common kernels are linear, polynomial, radial basis function, and Sigmoid [19].

C. Synthetic Minority Over-sampling for Regression

In most civil engineering problems, the acquired data is usually imbalanced. In classification problems, this usually means that the observation for one of the categories is substantially more that the other categories. In these cases, usually an oversampling (increasing the number of observations for minority groups) or undersampling technique (decreasing the number of observations for majority group) should be used [20].

However, in regression problems this is more complicated. Since in regression the target value is continuous and not categorical defining the minority data can be difficult. Also, in case of oversampling, unlike classification, the target value for newly synthesized data is not clearly given. SMOTER is an oversampling algorithm for this purpose. This algorithm consists mainly of three parts [21].

The first part is identifying the minority groups; this is done using a relevance function and a predefined relevance threshold. In this study, a Sigmoid function has been used for measuring the relevance of the data. The second part is synthesizing the new samples. This is achieved by randomly choosing one of the k-nearest neighbours of each minority observation and interpolating between these two values using a random number between 0 and 1. The third part calculates the target value. This is done by averaging the target values of the two-source observation based on the distance of the synthesized example from each source [22]. Fig. 15 illustrates the SMOTER process of creating new examples in a 2D problem.

D. Random Search Hyper-parameter Tuning

Usually, machine learning techniques have a set of predefined variables called hyper-parameters that control the process of training and finding the best solution. Therefore, finding the best set of this variables can highly affect performance of the models [23]. Random search (RS) is an algorithm that finds the best combination of these hyper-parameters. First, a range or a set of viable values should be defined for each hyper- parameter. Then, RS creates a grid-like space using those values. Next, it trains the model n number of

times and performs Cross-Validation (CV) on each to find the hyper-parameter combination corresponding to the model with highest performance [24]. It is seen that, if n is large enough, RS can find a combination of hyper-parameters with high performance, and it can be less computationally expensive in comparison to exhaustive algorithm such as Grid Search [25]. **Error! Reference source not found.** shows the hyper-parameter search space for SVM models.

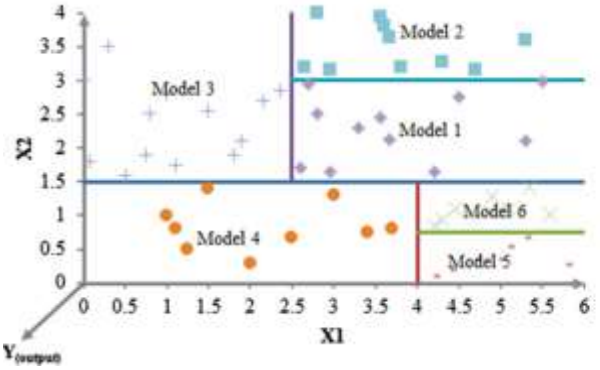


Fig. 14. A schematic view of the M5p model tree [12].

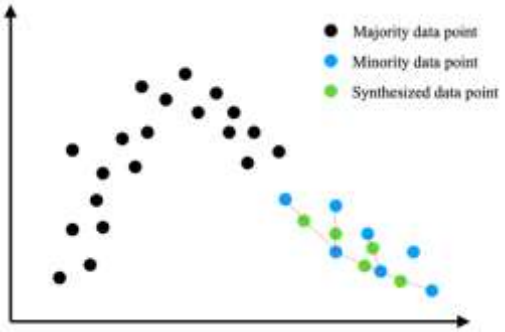


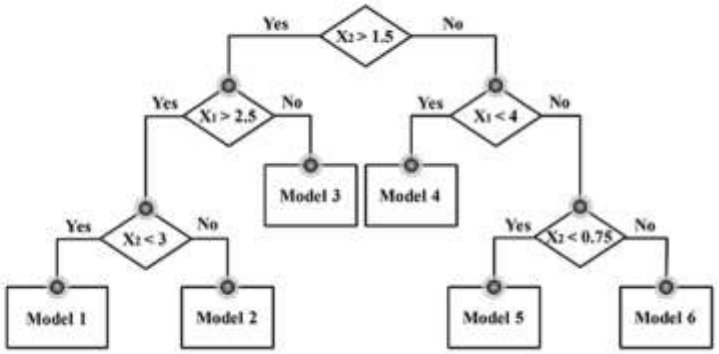
Fig. 15. Illustrates the SMOTER process of creating new examples in a 2D problem.

TABLE I. SEARCH SPACE FOR SVM HYPER-PARAMETERS

Hyper-parameter	Search Space
Regularization	0.1 to 100
Line-point distance importance	0.001 to 1
Kernel	Radial basis function, Sigmoid, Polynomial

E. Evaluation metrics

In this study, in order to evaluate the performance of the model two metrics are used to compare predicted values to observed values in the tests. They are coefficient of determination and Root Mean Squared Error RMSE which can be calculated by following equations respectively,



$$R^2 = \left(\frac{n\Sigma o_i p_i - \Sigma o_i \Sigma p_i}{\sqrt{(n\Sigma o_i^2 - (\Sigma o_i)^2)(n\Sigma p_i^2 - (\Sigma p_i)^2)}} \right)^2 \quad (6)$$

$$RMSE = \sqrt{\left(\sum_{i=1}^{n} \frac{(p_i - o_i)^2}{n} \right)} \quad (7)$$

Where p is the model prediction, o is the observed values, and n is the number of data points.

F. K-fold Cross-Validation

Sometimes, even by random selection of training data and testing data from the dataset, the training data can contain a small subset of data that does not is like the rest. However, due to this type of splitting these types of data cannot be found. The purpose of Cross-Validation (CV) is to make sure the model provides enough generalization on the used data set. The process starts by dividing the training data into k different folds and using one fold for evaluation and the rest for training. This step is going to be repeated k times until each fold has been used one time for evaluation and the rest has been used k times for training. Uniformity in the evaluation metrics between each test of this technique will ensure the generalization of the model.

G. Sensitivity Analysis

For any prediction problem, sensitivity analysis is the act of finding the effect of individual features and database size on the predictive model. Using this technique, it can be said that if the model performs better by using less feature and how much each feature affects the model performance in other words, how sensitive is the model to each feature. To perform sensitivity analysis each time, one feature is replaced with its mean on all data points and the model is developed. Then the evaluation metrics are calculated for that model. This process is repeated for all features [26].

III. DATA COLLECTION

In this study, a database of dynamic response of 21 clayey soil has been collected by Vardanega and Bolton [3] is used. The tests in this dataset were performed on various soils from different countries under different condition. Therefore, the results of study might reach to a good generalization. In this data base, 1219 data points of tests from different clayey soil is provided. The properties mentioned for each datapoint includes moisture content (w), initial void ratio (e_0), liquid limit (w_L), plastic limit (w_p), plasticity index (IP), effective confining stress (p'), shear modulus (G), maximum shear modulus (G_{max}), and shear strain (γ). The statistical characteristics of the dataset are given in TABLE I.

IV. METHODOLOGY

Python frameworks has been used in order to develop a prediction model for seismic response of clayey soil. A new hybrid method by merging M5p model tree and SVM is created and compared to SVM model. Fig. 16 shows the process used in this study to investigate the proposed models. The first step is importing the data and using SMOTER to balance the data based on G/G_{max}. Then, the balanced data is splitted into training and testing subset with proportion of 80% and 20%, respectively.

TABLE I. STATISTICAL CHARACTERISTICS OF THE DATA

Properties	γ	w	e_0	w_L	w_P	IP	p'	G/G_{max}
Mean	0.0068	0.46	1.23	0.70	0.32	0.38	236.62	0.55
sd	0.0153	0.32	0.82	0.26	0.12	0.18	157.13	0.32
Min	0	0.17	0.48	0.25	0.13	0.10	23.00	0.02
Max	0.0979	2.50	6.15	2.39	0.89	1.50	570.00	1.00
Variation coef.	2.2489	0.69	0.66	0.37	0.37	0.46	0.66	0.58
Skewness coef.	3.5983	2.73	2.47	2.39	1.09	2.58	0.35	-0.09

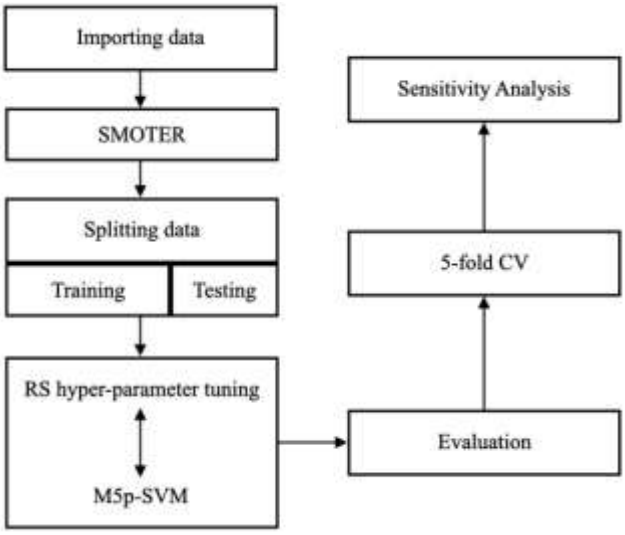


Fig. 16. Stepwise process adapted in this study.

After that RS hyper-parameter tuning and M5p-SVM or SVM approaches are used to find the best model. Then The model evaluated using testing data set and proposed performance metrics. Then to ensure enough generalization a 5-fold CV is performed on the model. Then the result of model is compared to the equation proposed by Vardanega and Bolton [3] for static adjusted curves

$$\frac{G}{G_{max}} = \frac{1}{1+\left(\frac{\gamma}{2.2\left(\frac{IP}{1000}\right)}\right)^{0.943}} \quad (8)$$

Then, A sensitivity analysis is performed on the data to find the importance of each feature in performance of proposed model.

V. RESULTS AND DISCUSSION

TABLE II compares performance metrics of developed model with hyperbolic equation in predicting the testing dataset. M5p-SVM shows the highest accuracy in prediction of G/G_{max} with 0.91 coefficient of determination. Hyperbolic formula is the next accurate model with R^2 value of 0.83 with 9% reduction in accuracy in comparison to M5p-SVm model, although both show acceptable performance. However, it is completely different for simple SVM model. R^2 value for SVM is 0.28 which is considered very low and shows that this model is incapable of predicting seismic behavior of clayey soils. It is interesting to notice the 69% difference in R^2 value of M5p-SVM and SVM model. This is an indicator that for hybrid model, most of complexity of the problem is distinguished with M5p tree structure and it is converted to multiple simpler problems that can be handle by SVM model. TABLE III shows the result of 5-fold CV for M5p-SVM and SVM models. The CV results show the same trends as performance metrics. The average R^2 value for the hybrid model is 0.946 and for all folds are between 0.94 and 0.95. These results are in consistency with the performance for testing result. The uniformity in accuracy of M5p-model is an indicator of good generalization of the model over dataset. For Simple SVM model the average R^2 value for

five folds is less than 0.1 which is another sign of incapability of this technique in handling the problem under investigation. Fig. 4 and Fig. 5 illustrate the scatter plot of predicted versus observed G/G_{max} for testing and training splits, respectively. The closer the points to the 45-degree line shows the higher accuracy of the model. As it can be seen, The M5p-SVM shows the highest convergence toward 45-degree line. Interesting observation can be done for simple SVM model. The SVM model predicts the seismic behavior of clayey soil on multiple horizontal lines. Which shows oversimplifying the problem by the model. Comparing this with the predictions of M5p-SVM model shows how splitting the problem to multiple simpler ones can help increase the accuracy of the SVM model. This can might be applicable for other existing machine learning techniques. Therefore, hybridization of existing techniques with M5p algorithm can lead to higher accuracy and more robust predictive models.

TABLE II. PERFORMANCE METRICS FOR DEVELOPED MODELS AND EMPIRICAL EQUATION FOR TRAINING AND TESTING SPLITS.

Model	R^2	$RMSE$
Training split		
M5p-SVM	0.91	0.006
SVM	0.28	0.088
Hyperbolic	0.83	0.012
Testing Split		
M5p-SVM	0.94	0.005
SVM	0.08	0.095
Hyperbolic	0.82	0.011

TABLE III. PERFORMANC METRICS FOR 5-FOLD CV

Model	R^2	$RMSE$
Fold 1		
M5p-SVM	0.95	0.005
SVM	0.16	0.07
Fold 2		
M5p-SVM	0.94	0.005
SVM	0.03	0.096
Fold 3		
M5p-SVM	0.94	0.005
SVM	0.03	0.095
Fold 4		
M5p-SVM	0.95	0.005
SVM	0.033	0.097
Fold 5		
M5p-SVM	0.95	0.005
SVM	0.21	0.047

TABLE IV contains the results of sensitivity analysis for M5p-SVM model. Replacing the shear strain with its average value reduces the R^2value by 90%. Therefore, the shear strain is the most important variable for predicting G/G_{max}. However, the same process for moisture content, liquid limit and plastic limit does not change the accuracy of the model. This shows that there is no relationship between seismic behavior of clayey soils and these parameters, and they can be ignored in developing the model. Sensitivity analysis for plasticity index and effective confining stress reduces the R^2 value by 3% and 8% respectively. Accordingly, in order, shear strain, effective confining stress, and Plasticity index are mostly related to the G/G_{max}.

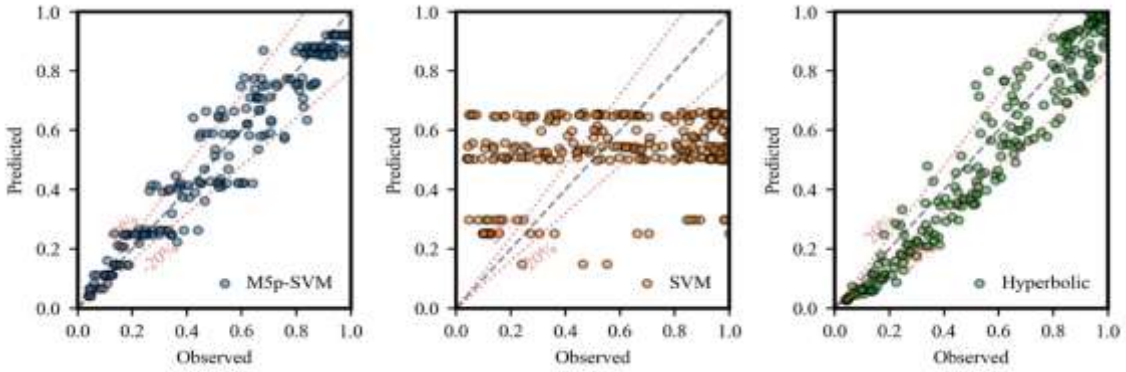


Fig. 17. Scatter plot of observed versus predicted G/G_{max} for developed models in testing phase.

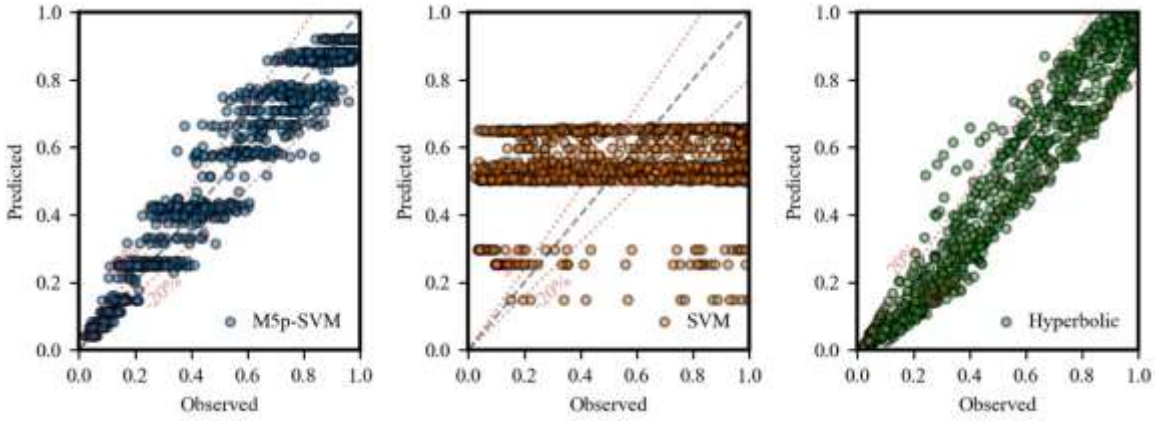


Fig. 18. Scatter plot of observed versus predicted G/G_{max} for developed models in training phase.

TABLE II. PERFORMANCE METRICS FOR DEVELOPED MODELS AND EMPIRICAL EQUATION FOR TRAINING AND TESTING SPLITS.

Model	R^2	*RMSE*
Training split		
M5p-SVM	0.91	0.006
SVM	0.28	0.088
Hyperbolic	0.83	0.012
Testing Split		
M5p-SVM	0.94	0.005
SVM	0.08	0.095
Hyperbolic	0.82	0.011

TABLE III. PERFORMANC METRICS FOR 5-FOLD CV

Model	R^2	*RMSE*
Fold 1		
M5p-SVM	0.95	0.005
SVM	0.16	0.07
Fold 2		
M5p-SVM	0.94	0.005
SVM	0.03	0.096
Fold 3		
M5p-SVM	0.94	0.005
SVM	0.03	0.095
Fold 4		
M5p-SVM	0.95	0.005
SVM	0.033	0.097
Fold 5		
M5p-SVM	0.95	0.005
SVM	0.21	0.047

TABLE IV. RESULTS OF SENSITIVITY ANALYSIS FOR M5P-SVM MODEL.

Parameter	R^2	*RMSE*
γ	0.08	0.18
w_0	0.91	0.008
w_L	0.91	0.008
w_p	0.91	0.008
PI	0.88	0.011
P'	0.84	0.012

VI. CONCLUSION

In this study, a hybrid model based on M5p model tree and SVM techniques are developed in Python programming language and its performance is compared to Simple SVM model and a hyperbolic equation. A random search hyper-parameter tuning technique were utilized in both M5p-SVM and SVM model to optimize the model.

Based on the results, proposed hybrid technique performance is superior to SVM and hyperbolic model. The Simple SVM model accuracy is 69% less than M5p-SVM model. This is an indicator that the idea of M5p algorithm dividing the big problem into multiple simpler problems lets the SVM to capture complexity of the matter much more effectively.

A sensitivity analysis was performed by replacing each input feature by the mean value of that parameter and keeping the other parameters the same. The results show that shear strain as expected is most important value for prediction of the G/G_{max}. Furthermore, effective confining stress and plasticity index are the next most important input parameters. Also, it was shown that moisture content, liquid limit, and plastic limit are not important for the developed model and are not effective in its accuracy.

REFERENCES

[1] A. Brunelli *et al*., "Numerical simulation of the seismic response and soil–structure interaction for a monitored masonry school building damaged by the 2016 Central Italy earthquake," *Bulletin of Earthquake Engineering*, vol. 19, no. 2, pp. 1181–1211, 2021.

[2] G. Chen, K. Liang, K. Zhao, and J. Yang, "Shear modulus and damping ratio of saturated coral sand under generalized cyclic loadings," *Géotechnique*, pp. 1–53, 2022.

[3] P. J. Vardanega and M. D. Bolton, "Stiffness of clays and silts: Normalizing shear modulus and shear strain," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 139, no. 9, pp. 1575–1589, 2013.

[4] L. Matešić and M. Vucetic, "Strain-rate effect on soil secant shear modulus at small cyclic strains," *Journal of geotechnical and geoenvironmental engineering*, vol. 129, no. 6, pp. 536–549, 2003.

[5] J. Zhang, R. D. Andrus, and C. H. Juang, "Normalized shear modulus and material damping ratio relationships," *Journal of geotechnical and geoenvironmental engineering*, vol. 131, no. 4, pp. 453–464, 2005.

[6] M. Bayat and A. Ghalandarzadeh, "Modified models for predicting dynamic properties of granular soil under anisotropic consolidation," *International Journal of Geomechanics*, vol. 20, no. 3, p. 04019197, 2020.

[7] M. M. Rahman, M. A. L. Baki, and S. R. Lo, "Prediction of undrained monotonic and cyclic liquefaction 6 behaviour of sand with fines based on equivalent granular 7 state parameter 8," 2014.

[8] S. Sharma, H. Venkateswarlu, and A. Hegde, "Application of machine learning techniques for predicting the dynamic response of geogrid reinforced foundation beds," *Geotechnical and Geological Engineering*, vol. 37, no. 6, pp. 4845–4864, 2019.

[9] S. Manafi Khajeh Pasha, H. Hazarika, and N. Yoshimoto, "An Artificial Intelligence Approach for Modeling Shear Modulus and Damping Ratio of Tire Derived Geomaterials," in *Advances in Computer Methods and Geomechanics*, Springer, 2020, pp. 591–606.

[10] J. R. Quinlan, "Learning with continuous classes," in *5th Australian joint conference on artificial intelligence*, 1992, vol. 92, pp. 343–348.

[11] B. Singh, P. Sihag, A. Tomar, and A. SEHGAL, "Estimation of compressive strength of high-strength concrete by random forest and M5P model tree approaches," *Journal of Materials and Engineering Structures «JMES»*, vol. 6, no. 4, pp. 583–592, 2019.

[12] A. Zahiri and H. M. Azamathulla, "Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels," *Neural Computing and Applications*, vol. 24, no. 2, pp. 413–420, 2014.

[13] M. Rezaie-balf, S. R. Naganna, A. Ghaemi, and P. C. Deka, "Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting," *J Hydrol (Amst)*, vol. 553, pp. 356–373, 2017.

[14] S. N. Almasi, R. Bagherpour, R. Mikaeil, Y. Ozcelik, and H. Kalhori, "Predicting the building stone cutting rate based on rock properties and device pullback amperage in quarries using M5P model tree," *Geotechnical and Geological Engineering*, vol. 35, no. 4, pp. 1311–1326, 2017.

[15] A. Behnood and D. Daneshvar, "A machine learning study of the dynamic modulus of asphalt concretes: An application of M5P model tree algorithm," *Construction and Building Materials*, vol. 262, p. 120544, 2020.

[16] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.

[17] J. Y. Park, Y. G. Yoon, and T. K. Oh, "Prediction of concrete strength with P-, S-, R-wave velocities by support vector machine (SVM) and artificial neural network (ANN)," *Applied Sciences*, vol. 9, no. 19, p. 4053, 2019.

[18] A. J. Smola, "Regression estimation with support vector learning machines," Technische Universität, München, 1996.

[19] S. M. Clarke, J. H. Griebsch, and T. W. Simpson, "Analysis of support vector regression for approximation of complex engineering analyses," 2005.

[20] B. Dimitrijevic, S. D. Khales, R. Asadi, and J. Lee, "Short-Term Segment-Level Crash Risk Prediction Using Advanced Data Modeling with Proactive and Reactive Crash Data," *Applied Sciences*, vol. 12, no. 2, p. 856, 2022.

[21] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "Smote for regression," in *Portuguese conference on artificial intelligence*, 2013, pp. 378–389.

[22] P. Branco, L. Torgo, and R. P. Ribeiro, "SMOGN: a pre-processing approach for imbalanced regression," in *First international workshop on learning with imbalanced domains: Theory and applications*, 2017, pp. 36–50.

[23] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of machine learning research*, vol. 13, no. 2, 2012.

[24] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Adv Neural Inf Process Syst*, vol. 24, 2011.

[25] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. de Carvalho, "Effectiveness of random search in SVM hyper-parameter tuning," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–8.

[26] H. Hamacher, H. Leberling, and H.-J. Zimmermann, "Sensitivity analysis in fuzzy linear programming," *Fuzzy Sets Syst*, vol. 1, no. 4, pp. 269–281, 1978.

A Hybrid Machine Learning Algorithm for Estimation of Tunnel Boring Machine Rate of Penetration

Ehsan Mehryaar
Department of Civil and Environmental Engineering
New Jersey Institute of Technology
Newark, USA
em355@njit.edu

SeyedArmin MotahariTabari
Department of Civil Engineering
University of Akron
Akron, USA
sm577@uakron.edu

***Abstract*—With a higher concentration on existing infrastructure around the word, tunnel-based transports are emerged as one of main choices in location with high population density due their less occupation of limited surface area. However, these projects are still dealing with significant cost and time overruns. Accurate prediction of tunnel bore penetration rate can be vital to optimal planning and subsequently decreasing financial risks of these projects. In this study, a hybrid solution based on Tree-structured Parzan Estimator and Extreme Gradient Boosting technique for prediction of rate of penetration is proposed. For unbiased comparison, the performance of the model is compared to the empirical equation in the literature based on the same database. An uncertainty analysis was performed to assess the reliability of the proposed model. A sensitivity analysis was performed to find the importance of each input feature for the proposed predictive model.**

Keywords—Extreme Gradient Boosting, Tree-structured Parzan Estimator, Tunnel Boring Machine, Rate of Penetration, Machine Learning

I. INTRODUCTION

In recent decades, with an increase in the population of urban areas and an increase in the need for transportation modes, tunnel-based transportation is becoming a more viable choice because it does not occupy the surface space that can be used for other social and economic activities [1]. However, tunneling projects are usually accompanied by high price tags and significant delays [2]. One crucial factor that affects the tunneling process is the penetration rate of tunneling machines (TBMs). The penetration rate of TBM has a direct impact on the scheduling of resources such as various materials and work forces, so it is a main tool in the planning of tunnel projects [3]. Therefore, accurate prediction of TBMs rate of penetration can lead to better planning and subsequently better time and cost efficiency of the projects.

Factors that affect penetration rate are generally divided into two categories of geological properties of the site and machine properties [4]. Most important geological properties are soil or rock unconfined compressive strength (UCS), tensile strength, and density and orientation of the faults and planes of weakness [4]. In recent years, researchers have used different input data (features) for their predictive models. Gao et al. [5] used only machine properties for developing their model including Torque, Velocity, Trust, the pressure of chamber at top left and bottom left of the machine. However, Minh et al. used different approach by only utilizing geological properties such as UCS, distance between planes of weakness (DPW), angle between tunnel alignment and planes of weakness (α), and brittle index (BI).

The initial predictive models were based on statistical analysis of existing data [6]–[8]. However, in recent years, by advancement of machine learning techniques and their promising application in civil engineering problems [9]–[11], researchers show significant interest in machine learning based techniques for TBMs rate penetration.

Salimi et al. [12] use a decision tree model for prediction of TBM penetration rate of Ghomrood water tunnel using geological data. Makaeil et al. [13] illustrated superiority of fuzzy Delphi analytic hierarchy process over fuzzy set theory and multifunctional fuzzy evaluation technique for prediction of Queens water tunnel TBM rate of penetration. Armaghani et al. [14] used Support Vector Machine (SVM) using Pahang-Selangor water transfer tunnel data.

In this study, a novel hybrid machine learning technique by merging Tree-Structured Parzan Estimator (TSPE) optimization method and extreme gradient boosting technique is developed to measure TBM rate of penetration. Data is acquired from Queens water tunnel [8]. The performance of the proposed model is compared to empirical equations given by Yagiz [8] using coefficient of determination and Root Mean Squared Error. A 5-fold Cross-validation is performed to assure generalization of the model. Then, an uncertainty analysis is performed to illustrate the 95% confidence error range of the model. Next, a sensitivity analysis is performed to determine the influence of each feature on the predictive model.

II. MACHINE LEARNING TECHNIQUES AND EVALUATION

A. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is an ensemble learning technique that uses the results of many decision trees as base learners to predict a target variable which are defined.

$$Y_i = \sum_{j=1}^{n} t_j(x_i), \;\; t_j \epsilon\, T \qquad (1)$$

Where Y_i is the prediction, x_i is the features, T is the function representing all decision trees and t is a single tree [15]. This means that XGBoost is an additive or forward stepwise technique and adds the results of a new tree two sum of previous ones to acquire new prediction therefore for n^{th} step of learning we have

$$Y_i^n = Y_i^{n-1} + \mathrm{t_n}(\mathrm{x_i}) \qquad (2)$$

The complexity of the tree is handled by a regularization term that uses vector mapping [16]

$$\Omega(t) = \gamma N + \frac{1}{2}\lambda \sum_{k=1}^{N} w_k^2 \qquad (3)$$

Where N is the number of terminal nodes (leaf), γ and λ are penalty factors corresponding to L1 and L2 regularization repectively, and w is the vector mapping function.

Then the objective function will be defined as

$$\mathrm{J}(\theta) = \sum_{i=1}^{M} l(y_i, Y_i) + \sum_{j=1}^{n} \Omega\left(t_j\right) \qquad (4)$$

Where y_i is the observed target value and Y_i is the predicted target value, and l is the loss function and defined as second degree Taylor expansion

$$l(x + \Delta x) \cong l(x) + l'(x)\Delta x + \frac{1}{2} l^{"}(x)\Delta x^2 \qquad (5)$$

Then for n^{th} step of learning objective function using additive feature of XGBoost would be [17]

$$\mathrm{J}(\theta)^n = \sum_{i=1}^{n} l(y_i, Y_i^{n-1} + t_n(x_i)) + \sum_{j=1}^{t} \Omega\left(t_j\right) \qquad (6)$$

B. Tree-structured Parzan Estimator Hyper-parameter Tuning

The Tree-Structured Parzan Estimator (TSPE) is a type of Bayesian optimization *technique* which can be categorized as Sequential Model-Based Optimization (SMBO) algorithm. In general, Bayesian optimization techniques try to create a surrogate function for probabilistic mapping of hyper-parameters to an objective function. Therefore, in Bayesian technique the goal is with each trial of hyperparameters refine the probability estimation of surrogate function and guess the next best hyper-parameter combination [18]. This is in contrast to grid search and random search algorithm that the next step hyper-parameter selection is without consideration of past selections [19]. The TSPE algorithm have the following steps [20]

1- Create a surrogate function of the objective function

2- Finding the hyper-parameter set with best result in surrogate function

3- Apply the selected set of hyper parameters to the actual objective function

4- Update the surrogate function with the result of the previous step.

5- Repeat steps two to four

The main advantage of TSPE over grid search or random search is to select the next set of hyper-parameters in a more intelligent manner to decrease the total number of trials. This can lead to higher efficiency of the technique [21]. Table I contains the hyper-parameter search space for XGBoost model.

C. Evaluation metrics

In order to evaluate the proposed model performance, two metrics of Coefficient of Determination (R) and Root Mean Squared Error (RMSE) is used.

$$R^2 = \left(\frac{n\Sigma a_i p_i - \Sigma a_i \Sigma p_i}{\sqrt{(n\Sigma a_i^2 - (\Sigma a_i)^2)(n\Sigma p_i^2 - (\Sigma p_i)^2)}} \right)^2 \qquad (7)$$

$$RMSE = \sqrt{\left(\sum_{i=1}^{n} \frac{(p_i - a_i)^2}{n} \right)} \qquad (8)$$

where p_i is the prediction, a_i is the actual value, and N is the number of observations.

D. Uncertainty analysis

In engineering problems, various sources of error can affect measurements of the parameters such as equipment errors, human errors, etc. Thus, uncertainty analysis is necessary for any predictive models in engineering. In this study, an uncertainty analysis based on distribution of errors is used. The error for each data point is defined by the difference between the prediction and the observed penetration value. When the number of data points are more than 30 it can be assumed that they follow a normal distribution [22]. Therefore, the mean and standard deviation of the errors can be calculated by

$$\bar{e} = \frac{1}{N}\sum_{i}^{N} e_i \qquad (9)$$

TABLE II. TABLE V. SEARCH DOMAIN OF HYPER-PARAMETERS FOR XGBOOST MODEL

Parameter	Search space
Maximum depth	3 to 18
Minimum loss function	1 to 9
λ	0 to 1
γ	40 to 180
Number of estimators	180

$$S_e = \sqrt{\frac{\sum_i^N (e_j - \bar{e})^2}{N-1}} \quad (10)$$

Using the standard deviation, a 95% margin of error can be calculated by

$$m = 1.96 S_e \quad (11)$$

Then, the confidence interval can be calculated [23].

$$CI = \bar{e} \pm m \quad (12)$$

E. Sensitivity analysis

Sensitivity analysis can be defined by investigation of influence of input features on the result of developed model. In other words, sensitivity analysis can help rank features based on their importance in the model. To perform the sensitivity analysis the model is re trained n times where n is the number of input features. In each trial, one variable is considered to have its mean value over all data points, and the rest of the variables are kept unchanged. This process is done for all variables [24]. The result of these trials would show the sensitivity model to input changes in individual input variables.

III. METHODOLOGY

Fig. 19 illustrates the process adapted in this study. To implement this algorithm a code using python libraries is developed. For dataset used is from Queens water tunnel available in literature [8]. The parameters available are UCS, DPW, α, Peak Slope Index (PSI), Brazilian Tensile Strength (BTS). Table II shows the statistical characteristics of dataset. The first step includes importing the data and splitting it into two subset of training and testing using 80% and 20% of data respectively. Then the training data is used to train the hybrid model. TSPE uses XGBoost to find the optimal hyper-parameters. Then, the best estimators would be evaluated using testing data and proposed performance metrics. Performance of the model is compared to Stepwise Multivariate Regression (SMR) derived [8] available equations in literature

$$ROP = 1.093 + 0.029PSI - 0.003UCS + 0.437Log(\alpha) - 0.219DPW \quad (13$$

After that, a 5-fold Cross-Validation is going to be done to find out level of generalization of the model. Then an uncertainty analysis is done to find 95% error bandwidth of the model. Finally, a sensitivity analysis is going to be performed to acquire importance of each input variable for the model performance.

IV. RESULTS AND DISCUSSION

In this section, performance of the proposed model for TBM rate of penetration is compared to the equation 13. Table III contains R^2 and $RMSE$ values for hybrid XGBoost and SMR models in training and testing splits. XGBoost provides an R^2 values of 0.88 and 0.94 for training and testing splits, respectively. Higher accuracy in the testing split compared to the training split is an indicator that the model is not overfitted.

TABLE III. TABLE VI. STATISTICAL CHARACTERISTICS OF THE DATA

Properties	**UCS**	**PSI**	**DPW**	**α**	**BTS**	**ROP**
Mean	150.05	34.58	1.02	44.72	9.55	2.04
sd	22.19	8.46	0.64	23.28	0.87	0.36
Min	118.30	25.00	0.05	2.00	6.70	1.27
Max	199.70	58.00	2.00	89.00	11.40	3.07
Variation coef.	0.15	0.24	0.63	0.52	0.09	0.18
Skewness coef.	0.63	1.43	0.17	0.01	-0.53	0.48

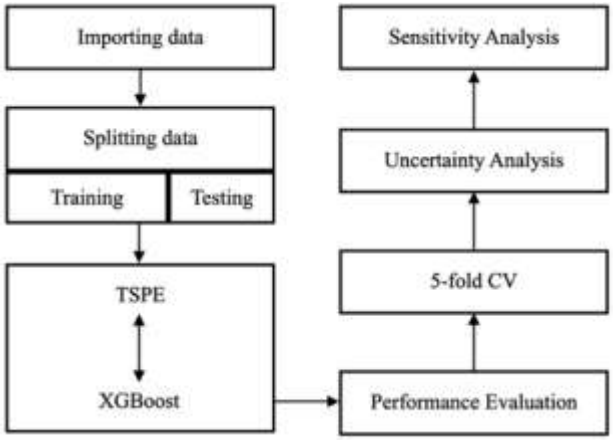


Fig. 19. Schematic illustration of the process implemented in this study.

TABLE IV. COMPARISON OF PERFORMANCE OF TSRP-XGBOOST MODEL WITH SMR MODEL

Model	**R^2**	**$RMSE$ (m/hr)**
Training		
XGBoost	0.88	0.014
SMR	0.64	0.044
Testing		
XGBoost	0.94	0.01
SMR	0.57	0.057

TABLE V. TABLE VII. PERFORMANCE METRICS OF THE XGBOOST MODEL IN 5-FOLD CV

Fold NO.	**R^2**	**$RMSE$ (m/hr)**
Fold 1	0.85	0.015
Fold 2	0.84	0.017
Fold 3	0.85	0.016
Fold 4	0.85	0.016
Fold 5	0.86	0.014

TABLE VI. TABLE VIII. UNCERTAINTY ANALYSIS RESULT FOR THE MODELS

Model	**Mean Error (m/hr)**	**95% Confidence Bandwidth (m/hr)**	**95% Confidence Interval (m/hr)**
XGBoost	-0.011	0.28	-0.29 to 0.27
SMR	-0.048	0.46	-0.51 to 0.41

The SMR model delivered an R^2 values of 0.64 and 0.57 for training and testing splits, which shows a 27% and 39% reduction in accuracy, respectively, compared to the XGBoost model. The same trend is applicable for $RMSE$ were the error values are substantially higher of SMR model compared to XGBoost. Table IV shows the result of 5-fold CV for XGBoost model. The average R^2 The value for 5-folds is 0.85 where each was between the range of 0.84 and 0.86. This is an indicator of a good generalization of the model over the dataset. Figs 2 and 3 scatter plot of predicted versus observed value of TBM rate of penetration for XGBoost and SMR model for training and testing splits ,respectively. The points on 45-degree line are indicator of 100% accuracy in prediction. It can be seen that for both training and testing splits XGBoost prediction provide better convergence toward 45-degree line. This is another indicator of higher accuracy of XGBoost model in comparison to SMR model.

TABLE VII. RESULTS OF SENSITIVITY ANALYSIS FOR XGBOOST MODEL

Parameter	R^2	$RMSE$ (m/hr)
Base model	0.94	0.01
UCS	0.86	0.02
PSI	0.46	0.08
DPW	0.47	0.07
α	0.75	0.04
BTS	0.89	0.02

Table V contains uncertainty analysis results for both XGBoost and SMR models. The average error for XGBoost model was -0.011 m/hr and for SMR model it was -0.048 m/hr. The negative value shows that both models underestimate the penetration rate. The absolute value of mean error for XGBoost is less than SMR model as expected by better accuracy of XGBoost model on performance metrics. The 95% confidence bandwidth for XGBoost and SMR model are 0.28 and 0.46, respectively. This means that 95% of the times it can be expected that the error of the models has value in a distance of less than the bandwidth from the mean error of the model. Therefore, smaller bandwidth shows

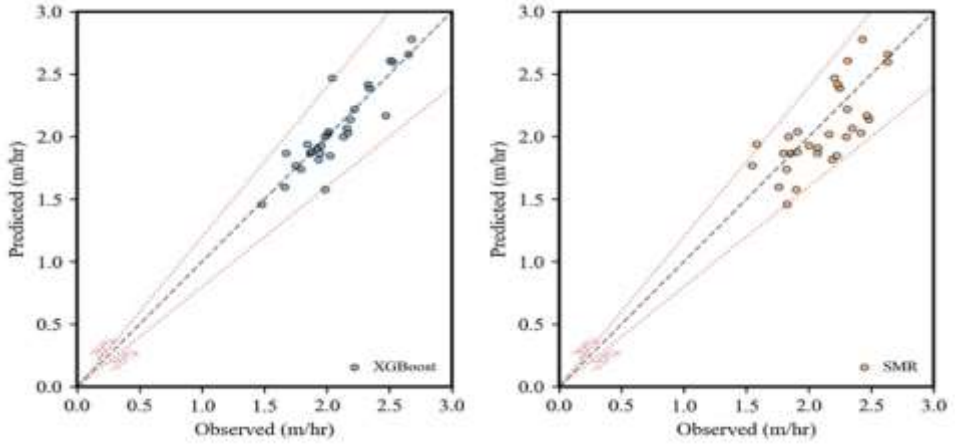


Fig. 20. Scatter plot of predicted versus observed ROP for testing split of the models.

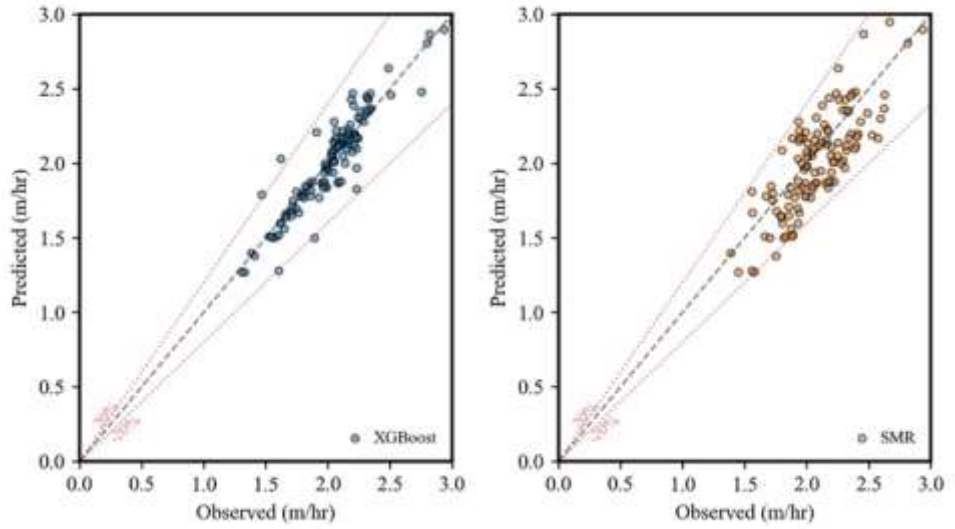


Fig. 21. Scatter plot of predicted versus observed ROP for training split of the models.

Higher confidence in the predictions of the models. The 95% confidence interval for the XGBoost model is -0.29 to 0.27 and for the SMR model it is -0.51 to 0.41. This again shows more robust performance of the XGBoost model compared to the SMR model.

Table VI provides the results of sensitivity analysis for the XGBoost model. By replacing the values of UCS, PSI, DPW, α, and BTS with their averages one at a time and retraining the model, the achieved R^2 values are 0.86, 0.46, 0.47, 0.75, and 0.89 which shows a reduction of 9%, 51%, 51%, 20%, and 5%. This shows that PSI and DPW are the most important input parameters for learning the complexity of the problem using the XGBoost model. α is somewhat important, UCS and BTS are the least important input variables for this matter.

V. CONCLUSION

In this study, a hybrid machine learning method based TSPE optimization and XGBoost techniques is proposed for prediction of TBMs rate of penetration. The performance of the model is compared of a statistically derived equation in the literature using the same dataset. A 5-fold CV was done to make sure that there was no data subset in training data that was ignored. An uncertainty analysis was performed to study the statistical characteristics of the model. A sensitivity analysis is done to find importance of each input variable in accuracy of the model.

Performance metrics showed substantially better performance of XGBoost model in comparison to SMR model. The R^2 value for training and testing splits were respectively 29% and 37% higher for hybrid model. Average R^2 value for 5-fold CV was 0.85 and all folds showed performance metrics close to each other, therefore, it can be concluded that the proposed model has sufficient generalization over the training dataset.

The performed uncertainty analysis illustrated smaller average error and smaller 95% confidence bandwidth of error for hybrid model compared to SMR. Hence, hybrid model provides more accurate and more certain predictions of the TBM rate of penetration.

The Sensitivity analysis showed that hybrid model mostly relies on PSI and DPW to capture complexity of the problem. However, sensitivity of the model to UCS and BTS were infinitesimal and they role in performance of the model were negligible. In conclusion, prediction of the proposed hybrid model were more accurate, more reliable than the SMR predictive equation.

REFERENCES

[1] P. Riera and J. Pasqual, "The importance of urban underground land value in project evaluation: a case study of Barcelona's utility tunnel," *Tunnelling and underground space technology*, vol. 7, no. 3, pp. 243–250, 1992.

[2] B. Flyvbjerg, M. K. Skamris Holm, and S. L. Buhl, "How common and how large are cost overruns in transport infrastructure projects?," *Transp Rev*, vol. 23, no. 1, pp. 71–88, 2003.

[3] H. Xu, J. Zhou, P. G. Asteris, D. Jahed Armaghani, and M. M. Tahir, "Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate," *Applied sciences*, vol. 9, no. 18, p. 3715, 2019.

[4] K. Thuro, "Drillability prediction: geological influences in hard rock drill and blast tunnelling," *Geologische Rundschau*, vol. 86, no. 2, pp. 426–438, 1997.

[5] X. Gao, M. Shi, X. Song, C. Zhang, and H. Zhang, "Recurrent neural networks for real-time prediction of TBM operating parameters," *Automation in Construction*, vol. 98, pp. 225–235, 2019.

[6] M. Alber, "Prediction of penetration and utilization for hard rock TBMs," 1996.

[7] P. J. Tarkoy, "Predicting tunnel boring machine (TBM) penetration rates and cutter costs in selected rock types," 1974.

[8] S. Yagiz, "Utilizing rock mass properties for predicting TBM performance in hard rock condition," *Tunnelling and Underground Space Technology*, vol. 23, no. 3, pp. 326–339, 2008.

[9] H. G. MELHEM and S. Nagaraja, "Machine learning and its application to civil engineering systems," *Civil Engineering Systems*, vol. 13, no. 4, pp. 259–279, 1996.

[10] P. C. Deka, *A primer on machine learning applications in civil engineering*. CRC Press, 2019.

[11] S. R. Vadyala, S. N. Betgeri, J. C. Matthews, and E. Matthews, "A review of physics-based machine learning in civil engineering," *Results in Engineering*, p. 100316, 2021.

[12] A. Salimi, C. Moormann, T. N. Singh, and P. Jain, "TBM performance prediction in rock tunneling using various artificial intelligence algorithms," 2015.

[13] R. Mikaeil, M. Z. Naghadehi, and S. Ghadernejad, "An extended multifactorial fuzzy prediction of hard rock TBM penetrability," *Geotechnical and Geological Engineering*, vol. 36, no. 3, pp. 1779–1804, 2018.

[14] D. J. Armaghani, E. T. Mohamad, M. S. Narayanasamy, N. Narita, and S. Yagiz, "Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition," *Tunnelling and Underground Space Technology*, vol. 63, pp. 29–43, 2017.

[15] R. Zhang, B. Li, and B. Jiao, "Application of XGboost algorithm in bearing fault diagnosis," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 490, no. 7, p. 072062.

[16] H. Zheng and Y. Wu, "A xgboost model with weather similarity analysis and feature engineering for short-term wind power forecasting," *Applied Sciences*, vol. 9, no. 15, p. 3019, 2019.

[17] T. Chen *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[18] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "BOA: The Bayesian optimization algorithm," in *Proceedings of the genetic and evolutionary computation conference GECCO-99*, 1999, vol. 1, pp. 525–532.

[19] R. Turner *et al.*, "Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020," in *NeurIPS 2020 Competition and Demonstration Track*, 2021, pp. 3–26.

[20] M. S. F. Erwianda, S. S. Kusumawardani, P. I. Santosa, and M. R. Rimadana, "Improving confusion-state classifier model using xgboost and tree-structured parzen estimator," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2019, pp. 309–313.

[21] H.-P. Nguyen, J. Liu, and E. Zio, "A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators," *Applied Soft Computing*, vol. 89, p. 106116, 2020.

[22] S. C. Gupta and V. K. Kapoor, *Fundamentals of mathematical statistics*. Sultan Chand & Sons, 2020.

[23] M. Najafzadeh, M. Rezaie Balf, and E. Rashedi, "Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models," *Journal of Hydroinformatics*, vol. 18, no. 5, pp. 867–884, 2016.

[24] A. Saltelli, "Sensitivity analysis for importance assessment," *Risk analysis*, vol. 22, no. 3, pp. 579–590, 2002.

Battling The Rise of Procrastination With The Implementation of Agile Methodology.

Aisha Abdur Rahim
Department of Computer Engineering and Informatics
Middlesex University Dubai
Dubai, United Arab Emirates
aishaar283@gmail.com

Chinnu Mary George
Department of Computer Engineering and Informatics
Middlesex University Dubai
Dubai, United Arab Emirates
C.George@mdx.ac.ae

***Abstract*— Procrastination is a widespread problem. Many people deal with the problem of Procrastination. The individual who keeps delaying the work faces a lot of issues. The main reason for procrastinating is being afraid of work. There exists a good number of websites and application that helps overcome Procrastination. However, when these systems were analyzed, it was found that some gaps and weaknesses exist. These issues become the reason for the user not being interested in the website. There needs to be a solution to motivate individuals to complete their work or tasks. To solve this issue and help people overcome Procrastination, this research paper presents the "Procast" website. The research logs in detail from the start to the end. It is shaped using the Agile methodology. The website is created for the purpose of preventing Procrastination and includes all the steps present in the Agile methodology.**

Keywords—Time Management, Procrastination, Productivity, Procrastinator, Agile Methodology.

I. INTRODUCTION

Procrastination is a huge problem, and it has a significant impact on one's life. It is often used to refer to the habit of not completing the work on time and delaying it further. Macnaught [1] points out that "Procrastination" is a widely searched term in today's world. The question of "How to gain victory against Procrastination" is searched almost daily on Google [1]. In a fast-paced world, where everyone is trying to get ahead in life, a huge percentage of people fall prey to Procrastination and stay behind [1]. O'Donovan [2] says that Procrastination makes people lose wonderful opportunities and, on the other hand, gifts them with burden and guilt. Macnaught [1] points out that it is easy to waste time on tasks that are not necessary, but it is hard to start a task that is important. There are certain aspects to why people procrastinate. Hence, there is a need for a website that can help people stop procrastinating. This research focuses on making a website that will counter more than one factor that are probable reasons for Procrastination. It will be done in such a way as to provide the user with a good experience as well as help them gradually come out of Procrastination. The research aims to assist and influence people to overcome the habit of Procrastination. The objective is to develop a website that can be a source that helps people beat Procrastination. The website will ask the user to enter the important tasks of the day. The user can either edit or remove the tasks if required. The website will also let the users prioritize the tasks based on urgency. The users will be praised for all the tasks completed. The user will be able to see the tasks that they have completed as well as not completed.

A. Problem Definition

The problem addressed is Procrastination. According to Cherry [3], Procrastination simply means postponing the tasks that need to be completed. Almost everyone has chosen to procrastinate instead of completing their tasks at some point in time. Cherry [3] emphasizes that taking a break and not completing a task at times is completely fine. But, the problem arises when one day of delaying leads to a phenomenon in which the individuals start to delay all their essential tasks [3]. Ho [4] says that the reasons for Procrastination are various. Most of the time, it is because the tasks at hand are complicated or hard. In such a situation, the human brain would choose an easier task, leading people to delay the work. This delay in work continues until the deadline [4]. Often, having a vague goal also leads to Procrastination because the goals that are vital to be completed are unclear [4]. Brockie [5] adds that another reason that is a cause of Procrastination is present bias. This means that individuals often prefer to vote for short-term happiness instead of long-term satisfaction [5]. O'Donovan [2] states that there are a lot of harmful effects of procrastinating. Procrastinating for a long time leads to low self-esteem and immense pressure. People feel guilty about not doing the work at hand, due to which they cannot even enjoy their chosen activity. O'Donovan [2] concludes that the huge burden of completing a big task as the deadline approaches closer has a harmful effect towards health and, consequently, on life.

B. Structure

The paper is divided into 5 sections. Section I is Introduction which speaks about the problem addressed as well as the aims and objectives. Section II is Literature Review that talks about Procrastination and existing websites for prevention. Section III describes the analysis of the website, the suggested design diagram, and the proposed method. Section IV explains the test carried out to check whether the website matches the set aims and goals. Lastly, Section V is the Conclusion, where limitations and future scope are discussed.

II. LITERATURE REVIEW

Procrastination is a common problem that is a part of most people's life. Macnaught [1] points out that a study was conducted among 1,000 adults in the UK. The results have indicated that almost 84% of people procrastinate and put off things for later. Among the 84%, it was further found that 20.5% procrastinated on a daily basis [1].

A. Existing techniques

To tackle the problem of Procrastination, quite a few websites have been created. One website is **Skip Procrastination.** This website is based on scientific thought that social media can be used to share and achieve [6]. The website takes in the user's task that the user has been putting off for some time. Then the user is asked the reason why it is essential and should be completed. Finally, the user enters a tentative time that is required to do the tasks [7]. All the above information is phrased into a short tweet that can be shared. The website then starts a timer of 25 minutes to prompt the user to start on the tasks. This website is best suited for those users who have been procrastinating on an important task for a long time [6]. Another website is **Why Do I Procrastinate**, which is useful at times when the user is unaware of the reasons behind procrastinating. The website enables the user to enter one task that is to be done by the user [6]. After this, the website provides a number of questions that target the user's feelings towards the task. The answers are in the form of a slider, ranging from 1 to 5. The results let the user know the core reason for procrastinating and how it can be faced [6]. Another website is **Procrastination.com.** This website serves more on the educational ideas about Procrastination. The motive of this website is to educate the users, and it includes certain elements [8]. Firstly, it has an online course where users can enroll and learn step-by-step processes to tackle Procrastination. Secondly, it conducts webinars that are open to everyone. Lastly, it has a blog that provides more information about Procrastination [8].

B. Strength and weaknesses

Patkar [6] points out that a study was carried out by Cornell University to find out whether anti-procrastination applications work. The results suggested that such systems work as it helps drive the user to start working. It was also found that the students who use these have more chances of being productive than those who do not. Others had to depend only on willpower [6]. There have been no weaknesses pointed out through research or studies.

C. Gaps

After analyzing the existing websites, some gaps were found which hinder the purpose of the websites. For instance, the **Skip Procrastination** website can be used only once at a time. It focuses on one factor: making the user accountable by posting the task online [6]. There is a risk that the user would go back to procrastinating after some time. There can also be instances where the user may choose not to post the task online [7]. Similarly, the quiz results can be inaccurate on the **Why Do I Procrastinate** website. On the other hand, if the answers provided by the users do not match the algorithm of the website, then there will be no results. The user's mood can also influence the outcome. The user may not answer truthfully, and hence the result can be incorrect [6]. Finally, some websites, such as **Procrastination.com,** only help by creating user awareness. This is not fruitful in situations where the user chooses not to read the content. Hence, the effort behind the website goes to waste. When Procrastination is concerned, it is better to develop a website. Custer [9] highlights that phone is a possible reason for Procrastination. It is more probable that the user may choose to use their phone for other purposes instead of using anti-procrastination applications [9]. Moreover, with the website being on the same machine the user is working on, there will be less need for the use of a phone. This will allow the website to be successful in being a companion that helps the users beat Procrastination. Considering all the above factors, the website will be designed to integrate all the elements and fulfill the gaps. This will be done to ensure that the website makes the user take action and is helpful.

III. METHOD

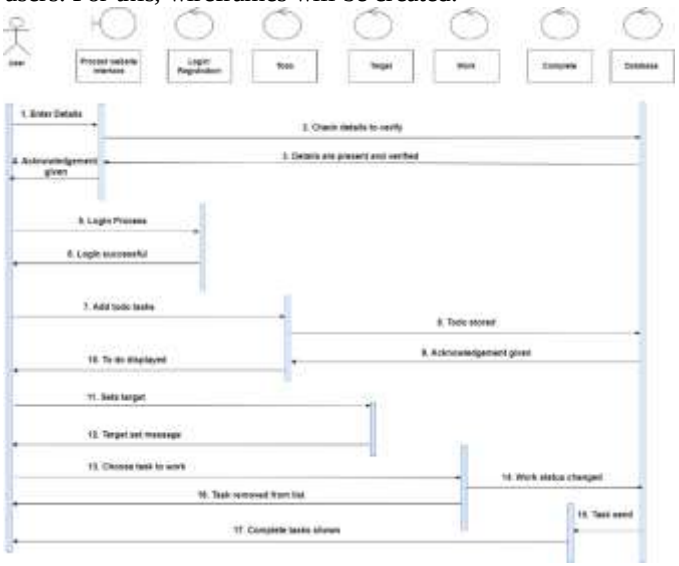
A. Analysis

From the analysis of the websites mentioned, the core feature that will be present in the **Procast** website is to allow the users to add in all their tasks. This way, all the tasks the user has to complete will be noted and presented in one place. Another feature will be to appreciate the user completing the tasks. This will work as motivation for the user to do their work. Yet, there can occur situations where the user may get bored and stop utilizing the website. To prevent this from happening, the website will have a feature that will allow the users to set a target of tasks that the user wants to complete out of all the tasks present. It will not be a problem if the user sets the target too low from the total number of tasks to complete. This is because the website's main aim is to get the users to do some of their work instead of none. This habit will gradually increase, and the user can focus and complete their work without procrastinating. The **Procast** website will be developed mainly, keeping students as the primary audience. It can be used by various types of people, but the reason for choosing students as the target audience is due to the fact that among all groups, students often participate in Procrastination the most. For instance, postponing finishing reports or studying for an exam [10]. The website can help students track their tasks and complete the work before the deadline. Anybody can register on the **Procast** website and use it for free. This allows everyone to access and utilize the website to start the journey of overcoming Procrastination. If seen from the perspective of marketing, this could help the website accommodate more users and become popular and well-known.

B. Design

The design is done following the Agile Model. Rough sketches are made to visualize the system, which will be modified with each iteration [11]. In Agile, there can be two approaches to doing the design. These include Software and User Interface design. In the first case, the requirements are considered, and

diagrams are formulated to depict how the system will be designed in order to fulfill the requirements. This is achieved by creating UML diagrams [11]. For this, various UML diagrams are present, among which **Sequence Diagram** is chosen for the research. It is described below. On the other hand, the interface design is considered when doing the UI design. It depicts how the system's front-end will look to the users. For this, wireframes will be created.



ce Diagram

IBM [12] explains that Sequence diagrams help depict the interactions in the system. It also showcases the messages passed to and fro between the various components of the system. This helps understand the sequence in which the system will react to the inputs received [12]. The Sequence diagram in Fig. 1 picturizes the general interactions on the **Procast** website. It contains all the interactions that will be done by the user on the website, as well as includes the responses from the system. As seen in Fig. 1, the Sequence diagram depicts two different interactions of the **Procast** website. For instance, these interactions can relate to the difference between registered and new users. In the case of a new user, as there exist no details previously, the user will have to register, which will save the data in the database. On the other hand, the registered users can log in with the account details created and access the website. The tasks present will be retrieved from the database and showcased.

C. Proposed Method

According to Ehrens [13], Implementation is when the ideas and requirements are practiced, which then leads to the development of the system that is anticipated. Before the implementation, it is crucial to plan the manner in which it will be executed. This will guarantee that the process does not fail and the system is developed [13]. The implementation of the **Procast** website is explained in this section. The development of the website is to be done with Node.js and Express. HTML is used to contain all the content to be displayed, whereas CSS is used to design the pages. “ejs” is the view engine used to render the HTML pages. Fig. 2 below shows the To-do page of the **Procast** website. The user can add all the tasks to be done on this page, which are then displayed together. The user can either edit or delete the tasks if any inaccuracies occur. At the same time, as showcased in Fig. 2 below, the user can take the challenge and set a goal of the tasks to be done out of the total number of tasks.

Fig. 2. To-do Page

All the tasks the user completes are moved to the Completed To-do page, which is depicted in Fig. 3 below. On this page, the user can unlock a reward, which becomes enabled only if the user meets the target set on the To-do page, as seen in Fig. 3. This motivates the user to push forward and complete all the tasks present on the to-do list. The reward ranges from an informative article that the user can read to a website that the user can browse. The rewards are meant to be a source of both providing a break and fun to the user.

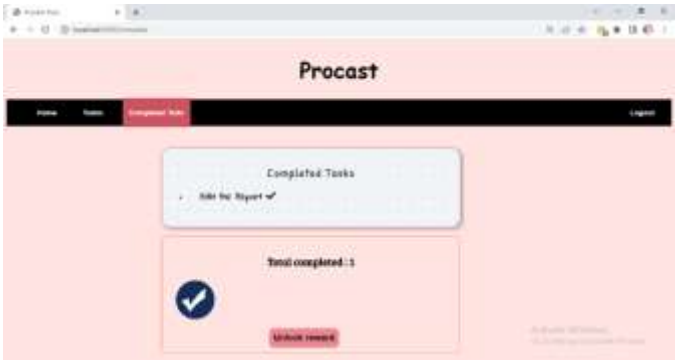


Fig. 3. Completed To-do Page

IV. RESULT AND DISCUSSION

A. Testing

According to Guru99 [14], it is very crucial to conduct testing. It acts as a source that will help identify any types of errors or gaps in the system. In other words, testing helps in assessing whether the system that has been developed is up to the mark. If a system is well tested, then there are very slight chances that it will have some errors. For testing the website, five kinds of tests will be carried out.

1) *Test Cases*

Guru99 [15] says that test cases account for some actions that are done to check specific features of the system. Table 1 showcases the test cases and results for the **Procast** website.

TABLE I. TEST CASES AND RESULTS

Test Cases	**Test Type**	**Expected Outcome**	**Actual Outcome**	**Pass (P) / Fail (F)**
Ensure that all the obligatory fields are set as required.	**Functionality**	If the required fields are empty, then there should be an error message displayed.	The website sends error messages that show that all fields are required.	Pass
Ensure that the response time of the website is fast.	**Performance**	The website should not take much time to load.	It can be seen that the website is able to run as soon as the server is running on port 3000.	Pass
All the pages of the website have to be linked to each other	**Usability**	The pages should be connected so that the user can move around.	The website has a navigation bar, which has links to all the pages, which makes the movement around the website smooth for the user	Pass
Any sensitive data should not be printed on the screen when entering	**Security**	The password should not be visible while entering.	It can be seen that the passwords are hidden and encrypted so that it will not be shown while the user is entering.	Pass
Ensure that all the elements on the website are readable.	**Usability**	The content of the website should be understandable and visible.	The design of the website matches the expectations of the users, and hence the elements provided are self-explanatory.	Pass
The todos that has been posted by one user should not be visible to other users.	**Security**	The todos that belong to the logged-in user should only be retrieved and displayed.	The website retrieves data based on the user's logged-in email from the database. This only takes in data that belongs to the specific user	Pass
All the buttons and links provided on the website are working.	**Functionality**	The buttons and the links should be doing what the code has intended to happen.	All the buttons present on the website are functional and submit data relevant to it. At the same time, the links are working and lead to the page that has been chosen.	Pass
The images/gifs of the website are visible in other browsers.	**Compatibility**	No matter which browser the user opens the website with, all the content, such as images/gifs, must be visible.	The website has the same layout, and all the content is visible across different browsers.	Pass

V. CONCLUSION

Procrastination is a significant problem that a considerable number of people face every day. It creates a negative impact on the life of the person. This is because, along with the stress of not completing the tasks, the person also feels guilt, anxiety, pressure, and lower self-esteem. It becomes important to have websites that provide the user with the right amount of motivation and fun to make the user want to overcome Procrastination. This goal was fulfilled by the development of the **Procast** website. Some limitations were faced with regard to the chosen language, which is JavaScript, at the initial stages of the research. The addition of a database to the website was difficult, and hence, more information had to be acquired. This led to introducing a view engine and back-end source to make the addition of the database possible. For future work of the research, the suggestions provided by the participants during the user testing will be addressed. This is to be done in order to enhance the website to match the user's expectations and make it the user's first choice. Secondly, the setting target feature can be set to assign the tasks to be done randomly. This will provide a challenge for the user with the promise of a better reward. At the same time, the user will be motivated to complete the tasks as this feature will give completing the tasks the shape of a game. It is to be noted that the addition of this feature will be most suitable for users who have been registered members for some time.

ACKNOWLEDGMENT

I express my sincere gratitude towards Mrs. Chinnu Mary George, who has mentored and guided me from the start to the end. She has constantly provided me with encouragement, support, and motivation to keep going forward. It was through her feedback and suggestion that I was able to enhance this research paper. Next, I would like to thank my family and friends for encouraging me to continue doing this research. Their immense support has played a big role in keeping me motivated and focused. Lastly, I would like to thank God for allowing me to complete the research successfully. Without his blessing and assistance, this research would not have been a success.

REFERENCES

[1] Macnaught, S(2020) *How to Stop Procrastinating (And the Numbers)* Available at: https://www.microbizmag.co.uk/procrastination-statistics/ (Accessed: 27 October 2020)

[2] O'Donovan, K (2020) *8 Dreadful Effects of Procrastination that can destroy your life* Available at: https://www.lifehack.org/articles/productivity/8-ways-procrastination-can-destroy-your-life.html (Accessed: 26 October 2020)

[3] Cherry, K. (2020). *What is Procrastination?* Available at: https://www.verywellmind.com/the-psychology-of-procrastination-2795944 (Accessed: 25 October 2020)

[4] Ho, L(2020) *Why Do I Procrastinate? 5 Root Causes & How to Tackle Them* Available at: https://www.lifehack.org/articles/lifehack/6-reasons-on-why-are-you-procrastinating.html (Accessed: 25 October 2020)

[5] Brockie, A(2020) *Why do you Procrastinate?* Available at: https://medium.com/@andy.brockie/why-do-you-procrastinate-c32a52a51030 (Accessed: 25 October 2020)

[6] Patkar, M (2019) *Why do we procrastinate? 5 Science-Backed Sites to understand and overcome it* Available at: https://www.makeuseof.com/tag/procrastinate-science-backed-sites/ (Accessed: 28 October 2020)

[7] SkipProcrastination (2020) *Procrastinating, This will help you start within 2 minutes.* Available at: https://skipprocrastination.com/ (Accessed: 27 December 2020)

[8] Procrastination.com (2018) *Stop Procrastination, Start Living with Procrastination.com* Available at: https://procrastination.com/#definition (Accessed: 29 December 2020)

[9] Custer, C. (2019) *Stop Procrastinating and Start Studying Using Science.* Available at: https://www.dataquest.io/blog/stop-procrastinating-science-based/ (Accessed: 29 December 2020)

[10] Solving Procrastination (2021) *Academic Procrastination: Examples, Consequences, Causes and Solutions* Available at: https://solvingprocrastination.com/academic-procrastination/ (Accessed: 3 April 2021)

[11] Feoktistov, I. (2021) *Agile Software Development Lifecycle Phases Explained.* Available at: https://relevant.software/blog/agile-software-development-lifecycle-phases-explained/ (Accessed: 5 January 2021)

[12] IBM (2012) *Sequence Diagrams* Available at: https://www.ibm.com/docs/en/radfws/9.6.1?topic=diagrams-sequence (Accessed: 10 April 2021)

[13] Ehrens, T. (2015) *Implementation* Available at: http://searchcustomerexperience.techtarget.com/definition/implementation (Accessed: 17 April 2021)

[14] Guru99 (2021) *What is Software Testing? Definition, Basics & Types* Available at: https://www.guru99.com/software-testing-introduction-importance.html (Accessed: 23 April 2021)

[15] Guru99 (2021) *Test Case vs Test Scenario: What's the Difference?* Available at: https://www.guru99.com/test-case-vs-test-scenario.html (Accessed: 24 April 2021)

An Optimized Image Steganography with High Embedding Capacity Based on Genetic Algorithm

Hadeel Alazzam
Department of Intelligent Systems
Al-Balqa Applied University
Al-Salt, Jordan
hadeel.alazzam@bau.edu.jo

Orieb AbuAlghanam
Department of Computer Science
University of Jordan
Amman, Jordan
O.Abualganam@ju.edu.jo

Abdulsalam Alsmady
Department of Computer Engineering
Jordan University of Science and Technology
Irbid, Jordan
aralsmady15@cit.just.edu.jo

Esra'a Alhenawi
Department of Software Engineering
Al-Ahliyya Amman University
Amman, Jordan
e.alhenawi@ammamu.edu.jo

Laith Al Shehab
Department of Computer Engineering
Jordan University of Science and Technology
Irbid, Jordan
Laith.alshehab@gmail.com

***Abstract*— Due to the rapid growth of the internet information security and data confidentiality are one of the main concerns of any type of communication. Information hiding is one of the most common techniques that used to provide data confidentiality which attracts more attention. Steganography is one of the most essential techniques for concealing information. It is the science of concealing secure information within a carrier object during the transmission through the public channel where only the sender and receiver can recognize and detect it. In this paper, a new optimized Steganography based on a Genetic algorithm has been proposed. The proposed approach relies on a random generator to guess the index of characters from a dictionary, while the Genetic Algorithm is used to find the best seed for the random generator. Only the number of tries required by the random generator to guess the index of the character is saved in the image instead of the full message. The proposed approach outperforms the traditional Least Significant Bit (LSB) technique in terms of Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), and cosine similarity using various messages sizes.**

Keywords—Genetic Algorithm, Optimization, Steganography, Least Significant Bit (LSB)

I. INTRODUCTION

Steganography is referred to as the technique of hiding data inside data. The first data refer to the secret information that needs to be hidden while the second data refer to the containers that hold the secret data inside which might be an image, video, or audio. Steganography is used to conceal a message using a physical object such as an image while in the digital image the secret message can be hidden inside or on top of an object that is itself not secret [1], [2].

The term stenographia, which is a combination of the words steganos and graphia, which denotes covered and written, respectively, has its origins in ancient Greek [3]. Also, it is noteworthy that Steganography is to some extent related to cryptography though the former is not exclusive to computers. For instance, the hidden message might take the form of an invisible ink concealed in a letter or image [4].

The objective of using Steganography is to ensure that security is enhanced through obscurity. Unlike cryptography, Steganography is designed in such a way that it does not attract any attention. The importance of Steganography is based on the fact that apart from relaying the contents of a message, a hidden message is also being sent [5]. It is noteworthy that technological advancements have seen the evolution of digital Steganography. In particular, the concept involves the use of advanced coding to conceal a message inside a document, file, protocol, or program [6]–[8]. In the majority of cases, digital Steganography involves the use of media files to hide a secret message. Besides, media files are ideal for the transmission of files and messages of large sizes [9].

Steganography greatly helps to hide the existence of a message, and as such, it is widely used in areas where there is a need for covert communication. Its utilization is gaining recognition in areas of intelligence, law enforcement, and the military. But the most important aspect worth noting is that Steganography ensures that the secret message is not distorted in any manner [10]. In essence, Steganography is an important aspect used in security installations, especially in fields where confidentiality of information is of utmost importance.

The image that hides the data in steganography is known as the Cover Image because it covers the secret message and stego-image represents the secret data and the Cover Image together [11]. LSB insertion is a popular and widely technique is used in steganography. The LSB embedding approach implies that data can be concealed in the LSBs of the cover file in such a way that even the naked eye is unable to detect the hidden information. It's a technique that works in the spatial domain [12]. Fig.1 illustrates the basic operations for steganography process to hide the secret message in a cover image.

It is notable that digital Steganography is mostly implemented using different computer algorithms. On the other hand, to choose any algorithm there are different metrics that should be taken into consideration one of the most important metrics is the compression ratio for the Cover Image which can be lossy or lossless. The compression ratio is associated with the enhancement of imperceptibility for the stego-image [13].

Genetic algorithm is used to enhance the robustness and the efficiency of the hidden message [14]. Notably, the algorithm utilizes search meta-heuristics to generate the best solutions and optimizations. It is based on Charles Darwin's theory of natural selection and belongs to the larger category of evolutionary algorithms [15].

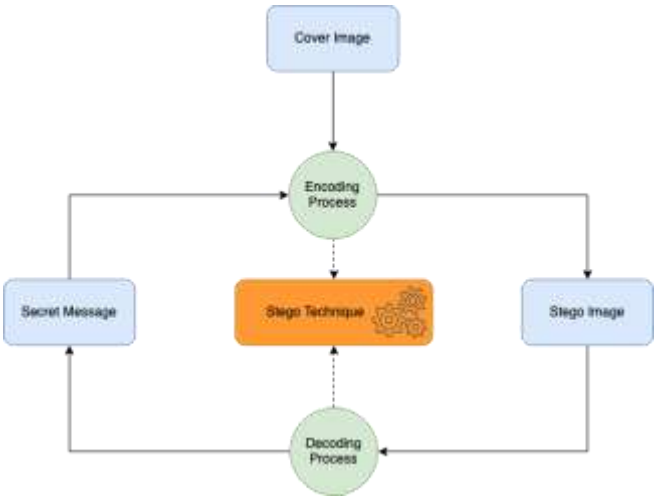


Fig. 1. Steganography Process

In this paper, a new Steganography approach has been proposed based on a Genetic algorithm. The proposed approach encodes the secret message inside an image by saving the number of flips required by a random generator with a predefined seed to produce the index of a character from a dictionary. Moreover, the proposed approach uses the Genetic Algorithm to find the optimal seed for a random generator for a certain message, that reduces the number of flips to target the index of the character. Which will reduce the number of bits required for saving the message.

The rest of this paper is organized as follows; Section II reviews the state-of-the-art related works of Steganography approaches that use the Genetic Algorithm, and Section III presents the proposed techniques. Section IV and Section V discuss the achieved results and concludes the paper respectively.

II. RELATED WORKS

Several proposals have been presented to conceal secret data based on different techniques to reach different goals. In [16] they use both cryptography and steganography to hide the secret data. Their main objective was focused on physical space on various storage media to reduce data transmission

time over the Internet. On the other hand, in [17] the proposed technique is based on steganography. It presents a modify secret data using a genetic algorithm and the least significant bit (LSB) has been proposed. The proposed model introduces a unique concept called a flexible chromosome, which allows GA to interpret chromosome values in a variety of ways. The results turn out that for 2 bit per pixel (bpp) and 3 bpp data hiding capacity, the suggested technique produces stego image with average PSNR values of 46.41 dB and 40.83 dB, respectively. On the other hand, the proposed technique is used both cryptography and steganography to hide data

Pramanik, Singh, and Ghosh (2020) propose a novel approach that involves the combination of bi-orthogonal wavelet transform and a Genetic algorithm to create image Steganography. In addition, the proposed technique also utilizes particle swarm optimization to enhance the image hosting the hidden message. In addition, a Genetic algorithm is used to generate the fittest hidden image from a selection of hidden messages [18].

Pandey (2020) proposes a novel approach used to enhance the security of medical data in transmission. The approach utilizes a bit mask-oriented genetic algorithm to encrypt data that is further embedded in images designed to be used in wavelet transformation [19].

Pramanik and Raja (2020) go on to suggest another approach that is designed to modify the genetic algorithm to improve its performance of the genetic algorithm. Further, the technique utilizes a hybrid approach that combines Discrete Ripplet Transform (DRT) and Fresnelet Transform (FT) to identify the best two sets of hidden messages [15]. The approach suggested by [20] is designed to be used for the purpose of enhancing the efficiency of the carrier image. In particular, a Genetic algorithm is strategically used to find the appropriate areas to embed the message. In addition, the technique incorporates the creation of a key file that is used to extract the hidden message.

Wazirali et al (2019) propose the development of a novel approach that is specifically designed to increase the least significant bits (LSB), which are used to match the carrier and the stego image. Further, Genetic algorithm is used to search for optimal solutions to ensure that the distortion in the carrier image is reduced regardless of the level of embedded capacity [21]. Al-Janabi, and Al-Shourbaji (2016) propose to develop a hybrid approach that is designed to hide more than one message simultaneously inside carrier images of the same size. In addition, a Genetic algorithm is used to generate a secret key as well as optimum matrix values [22].

The technique proposed by Shah and Bichkar (2017) utilizes the combination of a Genetic algorithm and a linear congruential generator. The approach is designed in such a way that the congruential generator and the algorithm are used to indent the coefficients matching the location where the message is to be hidden [23]. Finally, the authors in [24] propose the development of an approach that uses the combination of genetic algorithm and frequency domain to conceal messages in a cover image. Notably, the frequency domain is based on the concept of wavelet transform to identify the coefficients whereby the message will be embedded.

III. METHODOLOGY

In this paper, a new technique has been proposed to reduce the number of Least Significant Bits (LSB) that holds the secret message. This is done by using a random generator with a known seed. The random generator range is based on a dictionary index from a to z alphabet and include special characters [1-32] as shown in Table. I. The random generator will keep running until it hits the index of the first character of the secret message. Then, the number of runs (hits) will be held in the cover message for the first character. The procedure will continue until the secret message ends. This paper aims to find the optimal seed for the random generator such that the number of bits that will be used to hold the number of runs will be minimized. Which will reduce the

amount of effect of the secret message on the cover image. The optimal random generator seed will be found using the Genetic Algorithm. Fig.2 illustrates the full methodology for both encoding and decoding processes.

The encoding process will consist of the following:

- Dictionary Construction (the dictionary is fixed and is not a secret): the dictionary consists of the sequence of the alphabet as a key pair; the index and the character. The index of the character will be used as a target for the random generator in the encoding process instead of the character itself.
- The first 64 LSB from the first 64 bytes (16 pixels) will be reserved for the random generator seed.
- The next four LSB (4 bytes/ 1 pixel) will be used to hold the length of the number of flips (How many bits will be used to save the number of flips for each character).
- Apply the Genetic Algorithm (GA): the first population will be 100 random seeds. Each individual will be evaluated in terms of the maximum number of bits required to represent the number of runs of the random generator to target the index of a character. Equation 1 represents the fitness function calculation. The GA operators (selection, crossover, and mutation) will be used to produce the next generations. The output of the GA will be the individual that has the minimum fitness value. The value of the fitness function will be held in the four LSB of the previous step, after the seed value directly.

$$Number\ of\ bits = log_2\lfloor Max\ Counter \rfloor + 1 \quad (1)$$

- The solution produced by the GA will be used as a seed for the random generator. It will be used for both the encoding and decoding processes.
- Suppose the secret message is "Don't tell anyone!", then the random generator will be run until its value hit "4" with respect to the dictionary in Table I. A counter will hold the number of runs required to reach the index, then the value of the counter will be encoded in the cover image. The procedure will continue until the message end. When the message end, the last LSB of the predefined length specified in step 3 will hold zero only.

On the other hand, the decoding process has the following processes:

- Read the cover image and extract the random generator seed from the first 64 LSB.
- Extract the length of the counter.
- Run the random generator with the extracted seed with the counter value for each character.
- Retrieve the corresponding character from the dictionary based on the random generator value.
- Repeat the process until finding the zero value.

IV. RESULTS

A. Dataset

In this paper, the USC-SIPI Image Database has been used for evaluation purposes [25]. Six standard images have been selected that have been used in many research papers (Airplane, Baboon, Peppers, Sailboat, Tree and Splash) shown in Fig. 3. All images are bitmap images. The message

TABLE I. DICTIONARY EXAMPLE

Index	Char	Index	Char	Index	Char	Index	Char
1	a	9	i	17	q	25	y
2	b	10	j	18	r	26	z
3	c	11	k	19	s	27	Space
4	d	12	l	20	t	28	‘
5	e	13	m	21	u	29	!
6	f	14	n	22	v	30	(
7	g	15	o	23	w	31	)
8	h	16	p	24	x	32	-

payload or sizes used for evaluation are (1982, 3011, 3137, 3940, 5107, and 9914) bytes.

B. Performance Metrics

In this paper, the selected metrics used to evaluate our Steganography approaches are Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE), and Cosine Similarity. The following defines each metric with its corresponding formula:

- MSE: The Mean Square Error is an error metric used to represent the cumulative squared error between the compressed and the original image, the lower value of MSE is the lower error in the image. Equation 2 presents the formula used to calculate the MSE [26].

$$MSE = \sum_{M,N} \frac{|I_1(m,n) - I_2(m,n)|^2}{M*N} \quad (2)$$

Where M and N are the number of rows and columns in the input images.

- PSNR: Peak Signal to Noise Ratio commonly used to compute the quality of reconstructed images by compression. A high PSNR indicates that the reconstruction is a high quality [27]. The equation for PSNR is shown by Equation 3.

$$PSNR = 10log_{10}(\frac{255^2}{MSE}) \quad (3)$$

Where MSE is calculated using Equation 2.

- Cosine Similarity: is a measure to compute the similarity between two vectors by calculating the cosine of the angle between them [28]. Cosine Similarity Equation is presented in Equation 4.

$$Cosine\ Similarity = \frac{A\ B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i\ B_i}{\sqrt{\sum_{i=0}^{n} A_i}\sqrt{\sum_{i=1}^{n} B_i}} \quad (4)$$

Where A_i and B_i are components of vectors A and B respectively.

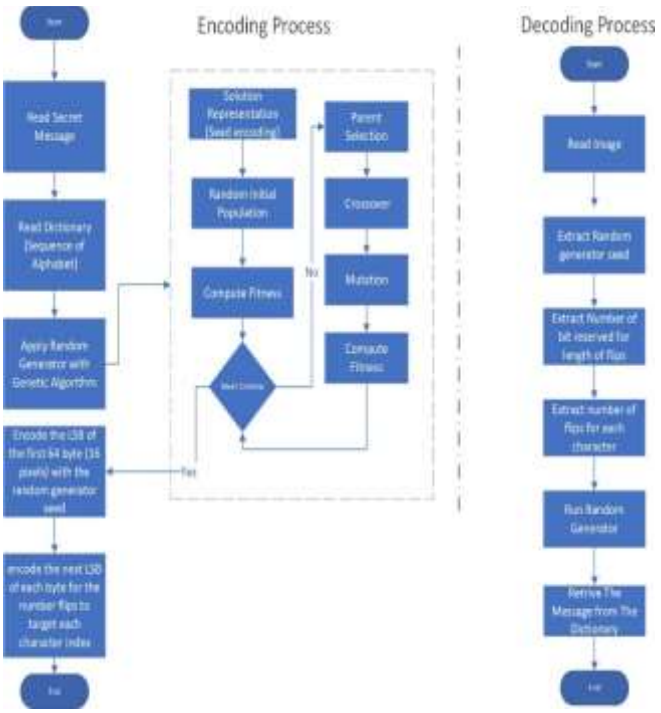


Fig. 2. Steganography Approach Workflow.

C. *Results Discussion*

In this section, the proposed approach has been evaluated and compared with the traditional Steganography technique based on LSB.

From Table II, it can be noticed that the experimental results turn out a noticeable enhancement for the proposed approach compared to the traditional approach in terms of MSE, PSNR, and COS. The results turn out that the proposed approach reduces MSE compare to the traditional approach in all data sets which are Airplane.tiff, Baboon.tiff, Peppers.tiff, Tree.tiff, Sailboat.tiff and Splash.tiff in all message sizes respectively.

Moreover, it can be noticed that as the size of the secret message increase the MSE value is increasing. Even though, the MSE value for the proposed is better than the traditional approach regardless of the message size.

PSNR reflects the quality of stego-image. It indicates that stego-image seems with high visual quality as the cover image. The proposed approach increases the PSNR in all datasets compared to the traditional approach.

Fig.3. Selected standard images.

V. CONCLUSION

In this paper, a random generator is used to generate the index of characters for the ”secret message” with respect to a predefined dictionary. A counter holds the required number of runs for the random generator to target the index. All counters are encoded in the cover image with the corresponding seed of the random generator instead of the message itself. A Genetic algorithm has been used to select the optimal seed for the random generator. An optimal seed generates the index of the characters with a minimal number of tries. Thus, reducing the number of bits that required to hide the message in the cover image. The proposed approach has been evaluated using six images from USC-SIPI dataset with various secret message sizes. The results shown that the proposed approach outperforms the traditional approach in terms of the MSE, PSNR, and Cosine Similarity.

REFERENCES

[1] S. Wendzel, L. Caviglione, W. Mazurczyk, A. Mileva, J. Dittmann, C. Kra¨tzer, K. Lamsho¨ft, C. Vielhauer, L. Hartmann, J. Keller et al., “A revised taxonomy of steganography embedding patterns,” in The 16th International Conference on Availability, Reliability and Security, 2021, pp. 1–12.

[2] O. AbuAlghanam, L. Albdour, and O. Adwan, “Multimodal biometric fusion online handwritten signature verification using neural network and support vector machine,” transactions, vol. 7, p. 8, 2021.

[3] M. Shyla, K. S. Kumar, and R. K. Das, “Image steganography using genetic algorithm for cover image selection and embedding,” Soft Computing Letters, vol. 3, p. 100021, 2021.

[4] Z. Abbas and M. Q. Saeed, “Image steganography using cryptographic primitives,” in 2021 International Conference on Cyber Warfare and Security (ICCWS). IEEE, 2021, pp. 124–131.

[5] M. Suresh and I. S. Sam, “Optimized interesting region identification for video steganography using fractional grey wolf optimization along with multi-objective cost function,” Journal of King Saud University-Computer and Information Sciences, 2020.

[6] A. A.-A. Gutub, "Adopting counting-based secret-sharing for e-video watermarking allowing fractional invalidation," Multimedia Tools and Applications, vol. 81, no. 7, pp. 9527–9547, 2022.

[7] O. AbuAlghanam, M. Qatawneh, W. Almobaideen, and M. Saadeh, "A new hierarchical architecture and protocol for key distribution in the context of iot-based smart cities," Journal of Information Security and Applications, vol. 67, p. 103173, 2022.

[8] O. Abualghanam, M. Qatawneh, and W. Almobaideen, "A survey of key distribution in the context of internet of things," Journal of Theoretical and Applied Information Technology, vol. 97, no. 22, pp. 3217–3241, 2019.

[9] N. Hamid, A. Yahya, R. B. Ahmad, and O. M. Al-Qershi, "Image steganography techniques: an overview," International Journal of Com- puter Science and Security (IJCSS), vol. 6, no. 3, pp. 168–187, 2012.

[10] P. Mehta, A. Nair, S. Edappilly, K. Manik, and S. Sahu, "A comprehensive study of ai-based steganalysis techniques on image and text documents," in Advances in Data and Information Sciences. Springer, 2022, pp. 53–63.

[11] T. AlKhodaidi and A. Gutub, "Refining image steganography distribution for higher security multimedia counting-based secret-sharing," Multimedia Tools and Applications, vol. 80, no. 1, pp. 1143–1173, 2021.

[12] B. Lakshmi Sirisha and B. Chandra Mohan, "Review on spatial domain image steganography techniques," Journal of Discrete Mathematical Sciences and Cryptography, vol. 24, no. 6, pp. 1873–1883, 2021.

[13] A. Hamza, D. Shehzad, M. S. Sarfraz, U. Habib, and N. Shafi, "Novel se- cure hybrid image steganography technique based on pattern matching," KSII Transactions on Internet and Information Systems (TIIS), vol. 15, no. 3, pp. 1051–1077, 2021.

[14] S. Pramanik, D. Samanta, S. K. Bandyopadhyay, and R. Ghosh, "A new combinational technique in image steganography," International Journal of Information Security and Privacy (IJISP), vol. 15, no. 3, pp. 48–64, 2021.

[15] S. Pramanik and S. S. Raja, "A secured image steganography using genetic algorithm," Advances in Mathematics: Scientific Journal, vol. 9, no. 7, pp. 4533–4541, 2020.

[16] O. F. A. Wahab, A. A. Khalaf, A. I. Hussein, and H. F. Hamed, "Hiding data using efficient combination of rsa cryptography, and compression steganography techniques," IEEE Access, vol. 9, pp. 31 805–31 815, 2021.

[17] P. D. Shah and R. S. Bichkar, "Secret data modification based image steganography technique using genetic algorithm having a flexible chromosome structure," Engineering Science and Technology, an Inter- national Journal, vol. 24, no. 3, pp. 782–794, 2021.

[18] S. Pramanik, R. Singh, and R. Ghosh, "Application of bi-orthogonal wavelet transform and genetic algorithm in image steganography," Multimedia Tools and Applications, vol. 79, no. 25, pp. 17 463–17 482, 2020.

[19] H. M. Pandey, "Secure medical data transmission using a fusion of bit mask oriented genetic algorithm, encryption and steganography," Future Generation Computer Systems, vol. 111, pp. 213–225, 2020.

[20] M. Nosrati, A. Hanani, and R. Karimi, "Steganography in image segments using genetic algorithm," in 2015 Fifth International Conference on Advanced Computing & Communication Technologies. IEEE, 2015, pp. 102–107.

[21] R. Wazirali, W. Alasmary, M. M. Mahmoud, and A. Alhindi, "An optimized steganography hiding capacity and imperceptibly using genetic algorithms," IEEE Access, vol. 7, pp. 133 496–133 508, 2019.

[22] S. Al-Janabi and I. Al-Shourbaji, "A hybrid image steganography method based on genetic algorithm," in 2016 7th International Conference on Science of Electronics, Technologies of Information and Telecommuni- cations (SETIT). IEEE, 2016, pp. 398–404.

[23] P. D. Shah and R. Bichkar, "A secure spatial domain image steganography using genetic algorithm and linear congruential generator," in International conference on intelligent computing and applications. Springer, 2018, pp. 119–129.

[24] A. Miri and K. Faez, "Adaptive image steganography based on transform domain via genetic algorithm," Optik, vol. 145, pp. 158–168, 2017.

[25] A. G. Weber, "The usc-sipi image database: Version 5," http://sipi. usc. edu/database/, 2006.

[26] B. Pattanaik, P. Chitra, H. Lakshmi, G. T. Selvi, and V. Nagaraj, "Contrasting the performance metrics of discrete transformations on digital image steganography using artificial intelligence," Materials Today: Proceedings, 2021.

[27] D. R. I. M. Setiadi, "Psnr vs ssim: imperceptibility quality assessment for image steganography," Multimedia Tools and Applications, vol. 80, no. 6, pp. 8423–8444, 2021.

[28] F. Li, H. Tang, Y. Zou, Y. Huang, Y. Feng, and L. Peng, "Research on information security in text emotional steganography based on machine learning," Enterprise Information Systems, vol. 15, no. 7, pp. 984–1001, 2021.

Table II. THE RESULTS OF THE TRADITIONAL LSB AND THE PROPOSED APPROACH MEASURING THE DIFFERENCE BETWEEN THEM.

		Traditional Approach			Proposed Approach			Difference Metrics		
	Message Size	**MSE**	**PSNR**	**COS**	**MSE**	**PSNR**	**COS**	**MSE**	**PSNR**	**COS**
Airplane.tiff	1982	0.010011037	68.12601286	0.999998897	0.005118052	71.03975653	0.999999436	-0.00489	-2.91374	-5.38879E-07
	3011	0.015293121	66.28584227	0.999998316	0.00764211	69.29867066	0.999999158	-0.00765	-3.01283	-8.42628E-07
	3137	0.01617686	66.04186146	0.999998218	0.008055369	69.06994918	0.999999113	-0.00812	-3.02809	-8.94445E-07
	3940	0.020098368	65.0991956	0.999997787	0.010075887	68.09797071	0.99999889	-0.01002	-2.99878	-1.10381E-06
	5107	0.026088715	63.96627679	0.999997127	0.013117472	66.95230204	0.999998555	-0.01297	-2.98603	-1.42857E-06
	9914	0.05048879	61.09885399	0.99999444	0.025295258	64.10041255	0.999997214	-0.02519	-3.00156	-2.77465E-06
Baboon.tiff	1982	0.010110219	68.08319784	0.999999318	0.004981995	71.15677106	0.999999664	-0.00513	-3.07357	-3.46092E-07
	3011	0.015338898	66.2728621	0.999998965	0.007776896	69.22274061	0.999999475	-0.00756	-2.94988	-5.10334E-07
	3137	0.01608785	66.06582354	0.999998914	0.008059184	69.06789303	0.999999456	-0.00803	-3.00207	-5.41831E-07
	3940	0.019959768	65.12924879	0.999998653	0.010175069	68.0554299	0.999999313	-0.00978	-2.92618	-6.60343E-07
	5107	0.026048024	63.97305569	0.999998242	0.013121287	66.95103925	0.999999114	-0.01293	-2.97798	-8.72389E-07
	9914	0.050566355	61.09218708	0.999996587	0.025246938	64.10871646	0.999998296	-0.02532	-3.01653	-1.70874E-06
Peppers.tiff	1982	0.010032654	68.11664534	0.999999121	0.005139669	71.02145228	0.99999955	-0.00489	-2.90481	-4.28486E-07
	3011	0.015293121	66.28584227	0.999998661	0.007785797	69.21777278	0.999999318	-0.00751	-2.93193	-6.57422E-07
	3137	0.015920003	66.11137209	0.999998606	0.007998149	69.10090892	0.9999993	-0.00792	-2.98954	-6.93728E-07
	3940	0.020347595	65.04567271	0.999998218	0.010206858	68.04188275	0.999999106	-0.01014	-2.99621	-8.88038E-07
	5107	0.026004791	63.98026989	0.999997723	0.012987773	66.99545684	0.999998863	-0.01302	-3.01519	-1.13992E-06
	9914	0.050395966	61.1068459	0.999995587	0.025244395	64.10915395	0.999997789	-0.02515	-3.00231	-2.20256E-06
Sailboat.tiff	1982	0.009883881	68.1815287	0.999999089	0.005054474	71.09404405	0.999999534	-0.00483	-2.91252	-4.45028E-07
	3011	0.01539739	66.25633258	0.999998581	0.007612864	69.31532279	0.999999298	-0.00778	-3.05899	-7.17332E-07
	3137	0.016091665	66.06479388	0.999998517	0.00802358	69.08712178	0.999999261	-0.00807	-3.02233	-7.43467E-07
	3940	0.020023346	65.1154371	0.999998155	0.010079702	68.09632679	0.999999071	-0.00994	-2.98089	-9.16301E-07
	5107	0.02576828	64.01994929	0.999997625	0.013056437	66.97255678	0.999998797	-0.01271	-2.95261	-1.17139E-06
	9914	0.050502777	61.09765101	0.999995346	0.02521642	64.11396923	0.999997676	-0.02529	-3.01632	-2.33012E-06
Spalsh.tiff	1982	0.010064443	68.10290618	0.999998704	0.005062103	71.0874936	0.999999348	-0.005	-2.98459	-6.44306E-07
	3011	0.015500387	66.22737832	0.999998004	0.007728577	69.24980842	0.999999005	-0.00777	-3.02243	-1.00103E-06
	3137	0.015968323	66.09821059	0.999997943	0.007979075	69.11127807	0.999998972	-0.00799	-3.01307	-1.02903E-06
	3940	0.020001729	65.12012815	0.999997424	0.010144552	68.06847506	0.999998693	-0.00986	-2.94835	-1.26962E-06
	5107	0.025847117	64.0066825	0.999996671	0.012982686	66.99715795	0.999998328	-0.01286	-2.99048	-1.65695E-06
	9914	0.050633748	61.08640282	0.999993478	0.025494893	64.06627161	0.999996716	-0.02514	-2.97987	-3.23782E-06
Tree.tiff	1982	0.040267944	62.08120901	0.999996243	0.020421346	65.02995997	0.999998094	-0.01985	-2.94875	-1.85193E-06
	3011	0.061538696	60.23932069	0.999994258	0.030385335	63.30416328	0.999997165	-0.03115	-3.06484	-2.90696E-06
	3137	0.064015706	60.06793819	0.999994027	0.031585693	63.13589946	0.999997053	-0.03243	-3.06796	-3.02607E-06
	3940	0.079966227	59.10173754	0.999992538	0.04002889	62.10706814	0.999996265	-0.03994	-3.00533	-3.72655E-06
	5107	0.103647868	57.97519988	0.999990328	0.052047729	60.96678572	0.999995143	-0.0516	-2.99159	-4.81488E-06
	9914	0.202067057	55.07584844	0.999981145	0.101308187	58.07435818	0.999990547	-0.10076	-2.99851	-9.4019E-06

Generation of Electrical Energy Through Human Traction on a Stationary Bicycle

C L Sandoval-Rodriguez
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
csandoval@correo.uts.edu.co

C A Ángulo-Julio
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
carlosaj@correo.uts.edu.co

O Lengerke
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
olengerke@correo.uts.edu.co

A D Rincón-Quintero
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
arincon@correo.uts.edu.co

A Rodriguez
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
alvarojrodriguez@correo.uts.edu.co

M A Castellanos-Carreño
Faculty of natural sciences and engineering
Unidades Tecnológicas de Santander
Bucaramanga, Colombia
malexandracastellanos@uts.edu.co

Abstract— **The exponential increase in energy demand in recent years and the environmental problems associated with the generation of energy from traditional sources such as coal, oil and gas have caused a global crisis that has motivated the development of alternative energy systems. Despite having one of the largest renewable energy generation infrastructures (hydroelectric) in Latin America, Colombia is no stranger to this energy crisis due to intense droughts caused by natural phenomena that directly affect energy production in the country. Here, we proposed to develop a system capable of generating electrical energy from mechanical energy through human traction when using a stationary bicycle in spinning class.**

Keywords—Energy generation, Renewable energy

I. INTRODUCTION

Globally, nearly 790 million people live without electricity, and another 3 billion use polluting fuels such as firewood or other biomass to cook or heat their homes. Access to affordable, reliable, sustainable, and modern energy is essential to ending poverty, and essential for many countries to meet their climate change mitigation goals [1–4].

Colombia has not been indifferent to this problem despite having a hydro generation infrastructure capable of contributing 75.3% of the national demand. This is how it has developed a regulatory framework to promote the development of renewable energies. Additionally, the country joined the Kyoto Protocol of the United Nations Framework Convention on Climate Change [5–10]. Therefore, renewable and alternative energies became a strategic option for Colombia.

In accordance with the above, to counteract the global energy crisis that is occurring due to the exponential increase in demand that has been generated in recent years, we propose the development of an integrated system for the conversion, storage, and distribution of clean energy within an educational institution [11–14].

In accordance with the above, in order to counteract the global energy crisis that is occurring due to the exponential increase in demand that has been generated in recent years, we propose the development of an integrated system for the conversion, storage and distribution of clean energy within a educational institution. This is how an electromechanical system was developed capable of capturing the mechanical energy generated using the spin bikes in the institution's gym and transforming it into electrical energy [15–18].

The system was designed in such a way that it is not necessary to modify or make adaptations to the spin bike, which makes it possible for the system to be used on any other spin bike. Likewise, the system will not affect the normal use of the bicycle for exercise due to its compactness and strategic location [19–23].

II. DESIGN

The operation of the generation system is activated by the force that the user gives to the pedals of the bike. The mechanical power generated by the user in the primary phase of the system is transmitted through the bike chain to the flywheel and then to the alternator shaft. The average power P developed by a person is around 314.4 Watts, which will be used by a 12 volts generator [24,25]. For this system, the expected current is

$$I = \frac{P}{V} = \frac{314.4\ \mathrm{W}}{12\ V} = 26.2 \quad (1)$$

A. *Primary mechanical* power *analysis*

On average, a person exerts about 30 Kg when pedaling horizontally. When pedaling, the perpendicular force F_A and torque T_A applied to the arm of a crank of length $d = 190$ mm is

$$F_A = m \cdot g = 30\ kg \cdot 9.81\ \tfrac{m}{s^2} = 294.3\ N \quad (2)$$

$$T_A = F_A \cdot d = 294.3\ \text{N} \cdot 0.19\ m = 55.92\ \text{Nm}. \quad (3)$$

Riding a spin bike at high cadences with little or no resistance does not adequately train the neuromuscular system [26]–[29]. Also, it is extremely difficult to maintain a high cadence (revolutions per minute [rpm] of the pedal cranks) while pedaling on the road. The most efficient cadence on hills is between 60 and 80 rpm.

If an average speed $V = 60$ rpm is achieved, the mechanical power P_M injected into the system would be

$$P_M = \tfrac{2\pi}{60} V \cdot T_A = \tfrac{2\pi}{60} 60 \cdot 55.92 = 351.34\ W. \quad (4)$$

With these values, the reactive forces can be obtained and the static in the main chain transmission system can be obtained from the drive sprocket or plate to the driven sprocket or gear that transmits the mechanical power to the flywheel (see Fig. 1).

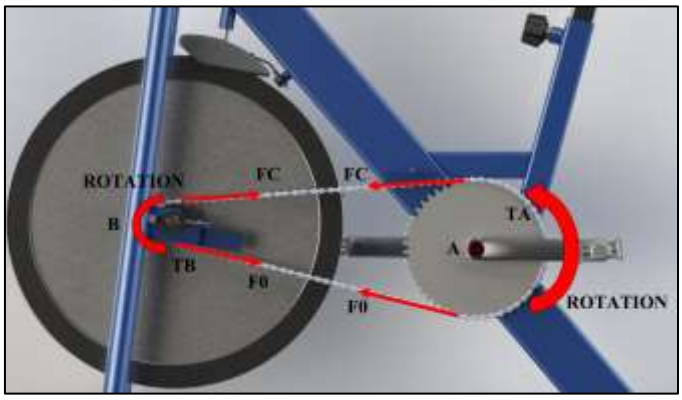


Fig. 1. Primary mechanical power analysis

The force F_C exerted on the chain, for a chainring of radius $r_A = 105$ mm, is

$$F_C = \frac{T_A}{r_A} = \frac{55.92}{0.105} = 532.5\ \text{N} \quad (5)$$

The torque T_B exerted on a gear of radius $r_B = 35$ mm is

$$T_B = F_C \cdot r_B = 532.5 \cdot 0.035 = 18.64\ \text{N} \cdot \text{m} \quad (6)$$

B. *Secondary mechanical power* analysis

For the transmission of secondary mechanical power (from the bicycle to the electric power generation system), two options were studied. The first option contemplates the use of belts and pulleys for the transmission while the second option contemplates the use of direct contact transmission by means of pulleys. It was decided to choose the second option since it uses fewer elements, reducing the percentage of transmission losses and the number of parts that need to be replaced due to wear. For the selection of the contact element, it was decided to use the rubber used for the manufacture of truck tires. This is a material with high adhesion capacity, resistance to high temperatures and wear, and capable of withstanding the mechanical power transmitted, the torque generated and the repulsion forces that are exerted on it [30], [31].

The magnitude of the torque in the driven sprocket or gear T_B is the same magnitude of the torque exerted on the flywheel T_C connected by the same shaft, so $T_C = T_B = 18.64\ \text{N} \cdot \text{m}$.

The drag force F_{dr} and torque T_D that will act on the contact element (with radius r_c=220 mm) coupled to the alternator shaft are

$$F_{dr} = \frac{T_C}{r_c} = \frac{18.64}{0.22} = 84.73\ N \quad (7)$$

$$T_D = F_{dr} \cdot 2r_c = 84.73 \cdot 2 \cdot 0.22 = 37.28\ N \cdot m \quad (8)$$

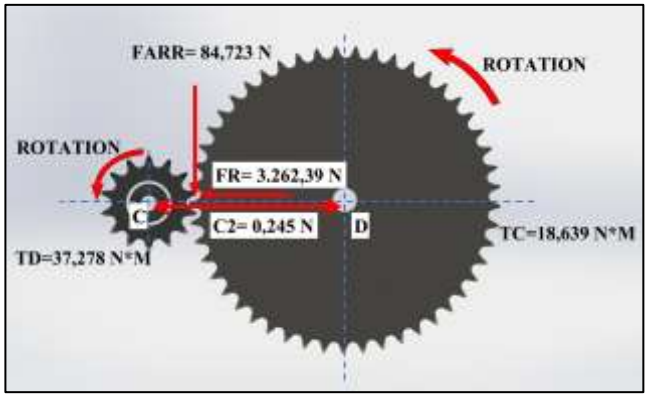


Fig. 2. Secondary mechanical power analysis.

C. *Electrical system*

As shown in Fig. 3, the electrical system is divided into four subsystems: generation, rectification, storage, and investment.

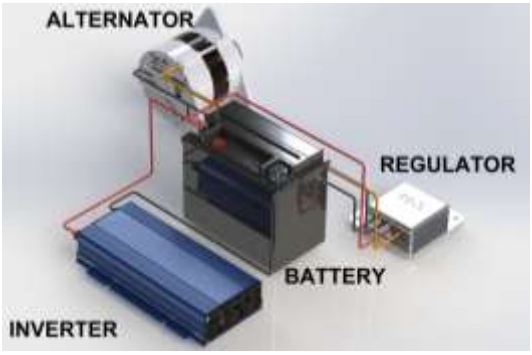


Fig. 3. Electrical system

The first subsystem considers the use of a generator with the capacity to take the mechanical power supplied by the user, i.e., $P_M = 351.34$ W. Taking into account the results in [32] and [33] , to carry out the transformation of mechanical energy into electrical energy, it was defined to use a Bosch KCB1 alternator [34], [35]. This 14 V alternator reaches currents of 50 A at speeds around 1500 rpm.

The rectification subsystem has a rectifier protected against short circuit, which together with the alternator generate 12.6 V.

In the third subsystem (storage) there is a 12 V – 600 A battery used for energy storage in vehicles and in the pre-excitation circuit. This battery will have the function of storing the duly rectified electrical energy provided by the alternator and supplying energy to the inverter [36].

The loads that could be fed with the energy stored in the battery are shown in TABLE IX. Considering a growth of 50%, it was decided to use a ISOOW inverter of 1500 W, that performs the DC-AC conversion by pulse width modulation. This inverter connects directly to the 12 VDC battery and provides a 110 VAC – 60 Hz output as well as a 5 VDC output.

TABLE IX. LOAD

Type of load	**Voltage [V]**	**Current [A]**	**Power [W]**
Lighting	110	1.45	160
Ventilation	110	0.73	80
Video	110	0.50	55
Audio	110	2.73	300
Electrical outlets	110	3.68	405
		TOTAL	**1000**

III. IMPLEMENTATION

To implement the proposed system, it was necessary to manufacture metal-mechanical elements for fastening and coupling it to the structure of the spinning bike. The structure is shown in Fig. 4, which is composed of a base and a coupling bridge.

Fig. 4. Structure of the system.

The base (see Fig. 5) allows the coupling and support of the designed elements and subsystems, as well as the support for the transmitted mechanical power, the generated torsional moment and the repulsive force that is exerted on the system.

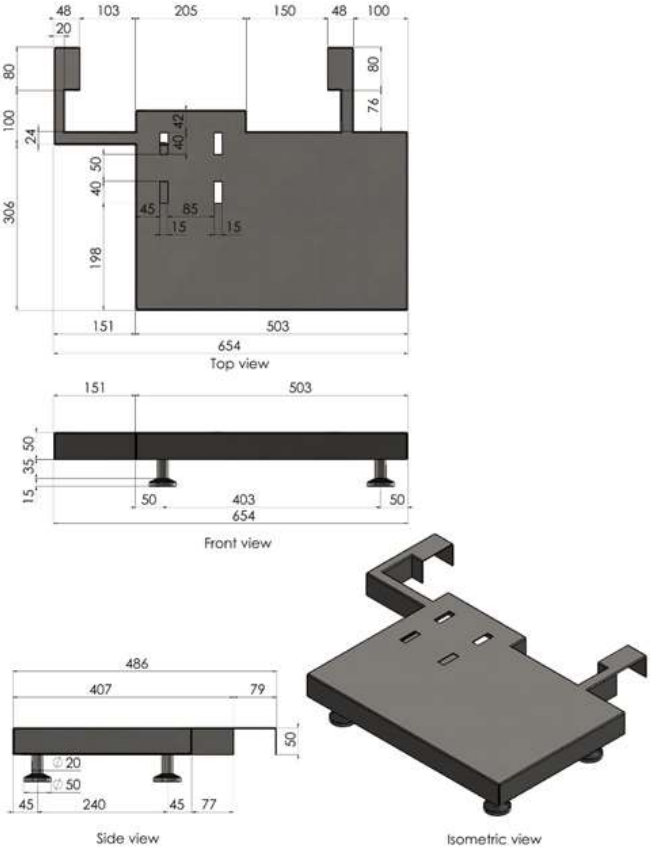


Fig. 5. Base of the system.

The coupling bridge (Fig. 6) allows the assembly of the generator, which in turn allows a movement parallel to the base, adapting to the different dimensions that may occur on spin bikes without losing its functionality.

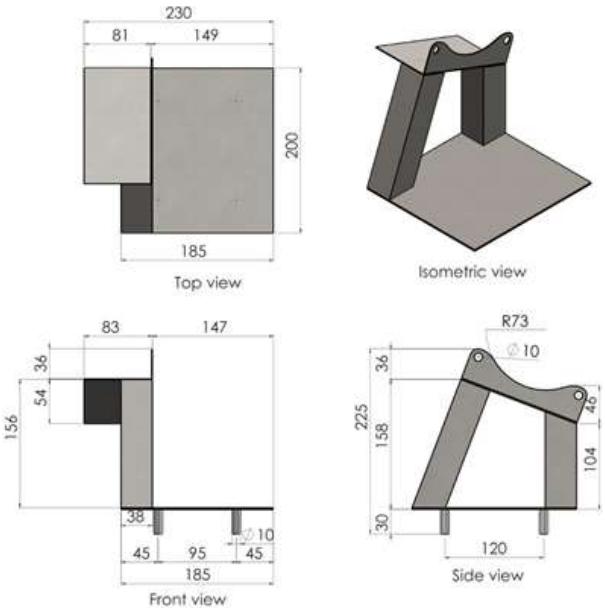


Fig. 6. Coupling bridge.

Furthermore, the contact flywheel for the alternator shaft which get into contact with the spinning bike flywheel is shown in Ref. 36.

Fig. 7. Contact flywheel.

IV. RESULTS

Some experiments were carried out by controlling the speed of the alternator V_D. The current obtained in each test is shown in TABLE X.

TABLE X. CURRENT

V_D [rpm]	Test 1	Test 2	Test 3	Test 4	Test 5	Power [Watts]
700	-2.5	-2.4	-2.2	-2.4	-2.5	-31.68
900	-2.4	-2.3	-2.1	-2.2	-2.3	-29.83
1100	-2.4	-1.8	-1.9	-1.7	-2.2	-26.40
1300	-2.1	-1.3	-1.5	-1.2	-1.5	-20.06
1500	1.0	2.1	1.7	1.9	1.0	20.33
1700	3.1	3.1	3.4	3.8	2.4	41.71
1900	4.3	4.1	4.8	4.9	3.6	57.29
2100	5.6	5.3	5.6	5.7	4.9	71.54
2300	6.2	6.2	6.8	6.8	6.2	85.01
2500	7.5	7.9	7.5	7.6	7.7	99.32
2700	8.5	8.7	8.7	8.1	8.2	111.41
2900	9.1	9.6	9.4	9.2	8.9	121.97
3100	10.0	10.0	10.0	10.0	10.0	132.00

Fig. 8 shows the average generation curve which relates the energy potential obtained using the spinning bike in conjunction with the designed system. The maximum theoretical speed of the alternator is 6000 rpm, but in user tests a maximum of 3100 rpm was achieved due to the resistance imposed by the magnetic field of the alternator, obtaining approximately 10 A and 13.2V.

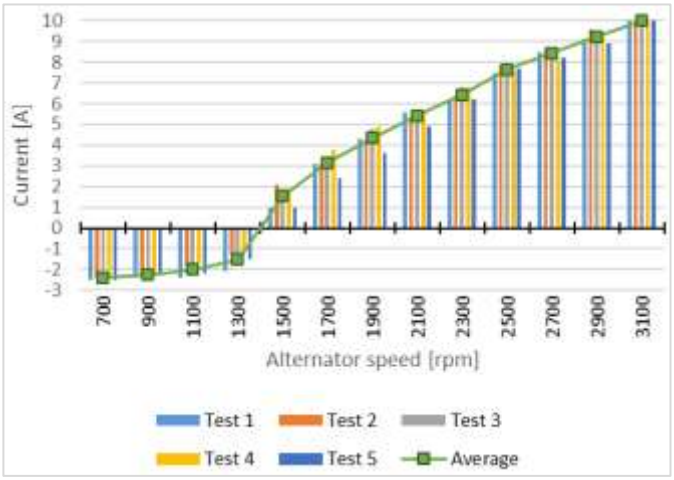


Fig. 8. Average generation curve.

From there it can be established that the system begins to generate electrical energy when the alternator reaches speeds between 1300 and 1500 rpm. Given the transmission ratio of the system is 26.4, these speeds are reached when the user of the bike achieves cadences between 49.2 and 56.8 rpm.

During the tests, the users were able to maintain cadences around 95 rpm (equivalent to 2500 rpm in the alternator), thus obtaining an average of 7.64A and 13V, which represents a generated power $P = 99.32$ W.

Each of the spinning bikes in the gym can be used for 10 hours a day, so the daily developed energy potential is 993.2 W·h

V. CONCLUSION

Here we develop a system capable of generating electrical energy from mechanical energy through human traction from a stationary bicycle. This type of system was considered in a previous work as an option for the transformation of mechanical energy to electrical energy in a closed space within the UTS.

The energy generated using the system developed here represents a clean, renewable, pollution-free resource, inexhaustible in nature and available to everyone. The system generates around 990 Wh per day, given that during the semester the gym works a total of 960 hours, the energy potential developed per semester will be around 950 kWh. This way, having a system capable of converting, storing, and distributing a free energy potential and with applicability in the institution allows the design of a closed space capable of being electrically self-sustaining.

REFERENCES

[1] P. Ribeiro, F. Fonseca, and T. Meireles, "Sustainable mobility patterns to university campuses: Evaluation and constraints," *Case Stud. Transp. Policy*, vol. 8, no. 2, pp. 639–647, 2020, doi: https://doi.org/10.1016/j.cstp.2020.02.005.

[2] J. Rich, A. F. Jensen, N. Pilegaard, and M. Hallberg, "Cost-benefit of bicycle infrastructure with e-bikes and cycle superhighways," *Case Stud. Transp. Policy*, vol.

9, no. 2, pp. 608–615, 2021, doi: https://doi.org/10.1016/j.cstp.2021.02.015.

[3] M. A. Kwiatkowski, "Regional bicycle-sharing system in the context of the expectations of small and medium-sized towns," *Case Stud. Transp. Policy*, vol. 9, no. 2, pp. 663–673, 2021, doi: https://doi.org/10.1016/j.cstp.2021.03.004.

[4] M. A. Durán-Sarmiento, L. A. del Portillo-Valdés, Y. J. Rueda-Ordoñez, C. Borrás-Pinilla, and D. C. Dulcey-Diaz, "Modeling and simulation of a braking energy regeneration system in hydraulic hybrid vehicles in the Colombian topography," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 755–766, 2021, doi: 10.21533/pen.v9i4.1986.

[5] J. Zhang, M. Meng, P. P. Koh, and Y. D. Wong, "Life duration of bike sharing systems," *Case Stud. Transp. Policy*, vol. 9, no. 2, pp. 674–680, 2021, doi: https://doi.org/10.1016/j.cstp.2021.03.005.

[6] N. Popovich, E. Gordon, Z. Shao, Y. Xing, Y. Wang, and S. Handy, "Experiences of electric bicycle users in the Sacramento, California area," *Travel Behav. Soc.*, vol. 1, no. 2, pp. 37–44, 2014, doi: https://doi.org/10.1016/j.tbs.2013.10.006.

[7] J. Sung, N. Ba Hung, S. Yoon, and O. Lim, "A study of the dynamic characteristics and required power of an electric bicycle equipped with a semi-automatic transmission," *Energy Procedia*, vol. 142, pp. 2057–2064, 2017, doi: https://doi.org/10.1016/j.egypro.2017.12.410.

[8] N. Ba Hung, S. Jaewon, and O. Lim, "A study of the effects of input parameters on the dynamics and required power of an electric bicycle," *Appl. Energy*, vol. 204, pp. 1347–1362, 2017, doi: https://doi.org/10.1016/j.apenergy.2017.03.025.

[9] B. E. Tarazona-Romero, Y. A. Muñoz-Maldonado, A. Campos-Celador, and O. Lenguerke-Pérez, "Optical performance assessment of a handmade prototype of linear Fresnel concentrator," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 795–811, 2021, doi: 10.21533/pen.v9i4.1987.

[10] J. G. M. Lázaro and C. L. S.- Rodríguez, "Design and set up of a pulverized panela machine," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 812–828, 2021.

[11] M. A. Halim, R. Rantz, Q. Zhang, L. Gu, K. Yang, and S. Roundy, "An electromagnetic rotational energy harvester using sprung eccentric rotor, driven by pseudo-walking motion," *Appl. Energy*, vol. 217, pp. 66–74, 2018, doi: https://doi.org/10.1016/j.apenergy.2018.02.093.

[12] N. Ba Hung and O. Lim, "The effects of operating conditions and structural parameters on the dynamic, electric consumption and power generation characteristics of an electric assisted bicycle," *Appl. Energy*, vol. 247, pp. 285–296, 2019, doi: https://doi.org/10.1016/j.apenergy.2019.04.002.

[13] Y. Yang, Y. Pian, and Q. Liu, "Design of energy harvester using rotating motion rectifier and its application on bicycle," *Energy*, vol. 179, pp. 222–231, 2019, doi: https://doi.org/10.1016/j.energy.2019.05.036.

[14] A. D. Rincón-Quintero *et al.*, "Manufacture of hybrid pieces using recycled R-PET, polypropylene PP and cocoa pod husks ash CPHA, by pneumatic injection controlled with LabVIEW Software and Arduino Hardware," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 844, no. 1, 2020, doi: 10.1088/1757-899X/844/1/012054.

[15] L. Bergmann, S. Leonhardt, D. Greven, and B. J. E. Misgeld, "Optimal assistive control of a pedal-electric drive unit," *Control Eng. Pract.*, vol. 110, p. 104765, 2021, doi: https://doi.org/10.1016/j.conengprac.2021.104765.

[16] D. Bucher, R. Buffat, A. Froemelt, and M. Raubal, "Energy and greenhouse gas emission reduction potentials resulting from different commuter electric bicycle adoption scenarios in Switzerland," *Renew. Sustain. Energy Rev.*, vol. 114, p. 109298, 2019, doi: https://doi.org/10.1016/j.rser.2019.109298.

[17] M. A. Heldeweg and Séverine Saintier, "Renewable energy communities as 'socio-legal institutions': A normative frame for energy decentralization?," *Renew. Sustain. Energy Rev.*, vol. 119, p. 109518, 2020, doi: https://doi.org/10.1016/j.rser.2019.109518.

[18] B. E. T. Romero, A. C. Celador, C. L. S. Rodriguez, J. G. A. Villabona, and A. D. R. Quintero, "Design and construction of a solar tracking system for linear fresnel concentrator," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 778–794, 2021, doi: 10.21533/pen.v9i4.1988.

[19] K. Kazemzadeh and P. Bansal, "Electric bike navigation comfort in pedestrian crowds," *Sustain. Cities Soc.*, vol. 69, p. 102841, 2021, doi: https://doi.org/10.1016/j.scs.2021.102841.

[20] M. Siman-Tov *et al.*, "A look at electric bike casualties: Do they differ from the mechanical bicycle?," *J. Transp. Heal.*, vol. 11, pp. 176–182, 2018, doi: https://doi.org/10.1016/j.jth.2018.10.013.

[21] A. Söderberg f.k.a. Andersson, E. Adell, and L. Winslott Hiselius, "What is the substitution effect of e-bikes? A randomised controlled trial," *Transp. Res. Part D Transp. Environ.*, vol. 90, p. 102648, 2021, doi: https://doi.org/10.1016/j.trd.2020.102648.

[22] G. Piazza, S. Bracco, F. Delfino, and S. Siri, "Optimal design of electric mobility services for a Local Energy Community," *Sustain. Energy, Grids Networks*, vol. 26, p. 100440, 2021, doi: https://doi.org/10.1016/j.segan.2021.100440.

[23] E. A. C. Quintana and B. E. Tarazona, "Characterization of mechanical vibrations in a metal structure using the transform Cepstrum," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 767–777, 2021.

[24] L. Celentano, D. Iannuzzi, and L. Rubino, "Detailed continuous and discrete–time models and experimental validation to design a power charging station for e-bike clever mobility," *Electr. Power Syst. Res.*, vol. 147, pp. 115–132, 2017, doi: https://doi.org/10.1016/j.epsr.2017.01.031.

[25] R. Zaripov and P. Gavrilovs, "Study of dynamic characteristics of electric bicycles," *Procedia Comput. Sci.*, vol. 149, pp. 307–313, 2019, doi: https://doi.org/10.1016/j.procs.2019.01.140.

[26] D. Iannuzzi, L. Rubino, L. P. Di Noia, G. Rubino, and P. Marino, "Resonant inductive power transfer for an E-bike charging station," *Electr. Power Syst. Res.*, vol. 140, pp. 631–642, 2016, doi: https://doi.org/10.1016/j.epsr.2016.05.010.

[27] W. Liu, H. Liu, W. Liu, and Z. Cui, "Life cycle assessment of power batteries used in electric bicycles in China," *Renew. Sustain. Energy Rev.*, vol. 139, p. 110596, 2021, doi: https://doi.org/10.1016/j.rser.2020.110596.

[28] N. B. Hung and O. Lim, "A review of history, development, design and research of electric bicycles," *Appl. Energy*, vol. 260, p. 114323, 2020, doi: https://doi.org/10.1016/j.apenergy.2019.114323.

[29] L. Betancur-Arboleda, P. Hulse, K. G. Domiciano, L. Krambeck, and M. Mantelli, "Experimental study of the thermal performance of a PCM in heat sinks," *Period. Eng. Nat. Sci.*, vol. 9, no. 4, pp. 744–754, 2021, doi: 10.21533/pen.v9i4.1991.

[30] M. Nematchoua, C. Deuse, M. Cools, and S. Reiter, "Evaluation of the potential of classic and electric bicycle commuting as an impetus for the transition towards environmentally sustainable cities: A case study of the university campuses in Liege, Belgium," *Renew. Sustain. Energy Rev.*, vol. 119, p. 109544, 2020, doi: https://doi.org/10.1016/j.rser.2019.109544.

[31] N. B. Hung, J. Sung, and O. Lim, "A simulation and experimental study of operating performance of an electric bicycle integrated with a semi-automatic transmission," *Appl. Energy*, vol. 221, pp. 319–333, 2018, doi: https://doi.org/10.1016/j.apenergy.2018.03.195.

[32] H. Zhu and Z. Pei, "Data-driven layout design of regional battery swapping stations for electric bicycles," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 13–18, 2020, doi: https://doi.org/10.1016/j.ifacol.2021.04.078.

[33] L. T. Hieu, N. X. Khoa, and O. T. Lim, "An investigation on the effective performance area of the electric bicycle with variable key input parameters," *J. Clean. Prod.*, vol. 321, p. 128862, 2021, doi: https://doi.org/10.1016/j.jclepro.2021.128862.

[34] A. Bigazzi and K. Wong, "Electric bicycle mode substitution for driving, public transit, conventional cycling, and walking," *Transp. Res. Part D Transp. Environ.*, vol. 85, p. 102412, 2020, doi: https://doi.org/10.1016/j.trd.2020.102412.

[35] S. Johny, S. S. Kakkattil, S. Sunny, K. S. Sandeep, and V. Sankar, "Design and fabrication of foldable electric bicycle," *Mater. Today Proc.*, vol. 46, pp. 9646–9651, 2021, doi: https://doi.org/10.1016/j.matpr.2020.07.157.

[36] N. B. Hung and O. Lim, "A simulation and experimental study of dynamic performance and electric consumption of an electric bicycle," *Energy Procedia*, vol. 158, pp. 2865–2871, 2019, doi: https://doi.org/10.1016/j.egypro.2019.01.937.

Analysis of IoT-Blockchain Technologies Integration into the Healthcare Ecosystem

Bassey Isong
Computer Science Dept.
North-West University
Mafikeng, South Africa
isong.bassey@ieee.org

Tshipuke Vhahangwele
Computer Science Dept.
North-West University
Mafikeng, South Africa
tshipukevhahangwele@gmail.com

Koketso Ntshabele
Computer Science Dept.
North-West University
Mafikeng, South Africa
koketso.ntshabele@nwu.ac.za

Adnan Abu-Mahfouz
Council for Scientific and
Industrial Research (CSIR)
Pretoria, South Africa
a.abumahfouz@ieee.org

***Abstract*—The healthcare ecosystem has experienced explosive development and transformations in recent years due to increasing advancements in technologies such as IoT and blockchain. This has resulted in efficiency, convenience, and cost-effectiveness as well as minimum resource usage. This paper performed a review of several studies on IoT and blockchain integration in the healthcare system to comprehend the solutions offered and the challenges that need attention. We found the existence of several solutions used for continuous monitoring, treatment, and diagnosis of patients remotely. Several techniques were utilized to protect healthcare data's security and privacy in transit or storage. Also, several issues were identified that requires attention to enhance the efficient deployment of blockchain in IoT system.**

Keywords— Healthcare, Patient, IoT, Blockchain, Privacy, Security.

I. INTRODUCTION

The Internet of Things (IoT) is a communication paradigm that ubiquitously connects countless devices to the internet to send and receive data using embedded sensors and actuators [1],[2]. IoT has brought about a significant impact on the advancement of communication. Particularly, IoT devices are growing exponentially and are everywhere, positively affecting our daily lives, the society and assisting in critical decision makings in industries. It has also been estimated that IoT devices will grow to billions soon. Today, IoT has been widely implemented in different domains such as healthcare, homes, transportation, agriculture, cities, telecommunication, traffic, energy production and distribution, offices, and so on [2],[3]. IoT's importance lies in real-world data and computer processing power to minimize cost and maximize efficiency and accuracy.

With the multiple solutions in different domains offered by IoT, the healthcare sector plays a critical role in the medical landscape and brings about effective and personalized healthcare evolution [1],[3]. IoT-based healthcare is the most sensitive and data-oriented service of the IoT using various devices. This is due to its application in the remote, continuous monitoring and maintenance of a patient's health status in real-time without visiting the hospital [1],[2]. However, due to the increase in the number of IoT-healthcare devices and the continuous node connection in the IoT network, it becomes prone to various data privacy and security issues [1-2][4]. This is due to the lack of security protocols and standards in most IoT devices to cope with current and future security issues [1]. Consequently, patients' medical data are compromised, and sensitive information is leaked to unauthorized persons. This poses serious concerns and to protect personal and device-generated data, several solutions have been proposed and developed such as blockchain-IoT solutions [1-10].

As one of the fast-growing and critical innovations, blockchain is a completely auditable and digital decentralized ledger used for information exchange, recording and storing of transactions on the peer-to-peer network securely between parties without third-party involvement [4],[11-14]. Patil *et al*. [13] considered it as a time-stamped sequence of blocks that all participating nodes collectively manage. A block can have several transactions in the chain, digitally signed and chained to the preceding block by a cryptographic hash function [4],[12],[13]. A valid hash is used for validation by all nodes and a new block is added after confirmation, making the transaction immutable. Currently, due to the wide range of interests in blockchain applications, the technology is divided into categories: public or permissionless, and consortium blockchains [4],[15],[16]. Its critical components for validating and accepting new blocks into the distributed ledger is the consensus protocol which includes proof of work(PoW), Proof of Authority(PoA), proof of stake(PoS), proof of familiarity(PoF), practical byzantine fault tolerance, etc. [4],[17-19].

Blockchain technology has gained widespread application in different sectors such as governments, insurance, voting, gambling, personal and medical information security, logistics and supply chain, secure IoT, data storage [4],[11],[12] and so on. For healthcare, many techniques have been developed such as electronic health records (EHR) protection and remote patient monitoring (RPM) based on IoT [1-10],[17-32]. However, several challenges surrounding IoT integration into healthcare are dominated by security and privacy [33],[34]. As a viable solution, the use of blockchain technology in healthcare can enhance transparency and secure communications between patients and healthcare providers [11]. Therefore, this paper brings together some of the current research efforts made towards integrating blockchain into the healthcare IoT systems as well as provides open research direction for future work.

The remaining parts of this paper are structured as follows: Sect. II presents the related works, Sect. III discusses the

analysis of existing IoT-Blockchain technologies in healthcare, Sect. IV and V present the paper discussion and research opportunities respectively, while Sect. VI concludes the paper.

II. RELATED WORKS

Some of the existing surveys and reviews are presented in this section. Hussien *et al.* [11] surveyed blockchain technology with a focus on the trends and research opportunities in healthcare. Similarly, Hasselgren *et al.* [35] studied the relationship between blockchain in healthcare and health sciences, while in [33], the use of blockchain to enhance processes and services in healthcare, provides identified open research challenges and opportunities. Sookhak *et al.* [36] also extensively surveyed the issues surrounding the use of blockchain and smart contracts for healthcare access control and proposed a granular access control method for adopting blockchain and smart contracts in healthcare services. Similarly, Sanka and Cheun[37] reviewed the issues of blockchain scalability while Abu-elezz *et al.* [38] reviewed the benefits and threats of blockchain in healthcare.

Accordingly, Himeur *et al.* [39] surveyed blockchain-recommended systems together with their challenges and future opportunities. A taxonomy for security and privacy issues was proposed, existing architectures were discussed and some future research areas were outlined. Tariq *et al.* [34] also performed a survey on blockchain and smart healthcare security where different security issues were raised in terms of how they can be addressed in an efficient, distributive, and scalable way. Shuaib *et al.* [30] reviewed the blockchain's self-sovereign identity (SSI) solution application in healthcare. The study outlined its merits and requirements as well as introduced a use case model.

The above-discussed studies are some of the existing related reviews in IoT-blockchain healthcare [11],[33-39]. This paper thus focuses on current work to elicit existing methods and solutions, and highlight potential research opportunities for efficient healthcare applications.

III. IOT-BLOCKCHAIN INTEGRATION IN THE HEALTHCARE

A. Methodology

This section presents the analysis of some of the existing research works and proposals on IoT and blockchain technologies in healthcare. Several relevant papers were collected and analysed based using the content analysis technique. Only papers that discussed IoT-blockchain integration were considered in this paper to discuss the different schemes employed to protect patients' healthcare data in the healthcare environment and the challenges that exist.

B. The Analysis of Proposed Systems and Techniques

This subsection discusses the considered studies based on the focus areas such as authentication, remote patient monitoring, product tracking and efficient data management based on IoT-blockchain applications. As summarized in TABLE I & II, the overall objectives were to protect healthcare data and patients' privacy.

1) *Healthcare data and device authentication:* Alzubi [5], presented a blockchain-assisted solution for medical IoT systems based on the Lamport Merkle Digital Signature (LMDS) to protect sensitive health data. This was to simplify the placement of cloud IoT applications and to generate and verify signatures [5]. Mechanisms such as LMDS generator (LMDSG) and LMDS verification (LMDSV) authentication were used for medical data security. Performance evaluation revealed good security improvement and outperformed other methods with a 25% reduction in processing time and overhead and a 7% increase in the level of security. Shukla *et al.* [7] also addressed the challenges of IoT healthcare data authentication and device identification. A blockchain-integrated platform of fog computing was suggested to transmit healthcare data between IoT devices, patients, doctors and fog nodes securely and reliably. A new Advanced Signature-Based Encryption (AES) algorithm was designed based on the Diffie-Hellman (DF) key exchange technique for encrypting and decrypting data using the blockchain and different cryptographic operations. The AES identifies both similar and diverse healthcare IoT devices, verifies data and destinations as well as authenticates sent IoT data from different devices and fog nodes using joint probability with random generation. Evaluations revealed AES with fog computing outperformed cloud, etc. with 91% accuracy in malicious node identification and 95% reliability over the cloud. Similarly, Ray *et al.* [6] proposed a blockchain-based lightweight simplified payment verification (SPV) for IoT-assisted e-healthcare. It deals with the issues of unsupported toolsets considered resource-constrained, new architecture demand, etc. associated with Bitcoins transactions. A patient must have a bitcoin wallet with a legitimate quantity of bitcoins and be fitted with a generic pulse sensor linked to the bitcoin lightweight IoT node (BLWN) for more analysis of e-health sensor data against medical caregivers who are paid in bitcoin.

Fotopoulos *et al.* [8] also suggested an efficient, highly secure and robust blockchain-enabled authentication mechanism for the Internet of Medical Things (IoMT) devices of diverse stakeholders. It uses an SSI, zero-knowledge proof and blockchain to provide an authentication system with revocation in the healthcare environment: the hospital, device and manufacturer. A patient or clinic's new medical device is authorised to exchange data with the system. The apiece device is validated using a unique public/private key pair for IoT devices and the gateway or central system connection. This ensures the reliability, security, privacy, etc. of the transmitted sensitive data while the SSI protects against impersonation and spoofing attacks in addition to the privacy and integrity of data. In the same vein, Quasim *et al.* [26] proposed a model to secure EHR based on the combined effort of blockchain and the LPWAN gateways. The model ensures healthcare data security and reliability where sensors are attached to the patient and captures data from the body, which is then sent to the EHR cloud system via mobile phones or right to the cloud if the patient is in the hospital. The encrypted data is then sent from the cloud to the communication channel, which is ultimately transferred to the blockchain. This network's peer nodes use wireless communication channels to deliver data and process it in the blockchain network.

TABLE I. SUMMARY OF IOT-BLOCKCHAIN IN HEALTHCARE SYSTEM

Ref.	Proposed Solution and Objective	Implementation	Consideration
[5]	Blockchain-assisted solution for medical IoT systems based LMDS for higher security.	Yes: CloudSim 3.0, LMDS, LMDSV, LMDSG, secured consensus technique	Immutability, privacy, authentication, data integrity and speed
[19]	IoT blockchain-enabled platform for a healthcare application for EHRs storage and examination using a hybrid e-health decentralization system to protect healthcare data.	Yes: Raspberry Pi 3, Ethereum Smart contracts, PoW and PoA.	Immutability, privacy, confidentiality, accountability, speed, transparency, decentralization, and energy consumption.
[26]	Blockchain framework for the security of EHR where data in the cloud is encrypted and sent to the blockchain integrated with LPWAN.	Yes: Blockchain, Cloud, LPWAN	Immutability, privacy, confidentiality, integrity and transparency
[1]	Blockchain-based smart contracts for protecting security and privacy issues of patient and sensors information and enhance on-time treatment	Yes: Raspberry Pi 3, body sensors, GSM module, GPS sensor.	Immutability, privacy, confidentiality, availability, transparency and integrity
[8]	Blockchain-based IoMT authentication framework using SSI for scalable and practical authentication for medical devices of various stakeholders	Yes: Hyperledger Aries and Ursa, Hyperledger Identity stack, SSI, Medical device, device vendor, Gateway vendor	Immutability, privacy, authentication, integrity
[9]	Decoupled blockchain-based scheme secure in-house health record transmitted from IoT devices to the edge nodes.	Yes: Blockchain, edge devices, incremental tensor train, Mhealth dataset	Immutability, privacy, integrity, speed and energy consumption
[3]	Secure IoT architecture based on blockchain for distributed and secure health data access and monitoring application	Yes: Hyperleder fabric, Smart Contracts	Immutability, privacy, authentication, integrity, accountability, and transparency
[24]	Integration of medical records into a distributed ledger using blockchain and a cryptographic hash for extra privacy protection.	Yes: React JS, Web3 Library, MetaMask, Ganache	Immutability, privacy, integrity and accountability
[4]	Blockchain-enabled mHealth system via wearable sensors for transparency, security, and privacy of remote monitoring healthcare data.	Yes: private Ethereum, IPFS distributed storage protocol	Immutability, privacy, data audit, authentication, data integrity, speed and accountability
[28]	A lightweight consensus technique and a decentralized patient software agent for the RPM system.	Yes: Blockchain, API gateway, IoT device gateway, IoT devices	Immutability, privacy, speed and energy consumption
[22]	BSDMF is based on IoMT for security and privacy of transmitted patient healthcare data, scalable accessible healthcare data	Yes: Blockchain, IoMT devices, Cloud server	Immutability, privacy, authentication, transparency.
[25]	A secure IoT-based COVID-19 vaccine distribution system for effective tracking of vaccine units	Yes: Blockchain smart contracts	Immutability, privacy, authentication, integrity, availability, speed
[23]	BIoTHR: a blockchain and swarm exchange method to protect the privacy of healthcare data from IoT devices to a backend server.	Yes: GnuPG, IPFS, GOlang	Immutability, privacy security, transparency, interoperability, access control, availability, decentralization, pseudonymity, data aggregating, low cost

Griggs *et al.* [29] also suggested smart contracts for securing and enabling real-time analysis, transmission and logging of data transactions in IoT-based healthcare. It used an Ethereum-based private blockchain that executes the smart contracts invoked by smart devices communicating with patients' IoT healthcare devices. The smart contracts assess all the medical sensors generated data and record all transactions on the blockchain for verification of electronic health records. It was evaluated for confidentiality, immutability, transparency, privacy, etc.[29]. Moreover, Wang and Song [31] proposed a blockchain framework embedded with attribute-based encryption (ABE) and identity-based encryption (IBE) achieve fine-grained access control, authentication, confidentiality, and integrity of EHR stored in the decentralized system. It does not generate a private key and a patient is responsible for granting their data access policy to the hospital and sending it to the blockchain data pool for consensus nodes' processing[31]. The hospital in turn encrypts the patients' healthcare data using ABE and IBE and submits it to the blockchain using an identity-based signature (IBS). The consensus nodes constantly monitor both submissions and verify them for completeness and validity.

Like [29], Dwivedi *et al.* [32] suggested blockchain-enabled data management and analysis. It uses a novel modified blockchain framework based on lightweight cryptography. Patients' wearable devices are granted network access with registration and verified identity. Then healthcare data is transmitted to the smart devices for formatting and aggregation before it is submitted to the specified smart contract for in-depth

TABLE II. SUMMARY OF IOT-BLOCKCHAIN IN HEALTHCARE SYSTEM

Ref.	Proposed Solution and Objective	Implementation	Consideration
[20]	Blockchain-enabled system for the security of remote patient records monitoring.	Yes: Ethereum (Solidity, Remix) Sensors, Oracle Meta mask, injected web3 and truffle environment, VDM-SL toolbox, C++	Immutability, privacy, authentication, integrity and accountability
[27]	DL embedded blockchain system for the security of image transfer and model for diagnosis in the IoMT landscape.	Yes: Blockchain, DL, GO-FFO algorithm, ECC, NIS-BWT technique.	Immutability, privacy, authentication, integrity and transparency
[6]	SPV method for lightweight IoT node-based system model for unsupported resource-constrained tools in Bitcoins transactions.	Yes: Bitcoin BLWN, IoT gateway, Bitcoin wallet, Bitcoin testnet3 APP, IoT cloud.	Immutability, privacy, authentication, integrity, accountability and speed
[18]	Blockchain-based collaborative medical decision-making with PoF consensus algorithm to protect EMRs as well as address the lack of confidence among stakeholders.	Yes: Private blockchain, PoF, Multichain 2.0 (alpha)	Immutability, security, privacy, confidentiality, reliability, speed, cost, scalability, throughput, energy consumption
[29]	Blockchain-enabled smart contracts for medical sensors management and facilitates secure analysis	Yes: Private Ethereum with Solidity, Consensus algorithm, Smart contracts	Immutability, confidentiality, availability, speed, traceability, privacy and transparency
[7]	A decentralized FC-based blockchain analytical and mathematical models with AES algorithm for IoT healthcare data authentication and device identification.	Yes: iFogSim, private Blockchain, Fog computing, Healthcare IoT devices, ASE algorithm, PoW	Immutability, privacy, authentication, integrity, availability, speed, and energy consumption
[17]	A decentralized patient agent controlled blockchain-enabled healthcare system for efficient consensus mechanism to storage device generated in RPM systems.	Yes: Smart devicesJava, Fog network, blockchain	Immutability, privacy, authentication, integrity, accountability, speed, energy consumption.
[2]	BlockMedCare: A secure healthcare system that integrates IoT with blockchain to protect RPM.	Yes: private Ethereum, Clique PoA, IPFS, proxy re-encryption, Remix IDE	Immutability, authentication, integrity, privacy and access control.
[31]	Secure EHR system based on attribute-based cryptosystem and blockchain to protect healthcare data	No: Based on Ethereum, ABE	Immutability, privacy, authentication, access control, integrity
[32]	Blockchain-enabled secure healthcare data management and analysis.	No: Smart Contracts	Immutability, confidentiality, authorization, privacy or anonymity, data integrity and availability

analysis using the threshold value. The value determines the health status: "Normal" and "Abnormal". Once "Abnormal", an event is created immediately by the smart contract and sends a notice to alert the health providers without storing the health data but through transaction alerts. Thippeswamy *et al.*[24] also suggested a secured and distributed blockchain network used to store patients' medical records instead of a centralized cloud-based system. Each patient has complete control over his/her medical records with the right to grant access. Moreover, two-step authentication is employed to ensure both security and privacy in the system while an interplanetary file system (IPFS) securely stores medical records. It was demonstrated using ganache and Meta mask wallets to set up the system locally.

2) *Patient's remote monitoring:* Pham *et al.*[1] developed a secured remote healthcare system to protect sensitive personal information, and device-generated data, and to ensure efficiency during an emergency involving patients. The hospital creates a smart contract, writes it to the blockchain, and then exposes its address to the public database for access by registered patients and doctors. All smart contract records were encrypted registration while the authorization smart contract authorizes the doctor to monitor the patient's health. The IoT sensor data are carefully filtered and the decision is made on whether to write data into blockchain or not. However, all abnormal sensor data are written immediately into the blockchain and triggers an emergency contract with doctors and hospital for on-time treatment [1]. Also, Taralunga and Florea [4] suggested a blockchain-based mobile health (mHealth) system for RPM. Patient privacy and security are protected by employing a wearable sensor that communicates with smart devices using peer-to-peer hypermedia and uses IPFS protocols to monitor patients and store health data respectively. Smart contracts were employed for data queries and accessing healthcare data, recording diagnostic, treatment, and therapy as well as notifying medical professionals and patients. The blockchain enhances data availability and transmission times. Attia *et al.* [3] designed and implemented a secure IoT framework based on blockchain for remote healthcare applications monitoring. The system monitors patients discharged from hospitals and is followed by medical staff via a connected wearable device. The sensors continually measure, collect and store health data on the blockchain network using smart contracts. The collected data is identified using semantically meaningful names based on the naming data networking protocol.

Pradhan *et al.* [20] also developed a blockchain-based RPM system for patients with chronic diseases to protect health data during transmission and data loss. Healthcare data is collected by IoT devices and the invocation of the smart contract method triggers data to be gathered, read and written in all transactions on the blockchain virtual network. Based on the data collected, a consortium oversaw executes smart agreements and with reasonable agreements, the subcontract writes the event on the blockchain and sends notifications for patents and the required help. Similarly, Ray *et al.* [23] developed an EHR servicing strategy in the IoT-Blockchain network to protect healthcare data. The scheme employed a double-layer private blockchain-IoT system to transmit patient-to-doctor EHR as well as to transmit patient-to-doctor EHR for diagnostic purposes using a secure swarm node of P2P communication. Monitoring and controlling the EHR flow between doctors and patients were based on a hybridized mechanism for encrypting and decrypting. Likewise, the swarm exchange techniques eliminate third-party involvement while ensuring enhancement, high security, availability transparency, privacy, etc.

Bhawiyuga *et al.* [28] suggested a novel platform based on IoT blockchain to enhance healthcare data protection. Wearable sensing devices collect patients' physical conditions and transmitted them via the IoT gateway devices using wireless connections. The gateway device then sends the data to the IoT blockchain platform via a blockchain API gateway acting as a client chain node. The peers then execute and send back the endorsed data to the API gateway. All collected endorsements from peers are then sent to create the block of all transactions received which then sends the block to all peers for final verification and commitment. Similarly, Azbeg *et al.*[30] developed BlockMedCare to support RPM of chronic diseases and to protect healthcare data. IoT healthcare devices collect patients' health data while blockchain smart contracts ensure secure data sharing, and access control and data are stored using the IPFS to provide data integrity. Likewise, security is achieved via combining re-encrypting proxy and blockchain for hash data storage. An evaluation was performed using diabetes as a use case and the smart contract interactions using Remix IDE.

3) *Efficient data management:* Under this category, Aujla *et al.* [9] suggested a decoupled blockchain-based technique to transmit healthcare data efficiently to edge devices to protect its integrity and privacy preservation. It employed detached blockchain block headers and ledgers to transmit the device-generated healthcare data to the edge. It then uses the incremental tensor train method to transmit the data from the edge to the cloud server [9] which ensures data duplication reduction and minimizes data error. The results obtained revealed effectiveness based on the block preparation and header generation times, approximation error [9], etc. Uddin *et al.*[17] also proposed a decentralized patient software to control the RPM system. It replicates patient agent software on smart devices, fog and cloud servers to provide healthcare data reliability, security and privacy during transmission. A blockchain-based PoS consensus algorithm running at the fog and cloud layers was designed. A new block is created at the fog layer for insertion and is confirmed into the blockchain by consensus protocol executing in the fog devices. The modified PoS outperformance of standard PoS with a significant reduction in the energy consumed and time to generate a block.

Yang *et al.* [18] suggested a blockchain-enabled cooperative medical decision-making system to enhance healthcare data protection and ensure transparency among healthcare stakeholders such as doctors, insurance providers and current and cured patients. It uses a PoF consensus algorithm to integrate medical decisions among the parties in terms of their medical verdicts and policies. A prototype was implemented and tested and the results revealed a showed in the chance of personal information being identified, and EMRs and decisions were adequately preserved. Frikha *et al.* [19] also proposed an IoT-enabled blockchain system for managing healthcare and fitness data and boosting the confidentiality of health records during storage and transmission. Here, wearable devices continually monitor and collect patients' health and fitness data. The data is then filtered, analysed and stored in the EHRs where a decentralized ledger is created as a smart contract. This facilitates collaborations among different healthcare stakeholders to provide timely diagnoses and treatments easily and gainfully. The implementation offers the option to apply different consensus algorithms such as PoW and PoA. However, PoA outperformed PoW in terms of patients' privacy preservation, energy consumption, etc.

Furthermore, Abbas *et al.* [22] proposed a secure data management architecture via IoMT based on blockchain. It securely manages the data between a personal server and the IoT devices as well as between the cloud and personal servers using blockchain. Healthcare data is made accessible to public health authorities, etc. via immunization registries, syndrome surveillance, and electronic laboratory reports [22]. It assists public health organizations in effectively tracking, preventing, and treating illnesses. Experimental results obtained show it achieved about 97.2% accuracy ratio, 98.3% average trust value, 15.6% latency ratio and 11.2% response time [22]. Similarly, Alqaralleh *et al.* [27] suggested a framework for secure image communication and diagnostics in the IoMT setting using deep learning and blockchain. It collects healthcare data from patients via IoT devices, safe transaction processing, hash value encryption, and data categorization [27]. The elliptic curve cryptography (ECC) was used for encryption while optimal key generation was achieved by hybridizing grasshopper with the fruit fly optimization (GO-FFO) algorithm. Additionally, the neighbourhood indexing sequence (NIS) was then integrated with the burrow wheeler transform (NIS-BWT) to encrypt the hash values while the deep belief network (DBN) classified and identified the presence of illness [27]. Its performance evaluation revealed classification effectiveness of about 99% accuracy, 97% sensitivity and 98% specificity.

4) *Tracking:* Benita *et al.* [10] proposed an authentic drug usage and tracking system to shirk duplicate and counterfeit drugs. While Marbouh et al. [21] suggested a data tracking system based on blockchain to address misinformation about the covid-19 pandemic. Both proposals were implemented using Ethereum smart contracts. Moreover, Rathee et al. [25] suggested a covid-19 vaccine distribution via an IoT-based system and blockchain. The blockchain network ensures security among IoT devices while vaccines are sent from the

vaccine supply firm to the vaccine supply units before being distributed to a variety of sites.

IV. DISCUSSIONS

The swift technological advancements have created tremendous impacts in our society and the healthcare sector is not isolated. IoT has significantly contributed to the healthcare evolution from traditional to smart where patients are continuously monitored in real-time, diagnosed, and treated remotely [4][5][8][27]. Moreover, the blockchain due to its transparency, immutability, traceability, decentralization nature[4],[9][19], etc. has been deployed to protect the privacy and security of healthcare data either in transit or stored. Several systems such as RHM and mHealth [4][9][20][29], track COVID-19 vaccine distribution and misinformation[21][25], counterfeit avoidance[10], etc. and healthcare data management techniques [32] have been proposed and developed that integrates IoT with blockchain. Therefore, this paper has performed a review of some of the studies and provided an analysis of each solution, its approaches, tools used and possible future works.

Based on the review performed, we found that IoT blockchain in healthcare has brought about transparency and communications transformation to the healthcare system [2][4][23], cloud storage[22][26] Fog and Edge computing [9][17], etc. and the integration SPV [20] for bitcoin transactions. In general, the proposed solutions have shown the effectiveness of IoT and blockchain technologies in the healthcare ecosystem due to their unique features such as immutability, efficiency, cost, etc.

V. FUTURE RESEARCH OPPORTUNITIES

This section presents some of the identified concerns that have been flagged for further investigation. The LMDS-based authentication methods presented by Alzubi [5] need to be scaled up and concurrently help a large number of patients as well as capture numerous vital signs without escalating the processing time and overheads. Likewise, the IoT blockchain-based framework for EHRs [19] has to be improved to support different sensors variety implementable on a wearable device with parameters that allow health workers effectively evaluate patients' health status [19]. Moreover, an extra security layer that allows data to be encrypted before being sent to the blockchain for storage should be integrated. Several encryption algorithms that are lightweight and have strong security should be explored. Also, PoW can be explored as well for implementing various nodes and its impact on energy consumption.

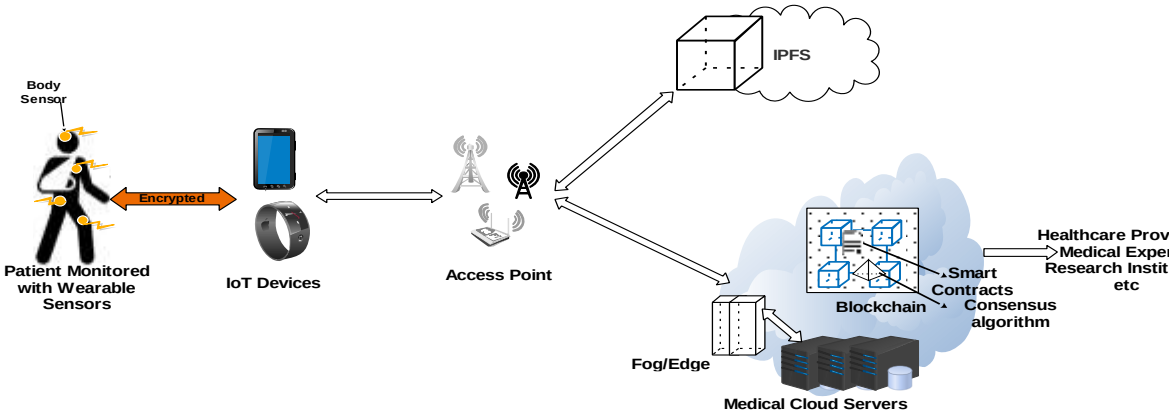


Fig. 2. IoT-Blockchain healthcare system generic framework

thereby enhancing its effectiveness, convenience, less cost as well as reduction of resources such as beds, doctors, nurses, etc. TABLE I and II present the review summary of solutions, implementation and security and privacy considerations. Accordingly, several solutions proposed or developed over the years, have a common solution architecture. (See Fig. 1). IoT devices are usually deployed as wearable devices in or around the patients' building or location, creating a network for monitoring, sensing, and transmitting data from sensors to other close-by devices or servers [4]. The gathered data are then used to get vital insights and make decisions. On the hand, the blockchain is used to enhance security, tracking, privacy and data management [9]. While several studies particularly focused on the security and privacy of health with Ethereum-based blockchain smart contracts [1][3][8][10][18-21][25][28][32], others added extra security to technology with cryptographic schemes such as hash[24], ECC [27], AES[7], LMDS [5], ABE [31], ASE[29], data storage with IPFS,

Quasim *et al.* [33] framework for integrating blockchain technology and LPWAN in healthcare needs further investigation such as boosting blockchain applications in cross border exchange of healthcare information via approaches like protocols, multi-jurisdictional health, standardization and data protection[26]. In addition, the blockchain's capability should be enhanced for a high volume of data storage, processing and analysis in a reasonable amount of time. Again, consensus algorithms, models for simulations and assessments of performance should be explored [26]. Similarly, Pham *et al.* [1] smart contract-based blockchain for remote healthcare monitoring systems should be extended with decentralized storage while Azberg *et al.* [2] suggested further research to explore the feasibility of Hyperledger blockchain in its solution in comparison to Ethereum-based solutions as well as the incorporation of artificial intelligence (AI) and more features added. According to [2], AI would help in the analysis of data,

prediction and prevention for efficient decision-making and treatments.

Shukla *et al.* [7] also suggested testing the ASE algorithm's ability to scalability limitation of blockchain and minimizing or eliminating the IoT-FC healthcare system's complexity with increased IoT and fog devices [7]. Again, the proposed technique's effectiveness should be explored in telesurgery, monitoring onshore and offshore oil and gas, etc. in the same vein, [8] suggested the enhancement of the proposed system with extra real-life requirements and properties such as business, efficiency, security and scalability. Also, Hyperledger Ursa can be enhanced with ABE for fine-grained medical data privacy as well as extended with the Hyperledger Indy network to like various hospitals [8]. On the same note, Aujla *et al.* [9] advocated for network performance improvement with networking technologies such as software-defined networks, 5G [9] and so on. This is also applicable to other proposed systems that utilized IoT-blockchain technologies. In [10], a suggestion has been made to extend the system using oracle time service to track drug expiry dates as well as integrate a QR code scanner with a mobile application. Other possible and efficient technologies can be explored as well.

Moreover, the framework in [3] needs to be enhanced further with extra features to achieve a full IoT-blockchain architecture to meet the needs of healthcare remote monitoring. Also, the mHealth framework in [4] can be enhanced with medical service payment using network tokens, secure access to healthcare data for clinical studies, etc. without integrating extra components [4]. Rathee *et al*.[25]'s vaccine distribution system should be extended by addressing issues such as verification delay, network scalability, special node selection, double spending threats, etc.[25] to effectively integrate blockchain with the smart-based application. Similarly, Ray *et al*. [23] suggested the proposed scheme be improved with integrated lightweight IoT devices, protocols, platforms [23], statistical monitoring, etc. In [20], medical data confidentiality can be improved by incorporating anonymizers in the system as well as finetuning the Hyperledger execution decisions in the system while Alqarallel *et al*. [27] suggested the need to enhance the DBN model performance with parameter tuning or any other feasible but efficient approach.

The proposed system in [18] can be improved to utilize previous decisions and learn from the environment for real-time decision-making as well as employing entity weighting factors in collaborative decision-making. Moreover, blockchain scalability issues should be addressed due to their adverse effect on real-life deployment. The proposed system should also be improved empirically with critical blockchain analysing factors such as block creation time, energy consumption, network latency, block finality time, and so on. Again, similar to [20], anonymizers could be integrated to improve privacy in the system with different transactions. Also, the Covid-19 pandemic tracking system [21] can be improved with additional smart contract functionalities as well as implementing decentralized applications for seamless participant interaction with Ethereum smart contracts [21]. Moreover, the solution is generic and can be adapted to other infectious diseases such as HIV, Malaria, TB, Ebola, etc. while in [32] an extension has been suggested to incorporate cryptographic components to improve the security of data. Though several proposed works have introduced several different cryptographic schemes and this remains an open research area.

VI. CONCLUSION

This paper has reviewed and analysed some of the selected papers on IoT-blockchain integration in the healthcare system. IoT-blockchain-based technical solutions were analysed and presented several research opportunities. The analysis as summarized in TABLE I and II show the existence of several solutions aimed at continuous monitoring, treatment and diagnosis of patients remotely as well as techniques to protect the security and privacy of sensitive EHRs. Consequently, IoT blockchain has transformed traditional healthcare into a smart one thereby, improving its efficiency, effectiveness, convenience, and cost as well as reduction of resources such as beds, doctors, nurses, etc. Moreover, while the amalgamation has been beneficial, there exist several issues that need to be tackled to ensure its performance and effectiveness such as blockchain scalability, improvement in privacy preservation such as the integration of anonymizers, as well as the addition of extra features to the proposed and future system such as specialized payment system, etc. Therefore, we recommend taking advantage of blockchain technology in sectors such as education, elections, etc. One of our future works will be the application of blockchain in registering and verifying vaccinated patients in healthcare.

ACKNOWLEDGMENT

This research was supported by the FNASRC, UDSC, the Department of Computer Science at the North-West University Mafikeng campus and the Council for Scientific and Industrial Research (CSIR).

REFERENCES

[1] H. L. Pham, T. H. Tran, and Y. Nakashima, "A Secure Remote Healthcare System for Hospital Using Blockchain Smart Contract," *2018 IEEE Globecom Work. GC Wkshps 2018 - Proc.*, 2019, DOI: 10.1109/GLOCOMW.2018.8644164.

[2] K. Azbeg, O. Ouchetto, S.J. Andaloussi, "BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security", *Egyptian Informatics Journal*, 2022, https://doi.org/10.1016/j.eij.2022.02.004

[3] O. Attia, I. Khoufi, A. Laouiti, and C. Adjih, "An IoT-Blockchain architecture based on hyper ledger framework for healthcare monitoring application," *2019 10th IFIP Int. Conf. New Technol. Mobil. Secur. NTMS 2019 - Proc. Work.*, 2019, DOI: 10.1109/NTMS.2019.8763849.

[4] D. D. Taralunga and B. C. Florea, "A blockchain-enabled framework for mhealth systems," *Sensors*, vol. 21, no. 8, pp. 1–24, 2021, DOI: 10.3390/s21082828.

[5] J. A. Alzubi, "Blockchain-based Lamport Merkle Digital Signature: Authentication tool in IoT healthcare," *Comput. Commun.*, vol. 170, no. April 2020, pp. 200–208, 2021, DOI: 10.1016/j.comcom.2021.02.002.

[6] P. P. Ray, N. Kumar and D. Dash, "BLWN: Blockchain-Based Lightweight Simplified Payment Verification in IoT-Assisted e-Healthcare," in *IEEE Systems Journal*, vol. 15, no. 1, pp. 134-145, March 2021, DOI: 10.1109/JSYST.2020.2968614.

[7] S. Shukla, S. Thakur, S. Hussain, J. G. Breslin, and S. M. Jameel, "Identification and Authentication in Healthcare Internet-of-Things Using Integrated Fog Computing Based Blockchain Model," *Internet of Things (Netherlands)*, vol. 15, p. 100422, 2021, DOI: 10.1016/j.iot.2021.100422.

[8] F. Fotopoulos, V. Malamas, T. K. Dasaklis, P. Kotzanikolaou, and C. Douligeris, "A Blockchain-enabled Architecture for IoMT Device

Authentication," *2nd IEEE Eurasia Conf. IoT, Commun. Eng. 2020, ECICE 2020*, pp. 89–92, 2020, doi: 10.1109/ECICE50847.2020.9301913.

[9] G. S. Aujla and A. Jindal, "A Decoupled Blockchain Approach for Edge-Envisioned IoT-Based Healthcare Monitoring," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 491–499, 2021, DOI: 10.1109/JSAC.2020.3020655.

[10] R. Benita, G. Kumar, B. Murugamantham, and A. Murugan, "Authentic Drug Usage and Tracking with Blockchain Using Mobile Apps," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 17, pp. 20–32, 2020, DOI: 10.3991/ijim.v14i17.16561.

[11] H. M. Hussien, S. M. Yasin, N. I. Udzir, M. I. H. Ninggal, and S. Salman, "Blockchain technology in the healthcare industry: Trends and opportunities," *J. Ind. Inf. Integr.*, vol. 22, no. November 2020, p. 100217, 2021, DOI: 10.1016/j.jii.2021.100217.

[12] M. J. M. Chowdhury, A. Colman, M. A. Kabir, J. Han, and P. Sarda, "Blockchain Versus Database: A Critical Analysis," *Proc. - 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng. Trust. 2018*, pp. 1348–1353, 2018, DOI: 10.1109/TrustCom/BigDataSE.2018.00186.

[13] P. Patil, M. Sangeetha, and V. Bhaskar, "Blockchain for IoT Access Control, Security and Privacy: A Review," *Wirel. Pers. Commun.*, vol. 117, no. 3, pp. 1815–1834, 2021, DOI: 10.1007/s11277-020-07947-2.

[14] J. Golosova and A. Romanovs, "The advantages and disadvantages of the blockchain technology," *2018 IEEE 6th Work. Adv. Information, Electron. Electr. Eng. AIEEE 2018 - Proc.*, pp. 32–37, 2018, DOI: 10.1109/AIEEE.2018.8592253.

[15] A. I. Sanka, M. Irfan, I. Huang, and R. C. C. Cheung, "A survey of breakthrough in blockchain technology: Adoptions, applications, challenges and future research," *Comput. Commun.*, vol. 169, no. December 2020, pp. 179–201, 2021, DOI: 10.1016/j.comcom.2020.12.028.

[16] X. Zheng, S. Sun, R. R. Mukkamala, R. Vatrapu, and J. Ordieres-Meré, "Accelerating health data sharing: A solution based on the internet of things and distributed ledger technologies," *J. Med. Internet Res.*, vol. 21, no. 6, 2019, DOI: 10.2196/13583.

[17] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "A Decentralized Patient Agent Controlled Blockchain for Remote Patient Monitoring," *Int. Conf. Wirel. Mob. Comput. Netw. Commun.*, vol. 2019-October, pp. 207–214, 2019, DOI: 10.1109/WiMOB.2019.8923209.

[18] J. Yang, M. M. H. Onik, N. Y. Lee, M. Ahmed, and C. S. Kim, "Proof-of-familiarity: A privacy-preserved blockchain scheme for collaborative medical decision-making," *Appl. Sci.*, vol. 9, no. 7, 2019, DOI: 10.3390/app9071370.

[19] T. Frikha, A. Chaari, F. Chaabane, O. Cheikhrouhou, and A. Zaguia, "Healthcare and Fitness Data Management Using the IoT-Based Blockchain Platform," *J. Healthc. Eng.*, vol. 2021, no. ii, 2021, DOI: 10.1155/2021/9978863.

[20] N. R. Pradhan, S. S. Rout, and A. P. Singh, "Blockchain Based Smart Healthcare System for Chronic -Illness Patient Monitoring," *3rd Int. Conf. Energy, Power Environ. Towar. Clean Energy Technol. ICEPE 2020*, 2021, DOI: 10.1109/ICEPE50861.2021.9404496.

[21] D. Marbouh *et al.*, "Blockchain for COVID-19: Review, Opportunities, and a Trusted Tracking System," *Arab. J. Sci. Eng.*, vol. 45, pp. 9895–9911, 2020, DOI: 10.1007/s13369-020-04950-4.

[22] A. Abbas, R. Alroobaea, M. Krichen, S. Rubaiee, S. Vimal, and F. M. Almansour, "Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things," *Pers. Ubiquitous Comput.*, 2021, DOI: 10.1007/s00779-021-01583-8.

[23] P. P. Ray, B. Chowhan, N. Kumar, and A. Almogren, "BIoTHR: Electronic Health Record Servicing Scheme in IoT-Blockchain Ecosystem," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10857–10872, 2021, DOI: 10.1109/JIOT.2021.3050703.

[24] M. N. Thippeswamy, B. M. Sai Kiran, P. R. Tanksali, M. Hegde, and P. R. Naik, "Blockchain based medical reports monitoring system," *Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020*, pp. 222–227, 2020, DOI: 10.1109/I-SMAC49090.2020.9243573.

[25] G. Rathee, S. Garg, G. Kaddoum, and D. N. K. Jayakody, "An IoT-Based Secure Vaccine Distribution System through a Blockchain Network," *IEEE Internet of Things Magazine*, vol. 4, no. 2, pp. 10–15, Jun. 2021, DOI: 10.1109/iotm.0001.2100028.

[26] M. T. Quasim, A. A. E. Radwan, G. M. M. Alshmrani, and M. Meraj, "A blockchain framework for secure electronic health records in healthcare industry," *Proc. Int. Conf. Smart Technol. Comput. Electr. Electron. ICSTCEE 2020*, pp. 605–609, 2020, doi: 10.1109/ICSTCEE49637.2020.9277193.

[27] B. A. Y. Alqaralleh, T. Vaiyapuri, V. S. Parvathy, D. Gupta, A. Khanna, and K. Shankar, "Blockchain-assisted secure image transmission and diagnosis model on Internet of Medical Things Environment," *Pers. Ubiquitous Comput.*, 2021, DOI: 10.1007/s00779-021-01543-2.

[28] A. Bhawiyuga, A. Wardhana, K. Amron, and A. P. Kirana, "Platform for integrating internet of things based smart healthcare system and blockchain network," *Proc. - 2019 6th NAFOSTED Conf. Inf. Comput. Sci. NICS 2019*, pp. 55–60, 2019, DOI: 10.1109/NICS48868.2019.9023797.

[29] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson, and T. Hayajneh, "Healthcare Blockchain System Using Smart Contracts for Secure Automated Remote Patient Monitoring," *J. Med. Syst.*, vol. 42, no. 7, pp. 1–7, 2018, DOI: 10.1007/s10916-018-0982-x.

[30] M.Shuaib, S.Alam, M. S. Alam, M. S. Nasir, "Self-sovereign identity for healthcare using blockchain", *Materials Today: Proceedings*,2021. Doi:10.1016/j.matpr.2021.03.083

[31] H. Wang and Y. Song, "Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain". *J Med Syst*. 2018 Jul 5;42(8):152. DOI: 10.1007/s10916-018-0994-6. PMID: 29974270.

[32] A.D. Dwivedi, G. Srivastava, S. Dhar, R. Singh, "A Decentralized Privacy-Preserving Healthcare Blockchain for IoT". Sensors 2019, 19, 326. Doi:10.3390/s19020326.

[33] T. McGhin, K. K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 135, no. September 2018, pp. 62–75, 2019, doi: 10.1016/j.jnca.2019.02.027.

[34] N. Tariq, A. Qamar, M. Asim, and F. A. Khan, "Blockchain and smart healthcare security: A survey," *Procedia Comput. Sci.*, vol. 175, no. 2019, pp. 615–620, 2020, DOI: 10.1016/j.procs.2020.07.089.

[35] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review," *Int. J. Med. Inform.*, vol. 134, no. December 2019, p. 104040, 2020, doi: 10.1016/j.ijmedinf.2019.104040.

[36] M. Sookhak, M. R. Jabbarpour, N. S. Safa, and F. R. Yu, "Blockchain and smart contract for access control in healthcare: A survey, issues and challenges, and open issues," *J. Netw. Comput. Appl.*, vol. 178, no. July 2020, p. 102950, 2021, DOI: 10.1016/j.jnca.2020.102950.

[37] A. I. Sanka and R. C. C. Cheung, "A systematic review of blockchain scalability: Issues, solutions, analysis and future research," *J. Netw. Comput. Appl.*, vol. 195, p. 103232, Dec. 2021, DOI: 10.1016/J.JNCA.2021.103232.

[38] I. Abu-elezz, A. Hassan, A. Nazeemudeen, M. Househ, and A. Abd-al Razzaq, "The benefits and threats of blockchain technology in healthcare: A scoping review," *Int. J. Med. Inform.*, vol. 142, no. August, p. 104246, 2020, doi: 10.1016/j.ijmedinf.2020.104246.

[39] Y. Himeur *et al.*, "Blockchain-based recommender systems: Applications, challenges and future opportunities," *Comput. Sci. Rev.*, vol. 43, p. 100439, 2022, DOI: 10.1016/j.cosrev.2021.100439.

Improved Application of Wavelet Transform in Protection of Multi-terminal HVDC Transmission Systems

F. Dehghan Marvasti
Department of Electrical Engineering
Yazd University
Yazd, Iran
farzad_dehghan@stu.yazd.ac.ir

A. Mirzaei
Department of Electrical Engineering
Yazd University
Yazd, Iran
mirzaei@yazd.ac.ir

M. Savaghebi
Department of Engineering Technology
Technical University of Denmark
DK-2750 Ballerup, Denmark
medi@dtu.dk

***Abstract*—Wavelet transform (WT) has proven to be a capable tool for protection purposes in high voltage direct current (HVDC) transmission lines due its desirable speed and accuracy. However, the process of waveform sampling can negatively affect the performance of WT and jeopardize the integrity of the WT-based protection principles in discrimination between internal and external faults. This paper investigates this phenomenon and proposes a method to mitigate that by analyzing the characteristics of fault-generated voltage travelling wave and modifying the sampled waveform. Analytical calculations and simulation studies are employed to fully investigate this matter. While analytical calculation establishes the basis to demonstrate and investigate this phenomenon, simulation study provides a more natural way of evaluating the detrimental impact of this phenomenon under real-time fault events. The proposed waveform modification method is implemented in a multi-terminal MMC-HVDC system to investigate its reliability and robustness. The simulation study exhibits improved performance for WT-based protection principles.**

Keywords— Fault-generated travelling wave, HVDC protection, Primary protection, WT-based protection

I. INTRODUCTION

Large-scale integration of renewable energy sources has become an increasing trend of harvesting electrical energy to meet the economic-benefit goals of green energy generation [1-2]. HVDC transmission lines incorporating modular multi-level converter (MMC) in multi-terminal systems are widely used for this purpose due to their distinct advantages such as independent active and reactive power control, fast controllability, and ability to reverse power with low difficulty [3-4]. Despite these advantages, MMC-HVDC systems have strict rules when it comes to the protection requirement, as the excessive fault currents can quickly damage the power electronic devices inside the converters [5]. Therefore, it is crucial to devise fast and reliable protection schemes for consistent operation of HVDC systems.

Traditional HVDC protection principles consist of travelling wave-based solutions, voltage derivative protection and current differential scheme [6]. Despite the difficulty in detecting high-resistance faults, travelling wave-based and voltage derivative methods are used as primary protection principles mainly for their fast response [7,8]. As backup protection, current differential protection method is utilized, which has enhanced selectivity under high-resistance faults [9,10]. However, sensitivity to the distributed capacitance of transmission line and slow operation are major deficiencies of this method [11].

To improve the performance of primary protection schemes, more advanced methods have been proposed in various studies, among which application of wavelet transform (WT) has proven to be useful in HVDC protection and fault location purposes and has been successfully verified in several studies. The research works presented in [12-14] use WT modulus maxima (WTMM) to discriminate between internal and external faults. In addition to using WTMM, the work presented in [14] also investigates the impact of various mother wavelets and wavelet decomposition scales on the performance of the adopted protection method. Using the generated WT coefficients, energy of the wavelet coefficients is used to identify fault scenarios in [15]. Extraction of high frequency components of the fault-generated transient voltages via WT is proposed in [16] for fault identification. A transient-voltage-based protection principle, focusing on the frequency-domain traveling-wave boundary characteristics, is proposed in [17], where the peak energy of the detail coefficients of WT is used for fault identification.

Similar to the aforementioned studies, WTMM and energy of WT coefficients of fault-generated voltage travelling waves are widely used for discrimination of internal and external DC faults. Despite the good performance verification, the impact of waveform sampling process on these methods, namely [12-17], is not properly discussed in these studies. Therefore, in this paper, it will be demonstrated that the process of waveform sampling, which is performed via employing a sampling window, can negatively affect the results of WT and consequently, degrade the performance of WT-based HVDC protection principles. Investigation on this matter is first conducted based on analytical calculation of line-mode fault-generated voltage travelling wave (LFVTW). After that, the detrimental impact of sampling process under various case studies are investigated via simulation study, and it is verified that based on the fault occurrence time and location, the sampling process can significantly affect the performance of WT-based protection methods. Finally, a waveform

modification method is proposed, which can mitigate the negative impact of sampling process and improve the application WT-based protection principles.

II. THEORY OF TRAVELLING WAVES

A. HVDC Test System

The schematic diagram of the multi-terminal MMC-HVDC test system and its parameters are presented in Fig. 1 and Table 1, respectively. The implemented overhead transmission lines are 400km, and are simulated using the frequency-dependent line model in PSCAD/EMTDC. The system is equipped with DC circuit breakers (DCCB) and current limiting inductors (CLI) for dealing with the DC faults. Measurement units and protective relays are installed at both ends of each line indicated by R_1-R_4 and R'_1-R'_4. Internal and external faults at various locations, as indicated by f_1-f_5, are considered for performance analysis of the proposed waveform modification method. The DC faults consist of single pole-to-ground (SPG) and pole-to-pole (PTP) faults. SPG faults comprise of positive pole-to-ground (PTG) and negative pole-to-ground (NTG) faults.

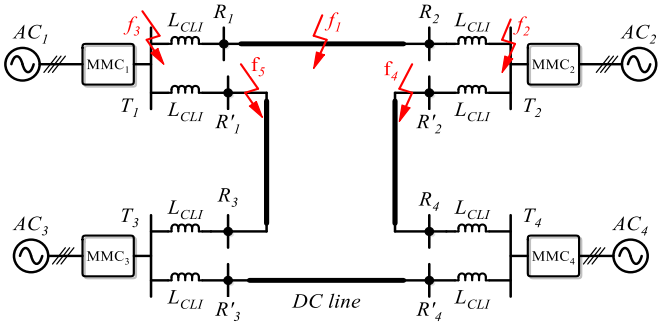


Fig. 1. Schematic diagram of the multi-terminal MMC-HVDC test system.

TABLE II. PARAMETERS OF MULTI-TERMINAL MMC-HVDC SYSTEM

System parameters	MMC1	MMC2	MMC3	MMC4
Rated power (MW)	2000	2000	2000	2000
Rated DC voltage, U_{dc} (kV)	±500	±500	±500	±500
Rated AC voltage (kV)	400	400	400	400
CLI (mH)	100	100	100	100
Converter arm inductor (mH)	40	40	40	40
Number of submodules per arm	225	225	225	225
Submodule capacitance (μF)	9000	9000	9000	9000

B. LFVTW at Fault Point

Majority of WT-based protection principles are primary protective solutions, which use DC pole voltage or LFVTW of the system for protective purposes [12-14]. Therefore, without the loss of generality, the analysis of this paper is performed on LFVTW measured at the sending terminal of the system (terminal R_1). To extract the line-mode component, first, a decoupling process should be performed on the pole voltages, as expressed in (1) [18].

$$\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u_p \\ u_n \end{bmatrix} \quad (1)$$

where the positive and negative pole voltages (u_p and u_n) are resolved into their equivalent zero-mode (u_0 and i_0) and line-mode (u_1 and i_1) components. After the decoupling process, the zero-mode ($\Delta u_{flt,0}$) and line-mode ($\Delta u_{flt,1}$) voltage variations at the fault point under PTG faults, can be expressed by (2) [18].

$$\Delta u_{flt,0} = \frac{-\sqrt{2}U_{dc}.Z_0}{s(Z_0+Z_1+4R_f)} \quad , \quad \Delta u_{flt,1} = \frac{-\sqrt{2}U_{dc}.Z_1}{s(Z_0+Z_1+4R_f)} \quad (2)$$

where U_{dc} is the rated DC pole voltage and R_f is the fault resistance. The Thévenin-equivalent zero-mode and line-mode impedances observed from the fault point are indicated by Z_0 and Z_1, respectively. Likewise, variation of the zero-mode and line-mode voltage components during NTG faults at the fault position are expressed by (3).

$$\Delta u_{flt,0} = \frac{\sqrt{2}U_{dc}.Z_0}{s(Z_0+Z_1+4R_f)} \quad , \quad \Delta u_{flt,1} = \frac{-\sqrt{2}U_{dc}.Z_1}{s(Z_0+Z_1+4R_f)} \quad (3)$$

For PTP faults, the zero-mode and line-mode voltage variations can be expressed as follows.

$$\Delta u_{flt,0} = 0 \quad , \quad \Delta u_{flt,1} = \frac{-\sqrt{2}U_{dc}.Z_1}{s(Z_1+R_f)} \quad (4)$$

Fig. 2 shows the equivalent line-mode circuit of the MMC-HVDC test system including terminal R_1 and R_2 under an internal PTG fault. The process of LFVTW extraction at terminal R_1 can be demonstrated using this figure, in which an internal DC fault, f_{in}, is used for the analysis.

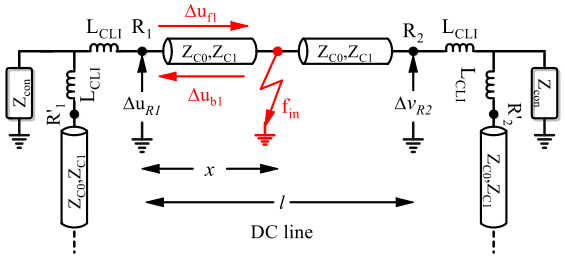


Fig. 2. Process of LFVTW calculation under an internal PTG fault

C. LFVTW at R_1 Under Internal Fault

LFVTW at terminal R_1 is the summation of the forward and backward voltage travelling waves originated from the fault point. Therefore, for an internal PTG fault (f_{in}) with resistance of R_f, LFVTW at R_1 can be expressed as follows.

$$\Delta v_{R1} = \Delta u_{b1} + \Delta u_{f1} = \Delta u_{flt,1}.e^{-\gamma_1(s)x} + \Gamma_1.\Delta u_{flt,1}.e^{-\gamma_1(s)x} \quad (5)$$

where Δu_{f1} and Δu_{b1} are the line-mode forward and backward voltage travelling waves, respectively. l and x are the line length and fault location, respectively. γ_1 is the line-mode attenuation coefficient of the line. Γ_1 is the line-mode reflection coefficient, which can be calculated based on (6).

$$\Gamma_1 = \frac{Z_{con} \,\|\, (sL_{CLI}+Z_{c1}) + sL_{CLI} - Z_{c1}}{Z_{con} \,\|\, (sL_{CLI}+Z_{c1}) + sL_{CLI} + Z_{c1}} \quad (6)$$

where Z_{c1} is the line-mode characteristic impedance. Z_{con} is the impedance of the converters, which is equivalent to a series RLC circuit and can be calculated based on (7) [19].

$$Z_{con} = R_c + sL_c + \frac{1}{sC_c} \tag{7}$$

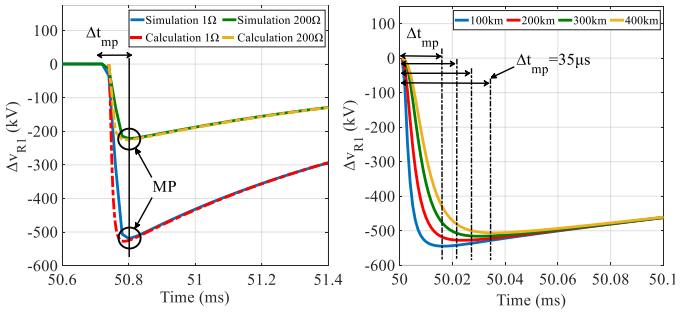
where $R_c = 2(R_{arm} + R_{on})/3$, $L_c = 2L_{arm}/3$ and $C_c = 6C_{sm}/N$. R_{arm}, L_{arm} and C_{sm} are the arm resistance, arm inductance and capacitance of the sub-module, respectively. R_{on} is the on-state resistance of the entire inserted sub-modules in each arm, and N is the number of inserted sub-modules. Vector Fitting employing a non-linear rational approximation can be adopted for modelling the line-mode propagation function ($e^{-\gamma_1 x}$) used in (5). Hence, the following form is employed [19].

$$e^{-\gamma_1(s)x} \cong F(s) = \sum_{k=1}^{n} \frac{C_k}{s - A_k} + D \tag{8}$$

where C_k and A_k are complex residues and poles, respectively, and D is a real constant. Finally, the frequency-domain expression of LFVTW as provided by (5) can be rewritten as follows.

$$\Delta v_{R1} = (1 + \Gamma_1).\Delta u_{flt,1}.F(s) \tag{9}$$

The time-domain expression of LFVTW can be extracted after applying the inverse Laplace transform to (9). To verify the accuracy of the analytical calculations, comparison with the simulation results of LFVTW under two internal PTG fault scenarios at 200km on the multi-terminal MMC-HVDC test system, is presented in Fig. 3(a).



(a) (b)

Fig. 3. Characteristics of LFVTW and a) impact of fault resistance on Δt_{mp}, b) impact of fault location on Δt_{mp}

Like other protection methods, WT-based protection principles use sampled quantities of a waveform to perform protective duties. These samples are obtained via waveform sampling process, which consists of taking measurements of a waveform at regular intervals and structuring the samples as an array of values. In regard to the WT-based protection principles discussed in this paper, the sampling process is performed on LFVTWs. It will be demonstrated that the sampling process can negatively affect the performance of WT-based protection principles by affecting the calculation of WTMM. Identifying the source of this phenomenon and mitigating its impact are the focus of this study. Therefore, LFVTW characteristics and application of Δt_{mp}, as shown in Fig. 3(a) and (b), are crucial in this regard, which will be further discussed in Section III-C.

Δt_{mp} is defined as the time it takes to reach the minimum point (MP) of LFVTW, which is calculated from the last steady-state sample of the waveform up until the MP moment as shown in Fig. 3(a). Considering that the system parameters are fixed during the operation of MMC-HVDC system, the main variables of (9) that may affect Δt_{mp} are fault location (x) and fault resistance (R_f). As shown in Fig. 3(a), fault resistance has minimum impact on the length of Δt_{mp}. On the other hand, according to Fig. 3(b), as the distance from R_1 to the fault location increases, so does the length of Δt_{mp}, which reaches its maximum value of $35\mu s$ for the fault scenario at 400km.

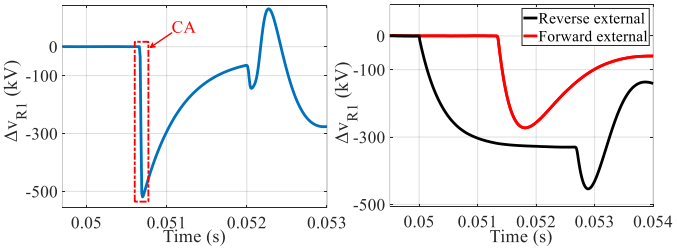
III. IMPACT OF SAMPLING PROCESS AND PROPOSED SOLUTION

A. *Typical WT-based Protection Principle*

A comparison of LFVTWs under 1-Ω internal and external PTG faults are presented in Fig. 4(a) and (b). By comparing the LFVTWs in Fig. 4(a) and (b), it can be seen that the severity of voltage drop rate under the internal fault is significantly higher than that under the external fault, which is directly related to the impact of CLI on the waveforms of the external faults. As the result of this, noticeably higher WTMMs are observed under internal faults than external faults. Therefore, extracting WTMMs at the 1st scale of decomposition and comparing them with a setting threshold is a widely used protective criterion for fault identification in many studies [14-17]. A general form of this criterion can be presented by (10).

$$|WTMM_{VR1}| > k \times th \tag{10}$$

where $WTMM_{VR1}$ is the WTMM obtained from applying WT to the LFVTW at terminal R_1. The setting threshold is denoted by *th*, which can be determined from the WTMM of a low-resistance external fault, and *k* is the reliability factor.



(a) (b)

Fig. 4. Comparison of LFVTWs under a) internal fault, b) external faults.

WTMMs of the reverse and forward 1-Ω external faults with the waveform presented in Fig. 4(b) can be calculated to determine the setting threshold (*th*) in (10). To ensure the practicality, a sampling frequency of 20kHz is used for the measurements. Using *db2* mother wavelet, the WTMMs at scale 1 for the reverse and forward external faults are 24.1 and 30.8, respectively. Therefore, considering *th*=31 and a reliability

factor of 1.5 to compensate for the impact of noise and measurement errors, the setting threshold in (10) is 46.5.

B. Identifying Impact of Sampling Process

WT-base protection principles extract and use WTMM for protective objectives, which is obtained after applying WT on a collection of sampled data. The samples are gathered via using a sampling window and applying a sampling frequency, typically, in range of 10-20kHz. Moreover, majority of these principles extract the WTMMs at the 1st scale of decomposition to focus on the high-frequency components of the travelling waves [12,13]. Therefore, based on the sampling frequency and the position of the samples inside the sampling window, WT can produce various WTMMs. This phenomenon can affect the performance of WT-based protection principles to the point that reliability of these methods can noticeably decline. To illustrate, there is an area marked as the critical area (CA) in Fig. 4(a) that directly affect the results of WTMM. This means that the shape of waveform inside the CA is crucial in the determination of WTMMs, which is again directly related to the position of the sampled data. To demonstrate this phenomenon, an internal PTG fault at 200km with resistance of 300 Ω is considered, and its LFVTW with a sampling frequency of 1MHz is calculated from (9). Then, totally 10 waveforms with a sampling frequency of 20kHz are extracted from the original 1-MHz signal in a way that there is a 1-μs time shift between the sampling intervals of two consecutive waveforms. These 10 waveforms are depicted in Fig. 5(a).

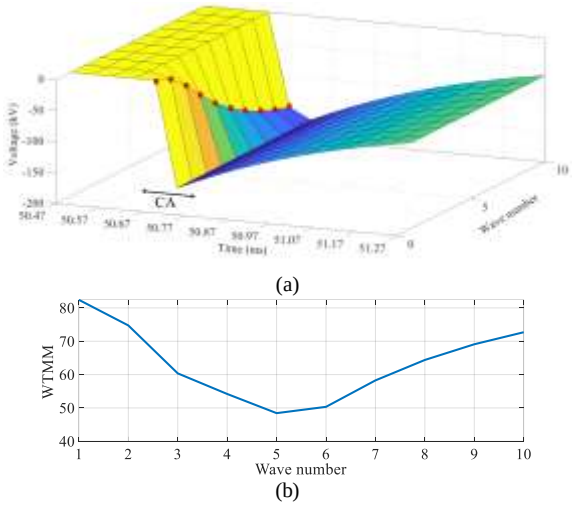


Fig. 5. a) LFVTWs with different sample structures (300-Ω internal PTG fault, 20kHz), b) resulted WTMMs

As marked with red dots in Fig. 5(a), it can clearly be seen that the samples inside the CA are located in different positions. After applying WT to the waveforms using *db2* mother wavelet, various WTMMs at scale 1 are obtained, which vary from 48.4 to 82.4, as shown in Fig. 5(b). The value of WTMM is highly dependent on the level of voltage drop that can be measured from two consecutive samples inside the CA. Therefore, depending on the position of samples inside the CA, different levels of voltage drop is observed, which can affect the results of WTMM. For example, as shown in Fig. 5 (a) and (b) the waveforms with higher levels of voltage drop between two consecutive samples (1st and 10th waveforms) are producing higher WTMMs compared to those with lower voltage drops (5th and 6th waveform). The high variation of WTMM observed in Fig. 5(b) can cause difficulty for WT-based protection principles as distinguishing internal high-resistance faults from external low-resistance faults becomes harder. As the fault resistance increases, the magnitude of WTMMs under internal faults declines. Consequently, WTMMs under high-resistance internal faults can come very close to those under low-resistance external faults. For example, the 5th and 6th waveforms in Fig. 5(b) are producing WTMMs of 48.4 and 50.3 respectively, which are very close to the threshold in (10) and may be identified as an external fault. Therefore, WT-based protection methods can face difficulties in identifying some of the internal high-resistance faults. It can be generally concluded that WT-based protection principles are reliable up to a certain fault resistance, which is around 300 Ω in this study. However, by implementing the proposed waveform modification method, which will be discussed in the following subsection, the negative impact of sampling process can be significantly mitigated.

The impact of sampling process on WT-based protection principles, even under a high sampling frequency, is significant. This can be verified by repeating the process of waveform creation explained for Fig. 5(a). Considering a sampling frequency of 100kHz instead of 20kHz, the LFVTW waveforms under a 1-Ω internal fault at 200km are depicted in Fig. 6(a), and the resulted WTMMs are shown in Fig. 6(b). It can be seen that the variation of WTMM is high and ranges from 124.4 to 231.

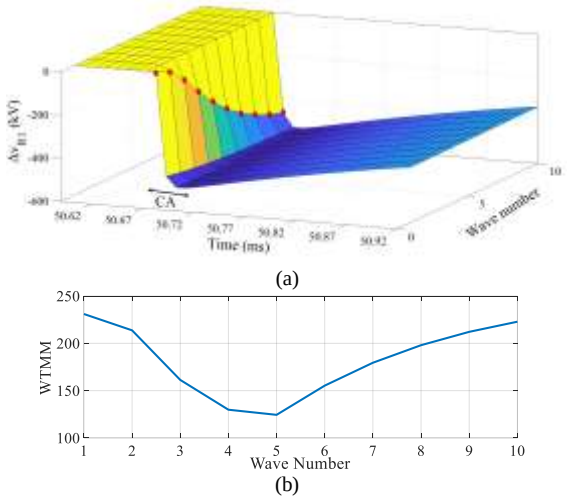


Fig. 6. a) LFVTWs with different sample structures (1-Ω internal PTG fault, 100kHz), b) resulted WTMMs

C. Proposed Solution

Now that the negative impact of sampling process is identified, a solution to that can be proposed. Referring back to Fig. 5(a) and (b), it can be seen that the maximum and minimum values of WTMMs are allocated to the 1st and 5th LFVTW waveforms respectively, which are depicted again in Fig. 7. As mentioned before, the main reason for the difference of WTMMs in these cases is the position of samples in the CA. According to Fig. 7, the blue waveform (BL) with the highest WTMM has only two samples in its CA, which translates into a higher voltage drop rate in comparison to that of the red waveform (RD), which has three samples in its CA. Therefore,

if the samples of RD are readjusted in a way that two consecutive samples reflect the highest amount of voltage drop (similar to BL), it can be expected to calculate higher WTMM and consequently, significantly eliminate the impact of sampling process. Therefore, the goal is to propose an algorithm to rearrange the samples in a way that the first sample inside the CA reflects the steady-state value of LFVWT, which is close to zero, and the second sample be as close as possible to the MP of LFVTW. The algorithm of the proposed waveform modification method is presented in Fig. 8, which can be explained based on the RD waveform in Fig. 7 as follows.

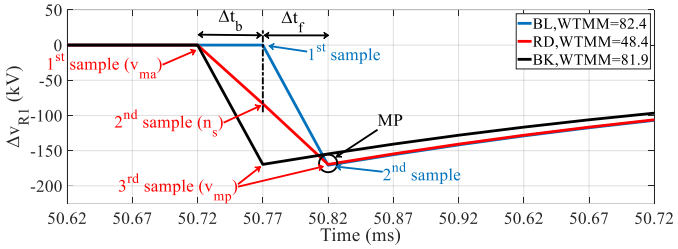


Fig. 7. Position of samples in CA of LFVTWs with highest and lowest WTMMs

First the starting sample of the algorithm should be indicated. Since majority of HVDC protection principles use a start-up unit (SU) to distinguish between normal and faulted condition [14], [16], the sample at which the SU is satisfied is used as the starting sample (n_s), which is the 2nd sample of RD in Fig. 7. After which the LFVWT at R_1 is calculated; two sampling windows, Δt_b and Δt_f, each with the length of 50μs are considered to identify the maximum and minimum voltage quantities of the waveform. The maximum voltage quantity in Δt_b and the minimum voltage quantity in Δt_f are then extracted, which are denoted by v_{ma} and v_{mp} in Fig. 7, respectively. Finally, the entire samples before n_s are replaced with v_{ma}, the voltage sample at n_s is omitted and substituted with v_{mp}, and the rest of the waveform is formed via shifting the original waveform samples.

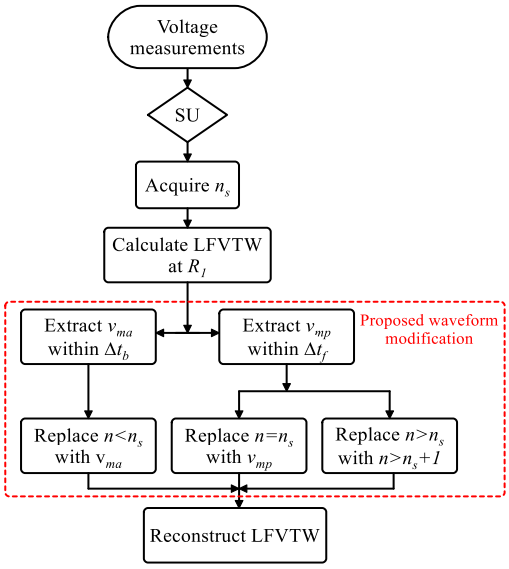


Fig. 8. Proposed waveform modification algorithm

After applying the proposed waveform modification on the RD, the modified version is obtained, which is represented by the black waveform (BK) in Fig. 7. In comparison to the WTMM of the RD (48.4), the WTMM of the modified waveform is enhanced to 81.9 and is very close to that of the BL (82.4). Therefore, by performing the proposed modification, BK has become very similar to BL in terms of the shape and resulted WTMM, and the negative impact of sampling frequency (as demonstrated in Section III-B) is mitigated. It is important to highlight that the main reason for choosing the two 50-μs sampling windows (Δt_b and Δt_f) is due to the conclusion derived at Section II-C and Fig. 3(b), where it was revealed that the maximum value of Δt_{mp} is 35μs in the case of fault scenario at 400km. Therefore, by assigning the 50-μs data windows, it is guaranteed that v_{ma} and v_{mp} can be extracted within Δt_b and Δt_f, and they closely represent the steady-state and MP values of the LFVTW respectively.

IV. SIMULATION RESULTS

As discussed in Section II, the position of samples inside the CA is crucial in determining the results of WTMM. In real world HVDC systems, where the sampling process is performed in real time, structure of the data inside the CA is directly related to the fault signature arrival time (FSAT) at terminal R_1, when the first samples are measured. Obviously, fault occurrence time is one of the factors that can affect FSAT. According to (9), fault location is another major factor, whose impact on FSAT can be determined via F(s). Therefore, it is essential to evaluate the performance of the proposed waveform modification method under fault scenarios with different fault occurrence times and fault locations to verify the application of the proposed method in real world fault scenarios. In this regard, numerous fault scenarios consisted of SPG and PTP faults with various fault occurrence times and different fault locations are simulated. The test system and the simulated fault scenarios are according to Fig. 1. Considering the 50-μs data window, the sampling frequency is chosen to be 20kHz. Some examples are provided here for further analysis.

Fig. 9 presents fault scenarios at 200km with resistance of 500 Ω when the fault occurrence time is different. The blue waveform presents the case where the fault happens at 50ms, and its generated WTMM is 57.5. According to (10), this can be reliably identified as an internal fault. However, if the same fault happens only 20 μs later at 50.02ms, the red waveform is generated, and the resulted WTMM declines to 32.9, which is consequently identified as an external fault. The main reason is due to the position of the samples inside the CA, which is directly related to the moment of FSAT at R_1. After applying the proposed waveform modification method on the red waveform, the modified black waveform is obtained, which has the WTMM of 55.5, and can be correctly identified as an internal fault. To investigate the impact of fault location on the sampling process and evaluate the proposed method in improving the performance, various PTG faults with resistance of 500 Ω at three different locations of 200, 250 and 300km are simulated. The measured LFVTWs at R_1 are presented in Fig. 10. It can be seen that the sampling process does not have a negative impact on the LFVTWs of the faults at 200 and 300km, and the resulted WTMMs are 57.5 and 56.1, respectively. This means that these

fault scenarios are reliably identified as internal faults. However, unlike the other cases, the fault scenario at 250km has three samples inside its CA and consequently, the resulted WTMM declines to 35.4. This means that this fault is mistakenly identified as an external fault. After performing the proposed modification, the waveform samples are readjusted (black waveform), and the WTMM is 55.1. Therefore, after the proposed modification, this case can be correctly identified as internal.

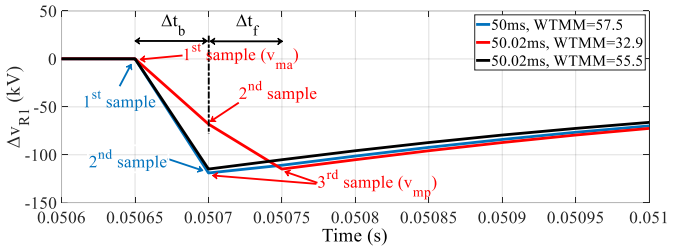


Fig. 9. Impact of fault occurrence time on sampling process, and performance of proposed method in mitigating the issue

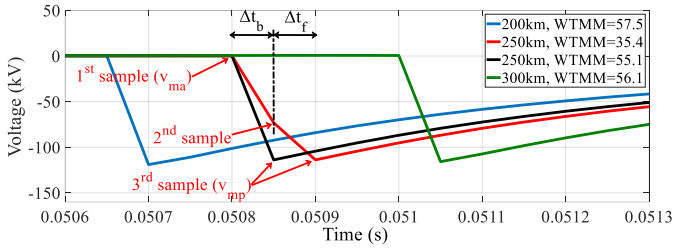


Fig. 10. Impact of fault location on sampling process, and performance of proposed method in mitigating the issue

Overall, it can be concluded that based on the fault occurrence time and fault location, the sampling process can deteriorate the performance of WT-based protection principles. However, the proposed waveform modification method can reliably mitigate this problem. It is important to highlight that the detrimental impact can be observed under some fault scenarios and does not exist under other circumstances. In other words, the impact of sampling process is quite arbitrary in nature. Therefore, highlighting this phenomenon and mitigating the issue is a strong point of this study.

V. CONCLUSION

WT is a popular tool for protective purposes in HVDC transmission lines due to its high speed and accuracy in fault identification. However, the process of waveform sampling can negatively affect its performance. Therefore, the main focus of this study is to identify the source of this detrimental impact and mitigate it via proposing a waveform modification method. First, analytical approach is used to demonstrate this phenomenon, and later simulation cases are studied to evaluate the performance of the proposed method. It is concluded that without the proposed modification, WT-based protection principles may face challenges in identifying internal faults with resistances higher than 300 Ω, while after implementing the proposed modification, reliable protection operation under high-resistance faults of up to 500 Ω is achieved. The validity of analysis is demonstrated on a multi-terminal MMC-HVDC transmission system.

ACKNOWLEDGMENT

The authors would like to thank Yazd Regional Electric Company for the support of this research.

REFERENCES

[1] A. Gandomkar, A. Parastar, and J. K. Seok, “High-power multilevel step-up DC/DC converter for offshore wind energy systems,” IEEE Trans. Ind. Electron., vol. 63, pp. 7574-7585, 2016

[2] P. Mitra, L. Zhang, and L. Harnefors, “Offshore wind integration to a weak grid by VSC-HVDC links using power-synchronization control: A case study,” IEEE Trans. Power Del., vol. 29, pp. 453-461, 2014

[3] F. D. Marvasti and A. Mirzaei. “A novel method of combined DC and harmonic overcurrent protection for rectifier converters of Monopolar HVDC systems,” IEEE Trans on Power Del, vol. 33, no. 2, pp. 892-900, 2018

[4] O. G. Bellmunt, A. J. Ferre, A. Sumper, and J. B. Jane, “Control of a wind farm based on synchronous generators with a central HVDC-VSC converter,” IEEE Trans. Power Syst., vol. 26, no. 3, pp. 1632–1640, 2011

[5] S. Du, A. Dekka, B. Wu, and N. Zargari, Modular Multilevel Converters: Analysis, Control, and Applications, Wiley-IEEE Press Hoboken, New Jersey, United States, 2018

[6] A. Li, Z. Cai, Q. Sun, X. Li, D. Ren, and Z. Yang, “Study on the dynamic performance characteristics of HVDC control and protections for the HVDC line faults,” in Proc. Power Energy Soc. Gen. Meeting, Calgary, AB, 2009, pp. 1-5

[7] J. Sneath and A. D. Rajapakse, “Fault detection and interruption in an earthed HVDC grid using ROCOV and hybrid DC breakers,” IEEE Trans. Power Del., vol.31, no.3, pp:973–981, 2016

[8] W. Leterme, J. Beerten, and D. Van Hertem, “Nonunit protection of HVDC grids with inductive DC cable termination,” IEEE Trans. Power Del., vol.31, no.2, pp:820–828, 2016

[9] F. Dehghan Marvasti, A. Mirzaei, M. Savaghebi, and M. R. Jannesar “A pilot protection scheme for HVDC transmission lines based on simultaneous existence of forward and backward voltage travelling waves,” 2022 IEEE 13th International Symposium on Power Electronics for Distributed Generation Systems, Kiel, Germany, 2022

[10] M. Gamal Muhammad, D. Mourad Hafez, and A. Mahmoud Abd-Elaziz, “Novel HVDC transmission line protection scheme based on current differential,” 2018 Twentieth International Middle East Power Systems Conference (MEPCON), 2018, pp. 361-366

[11] X. Chu, “Transient numerical calculation and differential protection algorithm for HVDC transmission lines based on a frequency-dependent parameter model,” Int J Electr Power Energy Syst, vol. 108, pp.107-116, 2019

[12] V. Psaras, D. Tzelepis, D. Vozikis, G. P. Adam, and G. Burt, “Non-unit protection for HVDC grids: an analytical approach for wavelet transform-based schemes,” IEEE Trans. Power Del., vol. 36, no. 5, pp. 2634-2645, 2021

[13] Y. Zhang and W. Cong, “An improved single-ended frequency-domain-based fault detection scheme for MMC-HVDC transmission lines,” Int J Electr Power Energy Syst, Vol.125, 2021

[14] X. Pei, H. Pang, Y. Li, L. Chen, X. Ding, and G. Tang, “A novel ultra-high-speed traveling-wave protection principle for VSC-based DC grids,” IEEE Access, vol. 7, pp. 119765-119773, 2019

[15] B. Mitra, B. Chowdhury, and A. Willis, “Protection coordination for assembly HVDC breakers for HVDC multiterminal grids using wavelet transform,” IEEE Systems Journal, vol. 14, no. 1, pp. 1069-1079, 2020

[16] W. Xiang, S. Yang, L. Xu, J. Zhang, W. Lin, and J. Wen, “A transient voltage-based DC fault line protection scheme for MMC-based DC grid embedding DC breakers,” IEEE Tran. Power Del., vol. 34, no. 1, pp. 334-345, 2019

[17] B. Li, Y. Li, J. He, and W. Wen, “A novel single-ended transient-voltage-based protection strategy for flexible DC grid,” IEEE Tran. Power Del., vol. 34, no. 5, pp. 1925-1937, 2019

[18] J. Suonan, S. Gao, G. Song, Z. Jiao, and X. Kang, “A novel fault-location method for HVDC transmission lines,” IEEE Trans. Power Del., 106463, pp: 1203-1209, 2010

[19] C. Zhang, G. Song, A. P. S. Meliopoulos, and X. Dong, “Setting-less nonunit protection method for DC line faults in VSC-MTdc systems,” *IEEE Trans. Industrial Electronics*, vol. 69, no. 1, pp. 495-505, 2022

Remote Work Alerting with Artificial Intelligence Technics

Nevra Akbilek
Department of Industrial Enginnering
Engineering Faculty Sakarya University
Istanbul,Türkiye
nakbilek@sakarya.edu.tr
ORCID:0000-0002-9525-1755

Yunus Akkaya
Database Assistant
Research Center
CMC_Turkey
Istanbul, Türkiye
Yunus.Akkaya@cmcturkey.com
ORCID: 0009-0000-0181-1401

Abstract—**Ensuring and maintaining the efficiency and security of company processes has become an essential issue in the remote working system, which started with Covid-19 and became widespread, especially in the service sector. According to the banking regulation supervision agency customer representative in a call center must be an actual employee, and the working environment must be appropriate and performance adequate. So, a system is proposed in which webcam images of customer representatives working from home can be analyzed, and anomalies can be detected. For this, an artificial neural network model has been established that performs real-time image processing and facial recognition, reality, and vitality analyses. In the study results, an artificial intelligence-based model was established to predict whether the person displayed on the camera during remote work is a natural person, whether there is a fake, and whether there is more than one person using viola jones and CNN algorithms. The answers to all these questions are sought, and the necessary actions are taken by reporting the statistics obtained.**

Keywords— remote working, Image processing, Face recognition, Liveness determination, person identification, biometrics.

I. INTRODUCTION

In the remote working system, which has become a necessity with the pandemic period, some employees return to the office, while most of them continue to work from home. There are two reasons for this: the firm wants to cut costs, and the employees do not want to come to the office. In addition, in today's world, a system of working from home is carried out through platforms connected to employees who are independent of the company. All of these have forced companies to keep up with this transformation. This research aims to analyze how efficient the remote working system is with the example of a call center company. A framework has been proposed in which the webcam images of the customer representatives working from home can be analyzed, and any anomalies or frauds can be detected.

Person recognition systems have undergone many developments from past to present and have been divided into branches. Fingerprint recognition, iris recognition, face recognition, or in some cases, security card-like applications are used at the beginning of the techniques used to identify people. In this study, face recognition is chosen because it is not disruptive, reasonably priced, continuous, a legal necessity, and more natural. Also, according to banking regulation, supervision agency customer representatives must be actual employees, and the working environment must be appropriate and performance adequate. Up-to-date artificial intelligence technologies will be used to carry out these operations.

There are some studies about remote working. In one of them, data such as keystrokes and clicks on the remote worker's computer were taken, and the encrypted data was transferred to the local database and then to a cloud database, and then entered into the SVM (support vector machine) model for training. The anomaly reports received were given to the authority as a printout. The project asked; is classified as successful, unsuccessful, or timeout. As input; mouse movements (clicking, wheel rotation), keyboard movements (ctrl c/v movements, function movements, Windows key presses, etc.), time spent at the computer, age, work experience, company, position, project information, program usage times were used [1].

Another study tried to determine whether the person was seen as a living person with CNN (convolutional neural networks) in the photographs. Three different databases are used in the study. These; are CASIA-FASD, Replay-Attack, and OULU-NPU. In the survey, cross-checking is done by training the model on one data set and testing it on another. Apart from that, the inputs are tested in RGB, blend, and disparity formats. The fusion layer is used at the beginning of the CNN model and in the middle of the model. As a result, it is seen that it is more effective and economical to use it in the first layer when the location of the fusion layer is changed [2].

Irbaz et al. (2021) introduced a face recognition model using Face NET and KNN for Labeled Faces in the Wild dataset, and they obtained a high accuracy rate for remote working [3]. Chen et al. introduced a bidirectional prediction to predict the same objectives with forwarding and backward networks for detecting surveillance video [4].

Beyond these, what is desired to be done in this study is to perform an efficiency measurement and to detect various fraud situations that may occur by controlling the working from home environment in real-time, using image processing techniques. Image recognition can be conducted through 2D and 3D methods. Chihaoui et al. (2016) introduced a detailed survey about 2d face recognition. 2D face recognition methods are classified mainly as global, local, and hybrid [5]. Global approaches included: linear technics and non-linear technics, and local technics: interest-point-based face recognition and Local Appearance- Based Face Recognition methods. Local Appearance- Based methods are used different properties are: Fourier transforms, Weber Law descriptor(WLD) [6], local binary pattern method(LBP) [7], Haar wavelets [8], Local phase quantization(LPQ) [9], Gabor coefficients [10]. We applied the local appearance-based face recognition method that uses Haar wavelets in this study.

The techniques used for three-dimensional face recognition can be divided into four headings, including branches under themselves. These; are global feature-based (based on global features), local future based (based on local features), deep feature-based, multimodal fusion (combining more than one model), non-specific conditions (based on some unchanged parameters, stance shape, mask, etc.) [11 ,12]

In this study, the viola-jones algorithm, which is used to detect faces according to general features, was used. Viola-jones algorithm is an algorithm developed by Paul Viola and Michael Jones in 2001 to detect a particular object from the images. Its operation is generally as follows; The algorithm examines the image piece by piece at a specified interval, using available images roughly consisting of columns and squares, and looks for eye, nose, and mouth expressions for this study. When it finds the necessary general shape, it defines it as a face. The general facial expression formed by these squares and rectangles is called haar [13].

II. MATERIALS AND METHODS

A. Viola-Jones Algorithm

Viola-jones algorithm is an algorithm developed by Paul Viola and Michael Jones in 2001 to detect a particular object from the images. Its operation is as follows; The algorithm examines the image piece by piece at a specified interval, using available images roughly consisting of columns and squares, and looks for eye, nose, and mouth expressions. It defines the image as a face when it finds the necessary general shapes. The general facial expression formed by these squares and rectangles is called haar [13].

Viola-jones algorithm has 4 main steps;

- Selecting the haar feature
- Preparation of the integral image
- Adaboost training
- Cascading classifiers

For this work, we use the frontal face cascade taken from GitHub, and for recognizing face we tried to simulate a call center agent default position. Here is some example for haars;

Fig. 1. Nose detection

Fig. 2. Eye detection

The face matching function is given in Equation 1. As seen in the image, how much the shape resembles the image to be detected according to its similarity to a specific shape is calculated according to the following formulation, and a threshold value is applied. As a result, step-by-step scanning is performed, and the result is obtained according to the total image as shown in Figure 3.

$$F_{Haar} = \frac{E(R_{black}) - E(R_{white})}{w \cdot h \cdot \sqrt{\left| E(R_{\mu})^2 - E(R_{\mu}^2) \right|}} \quad (1)$$

There are black pixels for negative values and white pixels for positive values. To detect the face in a video or an image the haars hovered over on frame, the areas under white pixels (positive values) are multiplied by the white pixel coefficient, and black areas under the black pixels (negative values) are multiplied by the black pixel coefficient. Moreover, that variables are added.

As a result, step-by-step scanning is performed, and the result is obtained according to the total image.

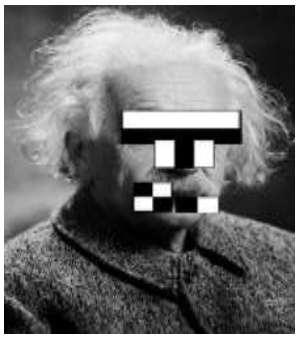


Fig. 3. Face detection with Haar shapes

Adaboost learning: The algorithm is used in viola jones's face or object detection for optimization. Considered is first boosting algorithm and has received the Gödel award. The steps;

- Determine y=1 for each positive example and y=0 for each negative example; the coefficients are set to -2/2m and 2/2n
- For each iteration, coefficients are optimized with the function;
- t, the weights probability distribution
- Every h haar is applied and the loss calculated.
- Set the haar weights for minimum loss
- Set the optimum classifier.

B. CNN Algorithm

A convolutional neural network is a name given to a neural network model that uses convolution in at least one of its layers. It is a sub-title of deep learning and is often used in visual or natural language processing projects. For this study, it was used in face detection and reality estimation.

A neural network is a method for simulating actual biological neural systems. The purpose is to find out the best function which is given the most similar outputs with the train variables. The results are constantly improving by forwarding and backward propagations. This system tries to work like a natural synaptic stimulation. When the outputs came to the threshold model gives the results. There are three main layers:

Convolutional layer: This layer is where the first inputs are given, and a mathematical convolution is performed on the information by transforming it into a matrix. Main features are tried to be extracted by performing a filter operation on the matrix. These properties are output from this layer to be processed in subsequent layers [14].

Pooling layer: Its primary purpose is to reduce the time cost of processing input from the previous layer. It works to reduce the complexity and connection paths that occur when extracting the feature map. It works separately on each feature map. Different pooling methods can be used according to the methods used.

Fully-connected (FC) layer: The inputs from the convolutional layers and the pooling layer are finalized in the final stage [15].

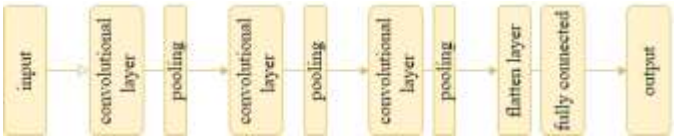


Fig. 4. CNN architecture

III. IMPLEMENTATION

After face detection, a data set consisting of previously prepared photographs is used to understand whose face it is. The face in the photograph registered to the system is scanned and the person's name is printed on the screen as a result of matching the appropriate person. Otherwise, the face is detected but tagged as unknown [16].

It aims to use a convolutional neural network model by using artificial intelligence to analyze the vitality of the face. In recent years, there have been many ways to fake facial recognition systems. When there is a different person in front of the camera, the system must detect it. Apart from this, it is aimed at preventing situations such as putting photos in front of the camera. To prevent attempts to deceive this type of face recognition system, it aims to understand whether the face on the screen belongs to the person it should be, even if it is a natural person, by training a model and making a real-time prediction. Several techniques can be used for this. One of these is to perform an anomaly detection based on facial movements, another to analyze by comparing three-dimensional or two-dimensional images of the face, or a reality analysis can be made by training the model with images modified according to pixel quality and real images [2].

TABLE III. PROCESSING STEPS

Solution Steps	Step content	Solution tool, program, etc
1	Data collecting	Kaggle, GitHub
2	Examining the data	Numpy, pandas
3	Development of neural network model	Cnn, Keras
4	Training the model	Tensorflow, Keras
5	Setting up the model for face detection	OpenCV
6	Training the model for face detection	Viola-jones haar cascade

To set up the desired system, real-time face recognition and status detection must be performed. For this, a face recognition

process was performed using the OpenCV module on python, and a CNN model was established for reality analysis. To solve the problem, the following steps are conducted below.

Four possible situations are considered—the first case: is the detection of known and unknown (not in database) images with the viola jones algorithm. The result of the algorithm application is printed on the screen in Figures 4, 5, and 6. When an unidentified face is seen using the application, it is detected but is printed to the screen as unknown, as shown in Figure 4. In other words, the image of the face is not registered in the database.

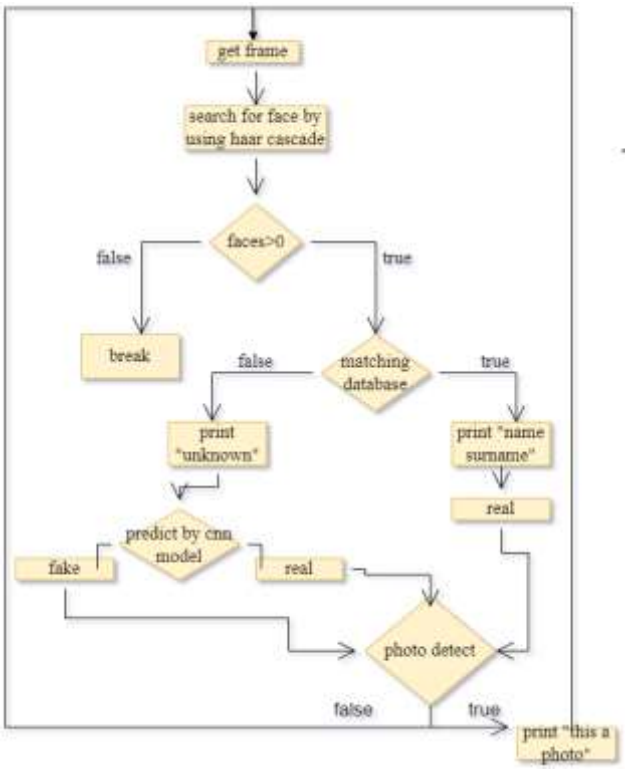


Fig. 5. Flow diagram of the proposed system

Fig. 6. Known real face

The second case is the estimation of whether the image is real/fake with the CNN algorithm. With this prediction, it is understood whether the image is an animation or a photoshop image. If it is an actual photo, the real message is printed on the screen, as shown in Figure 4. Otherwise, the fake message is printed on the screen, as given in Figures 5 and 6.

Fig. 7. Unkonwn real face

Fig. 8. Unkonwn fake face

In the third case, it is understood whether the image is a photograph or not. The inactivity of the picture is checked, and if it is still inactive for more than 10 seconds, the message "This is a photograph" is printed on the screen, as given in Figure 7. However, it returns to normal when motion is detected again.

Fig. 9. Unkonwn / photo

In the last case, it is checked whether there is more than one person. If there is more than one person, the relevant warning is printed on the screen, as shown in Figure 10.

Fig. 10. More than one face

IV. RESULTS

As a result of this work, an audit system has been tried to be created for companies with the remote working system. In the proposed face detection process, a comparison with the existing database has been conducted to detect whether the agent is the person or not with the image processing algorithms. For these processes, the Python Face_Recognition library and viola jones algorithm was used, and the desired outputs were obtained. To predict whether the received image is a situation to mislead the system, it was tried to indicate whether it is a natural human face with CNN. Also, if the extreme stasis, this is a photo warning related to the image printed on the screen. Also, there is more than one person, and the number of faces warning related to the image is printed on the screen.

REFERENCES

[1] M. Akpinar, M. F. Adak & G. Guvenc (2021). SVM-based anomaly detection in remote working: Intelligent software SmartRadar. *Applied Soft Computing*, *109*, 107457.

[2] Y. Abbas. U. Rehman, L. M. Po, M. Liu, Z. Zou, W. Ou & Y. Zhao (2019). Face liveness detection using convolutional-features fusion of real and deep network generated face images. *Journal of Visual Communication and Image Representation*, *59*, 574-582.

[3] M. S. Irbaz, A. Nasim, M. D. Abdullah & R. E. Ferdous (2022). Real-time face recognition system for remote employee tracking. In *Proceedings of the International Conference on Big Data, IoT, and Machine Learning* (pp. 153-163). Springer, Singapore.

[4] D. Chen, P. Wang, L. Yue, Y. Zhang, & T. Jia (2020). Anomaly detection in surveillance video based on bidirectional prediction. *Image Vis. Comput.*, *98*, 103915.

[5] M. Chihaoui, A. Elkefi, W. Bellil & C. B. Amar (2016). A Survey of 2D Face Recognition Techniques. *Comput.*, *5*, 21.

[6] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen; X. Chen, W. Gao WLD: A robust local image descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1705–1720.

[7] T. Ahonen, A. Hadid, M. Pietikainen Face recognition with local binary patterns. In *Computer Vision—ECCV 2004*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 469–481.

[8] P. Viola & M. J. Jones (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.

[9] V. Ojansivu, J. Heikkila Blur insensitive texture classification using local phase quantization. In Proceedings of the International Conference on Image and Signal Processing, Cherbourg-Octeville, France, 1–3 July 2008; pp. 236–243.

[10] R. Brunelli, T. Poggio Face recognition: Features versus templates. *IEEE Trans. PAMI* **1993**, *15*, 1042–1052.

[11] M. Li, B. Huang & G. Tian (2022). A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, *110*, 104669.

[12] Y. Zhanga, X. Wang, M. S. Shakeel, H. Wan & W. Kang (2022). Learning upper patch attention using dual-branch training strategy for masked face recognition. *Pattern Recognition*, *126*, 108522.

[13] M. V. Alyushin & L. V. Kolobashkina (2018). Optimization of the data representation integrated form in the Viola-Jones algorithm for a person's face search. *Procedia computer science*, *123*, 18-23.

[14] M. Billah, X. Wang, J. Yu & Y. Jiang (2022). Real-time goat face recognition using convolutional neural network. *Computers and Electronics in Agriculture*, *194*, 106730.

[15] K. Jia (2022). Sentiment classification of microblog: A framework based on BERT and CNN with attention mechanism. *Computers and Electrical Engineering*, *101*, 108032

[16] K. G. Shanthi & P. Sivalakshmi (2021). Smart drone with real time face recognition. *Materials Today: Proceedings*.

Developing a Spatio-Temporal Simulation Framework for Urban Energy Consumption Using Agent-based Modeling

I. Chun Chen
Department of Land Resources
Chinese Culture University
Taipei, Taiwan
cogiitri@gmail.com (ORCID: 0000-0002-4645-0125)

***Abstract*—The study analyzes urban energy consumption spatio-temporal patterns using agent-based modeling (ABM). The model uses the Local Indicators of Spatial Association (LISA) and the Geographically Weighted Regression (GWR) to overcome the problem of spatial heterogeneous interference in ABM simulation. In addition, the study also establishes features of preferences of various groups in energy consumption via the Decision Tree (DT). The energy consumption prediction process is executed in six modules via the transformation mechanism of living behavior and preferences of 907 agents in the studied area (Taipei City). The results of the better simulation are verified by Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The result in 2030 exhibits rising trend for energy consumption in five areas, with most rising agents concentrated in downtown areas, and declining trend for energy consumption in seven areas. The cumulative distribution function (CDF) of energy consumption prediction shows the number of agents exceeding 600 KWH/household, pointing to increase in regional energy consumption. Simulation results point to trend of further concentration of energy consumption in the future. Given possible shortfall in energy supply in the future, the study suggests to build small-scale power plants or introduce energy-conservation strategies in regions with concentrated energy consumptions, to assure stability of energy demand and supply there.**

Keywords—urban, agent-based model, Geographically Weighted Regression, Decision Tree, energy consumption behavior

I. INTRODUCTION

Rapid increase in global energy consumption in recent years has posed major environmental hazards. Energy consumption forecast models are common divided into two types, namely top-down and bottom-up types, in terms of econometric analysis, which needs a lot of statistical data for linear regression or nonlinear regression. However, the model has to overcome collinearity problems resulting from the causal relationships among variables, which may impede forecast accuracy [1]. In addition, urban patterns, in terms of socioeconomic variables, have complex nonlinear interactive effect, causing spatial heterogeneity. Meanwhile, it is not easy to collect reliable data on residents' energy usage habits, posing a major challenge for resource-consumption prediction involving interaction among economic element, human behavior, and resource flow in an urban system [2,3]. Therefore, to be more realistic, energy consumption forecast model needs high-resolution spatiotemporal information, including that on urban socioeconomic and human behaviors.

II. URBAN ENERGY CONSUMPTION SIMULATION

A. Main Variables of Urban Energy Consumption

Urban areas have accounted for 75% of global carbon dioxide emissions in the past decade, in line with their contribution to GDP worldwide. Because residents of different cities have different energy consumption patterns [4], changes in their behaviors could attain 10-85% savings in energy consumption. As a result, research on the energy consumption forecast model has expanded its coverage from buildings (micro level) to an entire urban area (macro level) [2].

Via literature review (SCI-E and SSCI), Zhang et al. (2018) found that existing studies on energy demand have focused on three major aspects, namely energy consumption, architectural design, and human behavior. Human behavior takes into account mainly residents' factors, including physiological (age and gender), psychological (personal preferences), social (jobs, household composition, education, and marital status), or environmental conditions (building type, outdoor temperature, air quality, and indoor temperature) [5, 6].

B. The Influence of Spatial Heterogeneity

A major problem for the prediction model on urban growth is how to overcome the influence of variables' nonlinear relationship, caused by spatial heterogeneity [7, 8]. Different variables have the problem of spatial autocorrelation in geographic information system, meaning while everything is mutually related, things of proximity have higher relevance. The aggregation and dispersion of socioeconomic characteristics of each spatial feature can cause heterogeneity interference. Spatial-temporal influence of heterogeneity is identified by the Local Indicators of Spatial Association (LISA)[9]. Meanwhile, the Geographically Weighted Regression (GWR) results on a fine geographic scale are presented and LISA is used to analyze optimal spatial bandwidth [10].

C. Agent-based Model

Agent-based model (ABM) can simulate a complex system with heterogeneous and diverse elements, which have interconnection enabling them to adapt to changes over time. The most important part of ABM result is how to accurately reflect the actual interconnection between agents and environment, thereby reducing the heterogeneity effect [11]. Urban resource simulation has a nonlinear relationship in space,

detrimental to the accuracy of ABM simulation. To better simulation of agent's action in space, it needs for ABM to establish relation attributes between agents and simulated environment, such as causal relationship, temporal relationship, and topological relationship.

Human behavior derives from a complex decision coupling process between people and the living environment (social, economic, and space). Socio-ecological system calls for a simulation process to explore the anthropocentric nature of the ecosystem service concept with inductive and deductive methods. Given the evolving effect of human behavior, social survey data and questionnaire results can manifest agents' actions and reactions. The theoretical agent-based model develops a simulation method to explore agents' behaviors, including emergency response and self-organization, as well as behaviors that are adaptive and probabilistic, optimizing, integrative, and inclusive in nature [12].

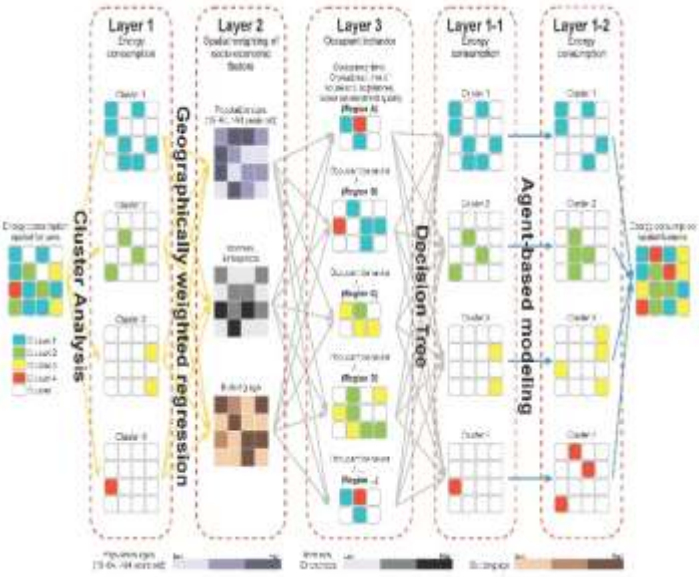
Existing studies have evaluated agents' behavioral preferences from the approaches of heuristics, utility maximization, adaptive learning, and typology method[12] and have developed several research models, such as Cellular Automata Modeling, Artificial Neural Network Modeling, Fractal Modeling, Linear/Logistic Regression, and Decision Trees (DT) [13]. DT, which is a kind of supervised machine learning, employs a heuristic method to identify agents' decision-making. DT can expose agents' decision preferences, thereby facilitating verification and validation of simulation results or exploration of decision pathways. Commonly applied in ABM, self-adaption method such as ANN lacks transparency in simulation. Therefore, DT appears to be the best model for studying the influence of agents' behaviors [14].

Fig. 1. Framework of agent-based model

III. METHOD AND MATERIALS

The study developed a small-scale geographic spatial model for urban resource metabolism, aiming to explore the influence of human behaviors (living habits) and socioeconomic conditions on energy consumption, elaborating on urban energy-concentration areas, as the basis for the formulation of regional energy management strategies. As shown in Fig. 1, a framework of agent-based model, the study first divides urban residents into several groups according to socioeconomic spatial data and then employs GWR to analyze the influence of those groups' socioeconomic conditions on energy consumption (model's first layer to second layer), followed by the employment of decision tree to explore residential comfort of various groups and types of their residences, as well as analysis of their habits and preferences in the use of home appliances (model's second layer to third layer). The study conducted questionnaire surveys on residents' energy consumption behaviors and analyzed energy consumption preferences with the decision tree method, setting data for agents' choice probability (T) of their habits and preferences in the use of home appliances. Fig. 1 sheds light on the change of groups' energy-consumption pattern by using ABM, with energy-consumption distribution shifting from the pattern on the left side to that on the right side.

Given the inability for comparing ABM simulation results with actual data and the difficulty in obtaining the data, as shown in many studies worldwide. The study concentrated on the data of Taipei city, which are available in complete information in many open databases, to overcome the above problem. The study employed the socioeconomic data of the target groups in 2010-2019 period, including energy consumption (average energy consumption per household), constructions (average age), economic activities (individual income tax, numbers of enterprises in various categories), and demographics (the three age brackets of 14 and less, 15-64, and 65 and over). It looked into changes in the attributes of the studied groups during the period and evaluated their changes. With 2019 as the baseline year, the study forecasts changes in energy consumption up to 2030 via simulation.



A. *Agent Cluster and Spatial Relations*

The study employed Taiwan's open data with better resolutions for identifying urban consumption patterns. With energy consumption data in use at the smallest spatial statistical level and socioeconomic data at the village level, the study employed data with scale consistency to look into features (agents) on the smallest scale for energy consumption simulation.

K-means algorithm, a common unsupervised algorithm for clustering analysis, was employed as reference in calculating the groups of residents. To divide all the observed values into k number of groups, the within-cluster sum of squares was calculated using each point, with the average within-cluster value of that group at the smallest, in order to identify the cluster that satisfied its smallest value. The optimal grouping results based on the smallest WSS are used to plot a broken line with the elbow method. In addition, the study clarifies the spatial aggregation and heterogeneity of the socioeconomic variables of each agent with LISA and then identifies with GWR the geographic weights between dependent variables (X_{hi}) and independent variables ($AMHEC_i$).

The study employed GWR to establish initial simulation Equation (1) for the energy consumption of different groups, containing i number of independent variable X_{hi} of spatial

statistical agent, spatial weight β_{hi}, intercept β_{oi}, and error value ε_i, as in:

$$AHEC_i = \beta_{0\,i} + \sum_{h=1}^{5} \beta_{h,i}(\mu_i, \nu_i) x_{h,i} + \varepsilon_i \quad (1)$$

B. Agent Energy Behavior Preference

Cluster analysis shows that survey respondents in the study were distributed across different groups. The sampling area is based on the statistical data of the Ministry of the Interior (MOI), which reveal population hotspots accounting for 21% of Taipei City's population. With the population hotspots as the parameter, the sampling rate stands at 0.15%. The questionnaire survey collected 773 valid copies of responses, offering answers to the following items: (i) basic information; (ii) residential type; (iii) electricity usage habits; and (iv) feelings about the indoor living environment. Then, the study used decision-tree analysis to identify agents' behavior preferences, concerning living comfort level and household appliances usage.

In the ABM scenario, the study used the final leaf blades (*l*) of different groups' decision tree paths (*d*) as agents' energy-consumption preferences in calculating the average household energy consumption and ratios of different groups (*g*) as the correction coefficient (*S*) and probability for choice of the leaf-blade path (*T*), with correction coefficient shown in Equation (2). Of various independent variables X_h in Equation (1), h includes: 1. building age, 2. the median number of individual income taxes, 3. the numbers of enterprises, 4. the number of 15-64 age group, and 5. the number of people aged 65 and older. Two scenarios were employed in the analysis of behavior preferences with decision-tree method (*d*): residential comfort decision tree (*a*) and home-appliances usage decision tree (*b*), set with residential comfort behavior $(S_{a,g,l}|T_{a,g,l})$ and home-appliances usage behavior $(S_{b,g,l}|T_{b,g,l})$, respectively. Equation (3) shows in:

$$S_{d,g,l} = \frac{AHEC_{d,g,l}}{AHEC_g} \quad (2)$$

$$AHEC_i = \beta_{0\,i} + \beta_{3,i}(\mu_i, \nu_i) x_{3,i} + \sum_{h=1}^{2} \beta_{h,i}(\mu_i, \nu_i)\left(S_{a,g,l}|T_{a,g,l}\right) x_{h,i} + \sum_{h=4}^{5} \beta_{h,i}(\mu_i, \nu_i)\left(S_{b,g,l}|T_{b,g,l}\right) x_{h,i} + \varepsilon_i \quad (3)$$

C. Agent Interaction Analysis

Given their different socioeconomic features, factors driving the transformation of different groups vary. The study judged mutual conversion among different groups with their overlapping features in the boxplot of four major socioeconomic data of past years. The intersection areas of the boxplot can be classified into four kinds, namely gradually increase, flat, gradually decrease, and no intersection. Groups with socioeconomic data exhibiting the first three features indicate the possibility of mutual transformation. Various agents' transformation probability (*P*) is calculated according to the number of overlapping units in the scope of overlapping numerical values of the first quartile (Q1/4) and third quartile (Q3/4) of the boxplots of a pair of groups' socioeconomic data in past years. A single influence factor with intersection has 10 transformation probabilities (*P*) (2010-2019). Without transformation, a specific group will retain its behavioral preference for energy consumption.

D. ABM Overview - Design - Details

The study employed GAMA (GIS Agent-based Modeling Architecture) platform to predict energy consumption for 907 specific agents in Taipei City. GAMA is a simulation platform consisting of such tools as integrated agent-based programming, geographic data management, flexible visualization tools, and modeling language GAML (Gama Modeling Language). In addition, the study referred to the guide of ABM ODD (Overview - Design - Details) [15] in developing a simulation framework.

- Purpose: The model aims to analyze the effect of agent energy consumption behaviors and predict temporal-spatial agents' consumption patterns in urban areas. According to simulation results, it is advised to establish small renewable energy facilities in energy consumption hot zones and formulate energy-saving strategies based on agents' behaviors.
- State variable and scale: The model sets 907 villages as the simulation boundary and evaluates urban energy-consumption patterns based on agents' transformation deriving from socioeconomic variables and agents' energy behaviors. The model comprises four hierarchical levels: village spatial polygon, agents, energy consumption, and buildings, while agents are characterized by state variables, namely income, enterprises, and age demographics.
- Process overview and scheduling: The model, with a one-year span, consists of six modules, including agent instantiation, agent transform probability (*P*), agent transform stop, agent behavior choice probability (*T*), and energy consumption prediction. Fig. 2 shows the research flow of the agent-based model. The study treats administrative areas (villages) as the individual boundary, with fixed scope and shape. Initial settings for administrative areas include such attributes as groups' energy consumption, respective independent variables, and GWR coefficient. Features of the boxplot of the socioeconomic state variables of administrative areas are used as reference for determining the possibility of transformation of agents' groups in administrative areas, with transformation principles being explained in III C. Therefore, with their state variables no longer overlapping, groups will stop transformation and conduct *T* and *S* selection with their original behavioral preferences. If groups' variables change, feature i may shift from the original group (*g*) to another group (*g'*). Energy consumption intensifies forecast hinges on the energy consumption behavior of the group (*g* or *g'*) to which feature *i* belongs. Behavioral features of different groups with different kinds of buildings (residence or business) are employed in probing the effect of comfort behaviors and home-appliances usage behaviors on energy consumption. Energy consumption forecast and

the influence of behavioral features are explained in III B.

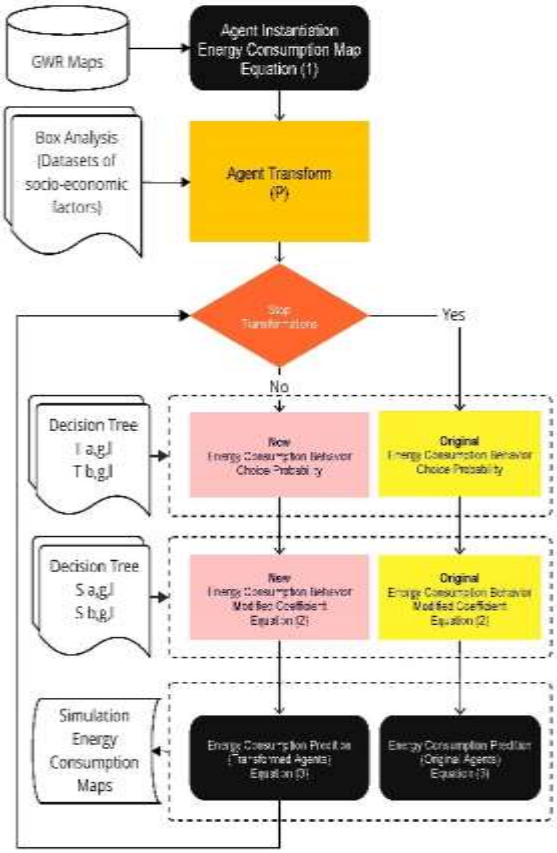


Fig. 2. Flow of agent-based model

- Design concept: Emergency is treated as the choice probability (*T*) of agents' energy consumption, the higher the *T* value, the higher the effect on agents' energy consumption (*S*). Interaction is regarded as the feasibility of agent alternation, for which the *P* value is set. According to the explanation in Section III C., each agent's transition can be influenced by at least 10 sets of *P* values of each socioeconomic variable. As a result, each set of *P* values can be interpreted as agent transform probability, which is related agent's characteristics, with the model setting the *P* value in a random access manner.
- Details: Fig. 2 shows the flow of setting agent instantiation and action parameter input and output.

IV. RESULTS AND DISCUSSION

Among the state variable features of the 907 agents of the studied area, except two agents with excessively low households (group 0) and 60 agents whose energy consumption data are outliers of overall data (group 5), the remaining 845 agents can be divided into four groups with K-means elbow method. Group 1 contains the largest number of agents (381 agents), all belonging to areas with the lowest average energy consumption per household, individual income tax (FLD04), and enterprise numbers (C_CNT). Group 4 has features similar to group 1, with more young/middle-aged and aged population, though, which affects the group 4 power consumption volume. Group 2 boasts the highest median individual income tax (FLD04) and group 3 has the largest number of enterprises (C_CNT), resulting in the highest energy consumption per household and much higher median individual income tax than group 1 and group 4. The study incorporated group 3 into group 2, renamed group 23, due to their largely similar features. Fig. 3 shows the spatial distribution of various groups and Table 1 exhibits the transformation probability(*P*) of various groups' agents.

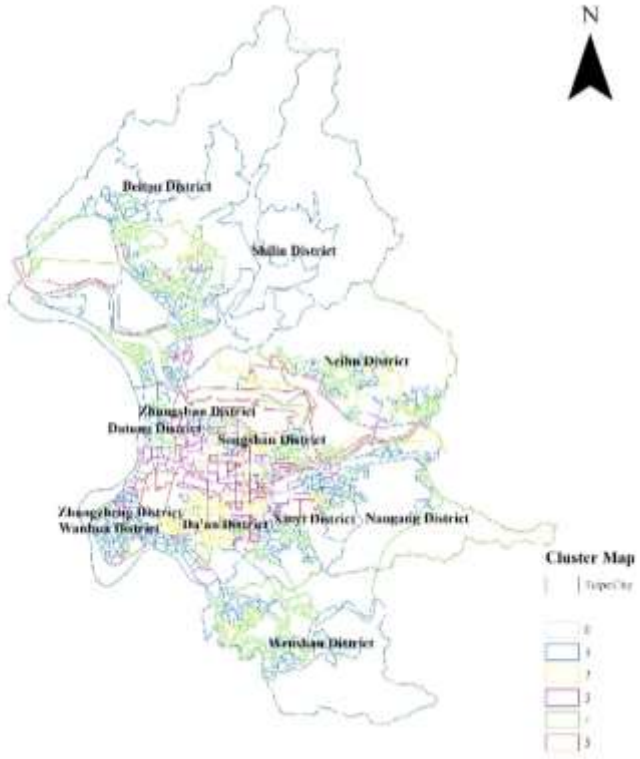


Fig. 3. Agent cluster map

TABLE I. AGENT TRANSFORM PROBABILITY

Group (*g*)	Group (*g'*)	State Variables	Transform Probability (*P*)	
			MIX	***MAX***
1	23	median number of individual income taxes	0.06	0.13
1	23	number of 15-64 age group	0.28	0.43
1	23	number of people aged 65 and older	0.28	0.43
23	1	median number of individual income taxes	0.02	0.05
23	1	number of 15-64 age group	0.11	0.16
23	1	number of people aged 65 and older	0.12	0.16
1	4	median number of individual income taxes	0.28	0.33
1	4	numbers of enterprises	0.41	0.49
4	1	median number of individual income taxes	0.18	0.22

Group (g)	Group (g')	State Variables	Transform Probability (P)	
			MIX	*MAX*
4	1	numbers of enterprises	0.19	0.21
23	4	median number of individual income taxes	0.07	0.13
23	4	numbers of enterprises	0.18	0.23
23	4	number of people aged 65 and older	0.26	0.35
4	23	median number of individual income taxes	0.22	0.32
4	23	numbers of enterprises	0.06	0.10
4	23	number of people aged 65 and older	0.33	0.43

A. *Model Accuracy*

The study conducted prediction on the average energy consumption per household of 907 agents during 2019-2030. Fig. 4 shows the cumulative distribution function (CDF) of the 2030 simulation value and 2019 actual value, with their CDFs crossing at 600 KWH/household. Compared with the 2019 CDF, the 2030 CDF has fewer agents with average household energy consumption lower than 600 KWH/household and more agents with average household power higher than 600 KWH/household, indicating an increasing trend for the number of agents with energy consumption growth.

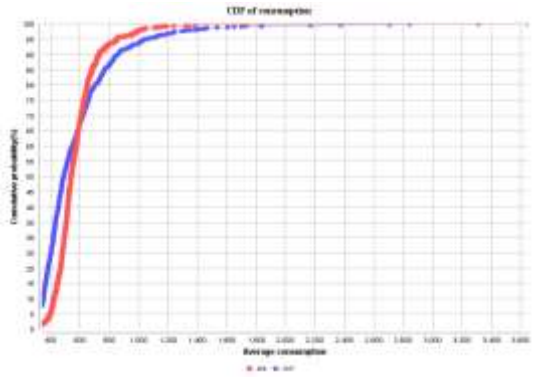


Fig. 4. Cumulative distribution function of study's prediction

Besides, the study carried out RMSE (root-mean-square error) and MAPE (mean absolute percentage error) verification with the 2019 actual value and forecast value, the smaller the two numerical values and more accurate the model's forecast. Model simulation is not good with MAPE exceeding 100%, compared with 10-20% for good simulation and 20-50% for reasonable simulation [16]. As shown in Table 2 on various groups' RMSE and MAPE values, group 3 has better model simulation, with RMSE and MAPE reaching 174.51 and 18.82%, respectively, while group 5 has the worst model simulation with RMSE and MAPE reaching 562.03 and 32.49%, since its power consumption value is an outlier in the studied area, with its agents mostly located in enterprise areas, which affects the accuracy of the ABM simulation significantly, due to focus of the study on residential energy consumption, excluding business energy consumption. It is advised to modify the model to include business energy consumption in the future.

TABLE II. PREDICTION ACCURACY

Group	Accuracy	
	RMSE	*MAPE (%)*
0	57.98	18.00
1	136.18	22.89
2	250.34	22.47
3	174.51	18.82
4	207.58	26.69
5	562.03	32.49

B. *Simulation Results*

The study employed ABM simulation to forecast average household power consumption of 907 agents in Taipei City during 2019-2030 (Fig. 5), grading energy power consumption into five levels, represented in colors, from green to red in ascending order. Fig. 5 shows that areas with higher power consumption and rising consumption in 2030 are situated in the center of the studied area (Taipei City's downtown areas), including Xinyi and Zhongzheng districts. The forecast also shows five districts with rising power consumption, namely Xinyi, Wanhua, Zhongzheng, Shilin, and Beitou, as well as seven districts with declining power consumption, namely Wenshan, Neihu, Daan, Nangang, Datong, Zhongshn, and Songshan.

Meanwhile, Fig. 6 exhibits various cluster (group) annual simulation results in a dashboard shape, visualizing the distribution of minimum and maximum values for the various clusters, using the average actual value in 2019 as the simulation standard value (target). Group 4 has a larger difference in the distribution of minimum and maximum values, with the average values of most agents' simulation values falling in the scope of 518-536 KWH/ household and only some agents having higher power consumption. In terms of various groups' annual simulation value cycle, group 2 has a shorter fluctuation in the simulation cycle of about three years. If a specific year's average forecast value (value at dashboard center) exceeds the target, it represents most agents' simulated energy consumption that year is higher than the simulation's start year, with the blue area in the dashboard surpassing the target line. Fig. 6 shows cluster 1's average simulation value exceeding the target every year, underscoring a rising energy consumption trend, different from cluster 2 with simulation value slightly lower than the target, cluster 3 near the target, and cluster 4 slightly higher than the target, all indicating small difference between annual simulation result and that of start year for the three groups.

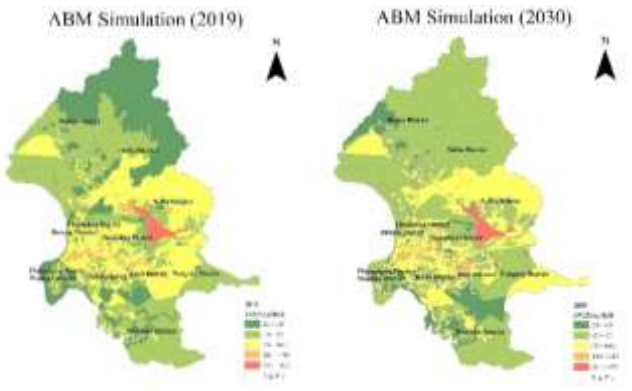


Fig. 5. Energy Consumption Map of ABM Simulation in 2019 and 2030.

Fig. 6. The Dashboard of Agent's Energy Consumption Prediction from 2019 to 2030

V. CONCLUSIONS

Uneven socioeconomic development, in terms of space and time, among various residential areas has caused heterogeneous features in urban resource consumption. ABM model, developed in recent years, can detect in detail changes in urban resource consumption. In order to assure the accuracy of its simulation results, it needs for the ABM model to take into account the transformation probability of agents' heterogeneous features and set the choice probability for agents' various behaviors.

The study analyzes urban energy consumption prediction patterns using agent-based modeling (ABM). For most areas by 2030 in the studied area (Taipei City), which means that there will not be much change in the city's energy consumption in the coming years, should its residents' energy consumption behaviors remain unchanged, such as for group 2, group 3, and group 4. Transformation of group 1 into group 4 has the highest probability rate, due mainly to change in the number of businesses, as evidenced by *P* value ranging 0.41-0.49, underscoring the significant influence of enterprises on the studied area. Meanwhile, the annual energy consumption forecast for group 4 shows the widening gap between the minimum value and maximum value, indicating possible gradual shift of its energy-consumption features to that of group 5. It is advised for future study to include business energy consumption behavior into the research scope, so as to augment the accuracy of forecast for overall regional energy consumption.

ACKNOWLEDGMENT

The author wishes to thank the students and research assistants (Kuang-Ly Cheng, Pao-Hsuan Huang, and Nai-Hui Wang) for implementing the ABM project. This work was supported by Grant MOST 109-2410-H-034-026.

REFERENCES

[1] N. Fumo and M. A. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," Renewable and Sustainable Energy Reviews., vol. 47, pp. 332-343, July 2015.

[2] M. Dijst et al., "Exploring urban metabolism-Towards an interdisciplinary perspective," Resour. Conserv. Recycl., vol. 132, pp. 190-203, May 2018.

[3] S. Basu, C. S. E. Bale, T. Wehnert, and K. Topp, "A complexity approach to defining urban energy systems," Cities, vol. 95, pp. 102358, December 2019.

[4] P. van den Brom, A. Meijer, and H. Visscher, "Performance gaps in energy consumption: household groups and building characteristics," Build. Res. Inf., vol. 46, pp. 54-70, May 2017.

[5] Y. Zhang, X. Bai, F. P. Mills, and J. C. V. Pezzey, "Rethinking the role of occupant behavior in building energy performance: A review," Energy and Buildings., vol. 172, pp. 279-294, August 2018.

[6] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," Renew. Sustain. Energy Rev., vol. 43, pp. 901-917, March 2015.

[7] G. Polinesi et al., "Population trends and urbanization: Simulating density effects using a local regression approach," ISPRS Int. J. Geo-Information, vol. 9, pp. 454, July 2020.

[8] A. Yu, J. You, H. Zhang, and J. Ma, "Estimation of industrial energy efficiency and corresponding spatial clustering in urban China by a meta-frontier model," Sustain. Cities Soc., vol. 43, pp. 290-304, November 2018.

[9] L. Anselin, "Local indicators of spatial organization -LISA," Research, vol. 27, pp. 1-25, 1995.

[10] I. C. Chen, K. L. Cheng, H. W. Ma, and C. C. W. Hung, "Identifying spatial driving factors of energy and water consumption in the context of urban transformation," Sustain., vol. 13, pp. 10503, September 2021.

[11] D. G. Brown and D. T. Robinson, "Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl," Ecol. Soc., vol. 11, pp. 46, 2006.

[12] M. D. A. Rounsevell, D. T. Robinson, and D. Murray-Rust, "From actors to agents in socio-ecological systems models," Philos. Trans. R. Soc. B Biol. Sci., vol. 367, pp. 259-269, 2012.

[13] C. S. E. Bale, L. Varga, and T. J. Foxon, "Energy and complexity: New ways forward," Appl. Energy, vol. 138, pp. 150-159, January 2015.

[14] C. Peter and M. Swilling, "Linking complexity and sustainability theories: Implications for modeling sustainability transitions," Sustain., vol. 6, pp. 1594-1622, March 2014.

[15] V.Grimm et al., "A standard protocol for describing individual-based and agent-based models," Ecol. Modell., vol. 198, pp. 115–126, September 2006.

[16] C. D. Lewis, Industrial and Business Forecasting Methods : A Practical Guide to Exponential Smoothing and Curve Fitting, London: Butterworth Scientific, 1982.

BLE Mesh using CODED PHY

Javier Silvestre-Blanes
DISCA. ITI.
Universitat Politècnica de Valencia (UPV)
Alcoy, Spain.
jsilves@disca.upv.es

Juan Carlos García Ortiz
Instituto Tecnológico de Informática (ITI)
Valencia, Spain
juagaror@iti.es

Víctor M. Sempere-Payá
DCOM. ITI
Universitat Politècnica de Valencia (UPV)
Valencia, Spain.
vsempere@dcom.upv.es

David Cuesta Frau.
DISCA. ITI.
Universitat Politècnica de Valencia (UPV)
Alcoy, Spain.
dcuesta@disca.upv.es

***Abstract*—The increasing use of wireless technologies in Industry 4.0 is a reality and a current necessity. While the foreseeable use of 5G networks will play an important role in this field, the use of different communication networks in Industrial Internet of Things (IIoT) networks is nowadays an active field of research, since it is considered that the industry of the future will be supported by a heterogeneous set of network technologies. In this paper we demonstrate how the use of CODED PHY coding in BLE mesh networks can reduce the number of relays, overload in the medium and energy use, as well as improve the range of these networks while improving the packet delivery rate. In a low node density grid scenario, a node reduction of 85% is achieved, while improving the PDR (Packet Delivery Ratio) obtained without using BLE Mesh forwarding mechanisms, reducing energy consumption and media saturation.**

Keywords— Bluetooth Low Energy, Mesh Networks, IIoT

I. INTRODUCTION AND RELATED WORK

The fourth industrial revolution, the so-called industry 4.0, will not only transform production methods, but will also have an impact on society in general. This revolution is based on emerging technologies such as Big Data Analytics and Internet of Things (IoT), allowing the development of an intelligent network that permeates all stages of production. These networks provide interoperability that can be categorized into vertical, horizontal and end-to-end integration. Vertical integration provides a connection between different subsystems within a company, including sensors, actuators, management and planning. In Industry 4.0, communication networks are becoming increasingly heterogeneous from the point of view of the technologies used [1] [2], and where wireless networks play an important role in achieving the required flexibility. The use of Industrial Internet of Things (IIoT) requires the use of Wireless Sensor Networks (WSNs [3], such as Zigbee or TSCH, based on 802.15.4 and 802.15.4e respectively) to provide communications to the sensors with the sink in charge of processing or transmitting the collected information to the next level. Mesh networks are used to provide device-to-device (D2D) communications (see Fig. 1), but they can also be used to form Wireless Sensor Networks. Bluetooth Low Energy (BLE) has interesting characteristics to be used in this type of application, since it has a low consumption and cost, while the range is greater than in other networks of this type, taking further advantage of its ability to form meshed networks. This type of

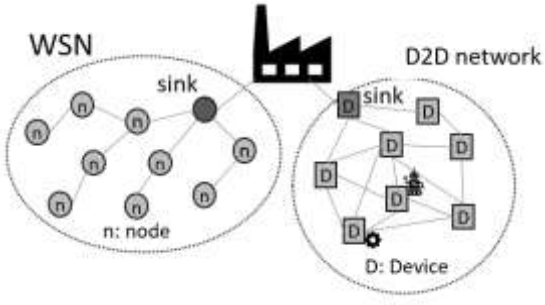


Fig. 22 Wireless network configurations in industry

network, the BLE mesh (Bluetooh Mesh - hereafter referred to as BM), has been used in precision agriculture applications [4], the use of robots in logistics [5], in industrial control applications [6] and industry 4.0 [2]. BM is based on BLE v4 advertisement, which can only use LE_1M encoding (Low Energy, 1 Mbps, see Table I) to maintain compatibility with 4.0 devices. This limits the ability to increase the range, as well as improve the robustness of communications, since these properties rely on the use of LE_CODED (Low Energy Coded) coding (see Table I). We used the coding used in LE_CODED, but without the rules used in Bluetooth 5.0, and denominate it as CODED PHY. We broadcast Bluetooth Low 4.0 compatible frames (advertisements) but using this radio mode, instead of using Bluetooth Low 5.0 compatible (extended) frames with this radio mode which is called LE_CODED. The main reason is to get the maximum communication robustness. Although the LE_CODED radio is designed to increase the range and reliability of the communications, this mode uses two packets for a message transmission. The first one is sent over the primary channels (see Fig. 2) and it contains a pointer offset to the second one that contains the data that is sent over a secondary channel. Sending information over secondary channels is more likely to be affected by interference from other protocols, increasing the likelihood of packet loss. The secondary channels can be affected considerably when WiFi networks coexist. Thus, the CODED PHY mode avoids these problems by using the robust radio mode while maintaining the same packet structure and operation of the BLE 4.0 specification. The CODED PHY mode increases reliability by using only one frame to send the content, instead of the two packets (pointer + data) used by BLE 5.0

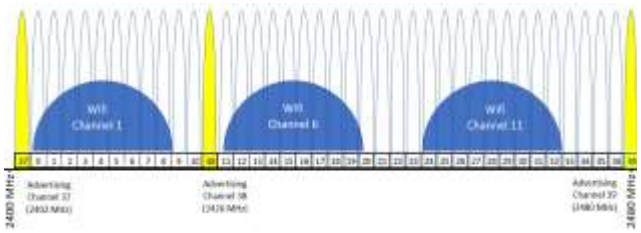


Fig. 23 Bluetooth channels

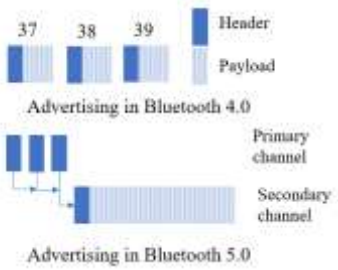


Fig. 24 Bluetooth advertising

LE_CODED specification (see Fig. 3). Finally, the CODED PHY could be implemented easily with a simple modification to the BLE Mesh protocol, while LE_CODED could require substantial modifications to some protocol layers.

TABLE I. PHYSICAL LAYER PROPERTIES AND CODING SCHEMES

	Coding		
	1M	***2M***	***CODED***
Symbol rate	1 Ms/s	2 Ms/s	1 Ms/s
Data rate	1 Mb/s	2 Mb/s	500 or 125 Kb/s
AccessHeader	Uncoded	Uncoded	S=8
Payload	Uncoded	Uncoded	S=2 or S=8
Range Multiplier	1	0.8	2 or 4
Sensitivity (dBm)	< -70	< -70	< -75 or <-82
BLE Versions	4.0-5.2	4.0-5.2	5.0-5.2

In [7] two methods for improving performance based on modifying the protocol stack to allow the transmission of larger data structures are presented. In [8] BLE 5.0 and LE_2M coding are used to implement industrial applications, proposing a more efficient flooding mechanism than the one proposed in the standard. BLE is considered a good candidate for several industrial applications and enhance the support of Real-Time communications [9]. In [10] the use of the encodings provided in BLE 5.0 for Intelligent Transport Systems Applications with a line topology was evaluated.

This paper proposes and analyses the use of CODEC PHY coding in BLE mesh networks focused on industrial applications, using a grid topology [4] [11]. The objective is to evaluate the following: the reduction in the number of devices (relays) compared to a network based on LE_1M; the reduction in the number of retransmissions necessary due to the coding and because they are meshed networks with different alternatives, which implies less energy expenditure and less saturation in the network; and the greater robustness and reliability of the communications. The remainder of this work is structured as follows. Section II presents an overview of BLE mesh. Finally, in section III the testbed, the measurement results and conclusions are presented.

II. BLE MESH

Mesh networks offer strong robustness and tolerance to node failures in a self-managed way, which makes their use advantageous in changing environments where the paths and channels for sending packets vary constantly. There are different technologies and products on the market that are based on this philosophy as Zigbee [12], Thread [13], Z-Wave [14] or WirelessHART [15]. One of them is by Bluetooth SIG and is called Bluetooth Mesh [16]. This specification is based on Bluetooth Low Energy and details a protocol stack differentiated in layers, from the highest level to the lowest (see Fig. 4):

- Model and Foundation Model Layer: are two layers that standardize the behavior of certain applications to ensure their interoperability. The first refers to applications/user behavior (of which there are different models). The second refers to specific models related to management and configuration.
- Access Layer: which oversees defining the format and fields of the application messages in a generic way for all models.
- Upper and Lower Transport Layer: in which the former manages message encryption and authentication while the latter oversees the packet segmentation and reassembly process.

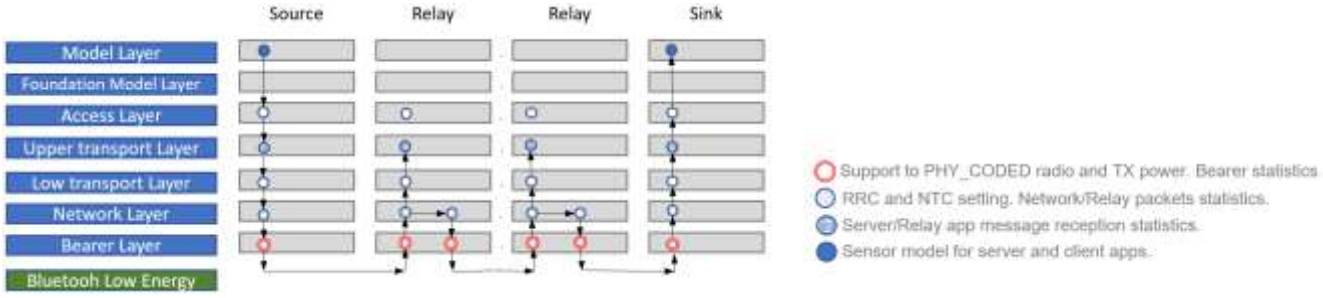


Fig. 25. BLE system architecture and modifications made at each layer

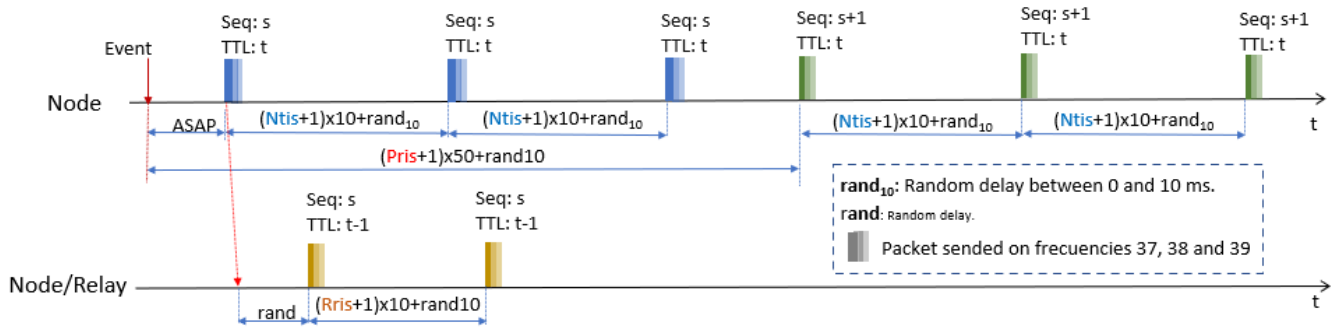


Fig. 26. Timing of message sending

- Network Layer: which is responsible for routing and forwarding packets, as well as encrypting them at the network level. Also, it implements the relay mechanism.
- Bearer Layer: which indicates how packets are sent and received through the available network interfaces. It is intended to support different types, although the current specification is based on the use of BLE 4.0 advertisements.

When using the Bluetooth Low Energy Bearer layer, the BLE Mesh protocol is placed on top of it. For compatibility reasons, the specification restricts the use of Bluetooth 4.x type frames to only radio mode LE_1M, not supporting in a normative way any other use or change on the Bluetooth Low Energy protocol itself. Maintaining this BM compatibility with Bluetooth 4.x makes it impossible to take advantage of the new improvements in communication robustness and range. Also, and due to that restriction, the maximum amount of application data sent in a packet is limited to 8-11 bytes (depending on the application and message type), although the protocol allows to send larger messages by the packet segmentation and reassembly mechanism implemented at Transport Layer.

To support a Mesh topology, the routing protocol is based on managed flooding, thus eliminating the use of complex routing algorithms and the need to make connections between nodes for both information exchange and route management. In this way, each message received by a node is rebroadcast by that node without prior knowledge of the number and status of neighboring nodes. Even so, and to avoid a possible saturation of messages in the network, some strategies are implemented to optimize performance, such as:

- Relay nodes: in which each node in the network initially acts as a repeater, but it is possible to establish if a node wishes not to be a repeater.
- Relay cache: in which each node of type Relay maintains a list of packets already received/diffused. If a packet is found in this list then it is determined not to broadcast it.
- Time To Live: where each packet has a limited maximum number of retransmissions. Once this number is reached, no more retransmissions are made.

Security is a key concept on the design of Mesh protocols. Bluetooth Mesh introduces several mechanisms at different layers to provide the maximum security against different attacks. Some of these are:

- Secure Device Provisioning: in which each node added to the network is performed by a secure process mechanism.
- Encryption, authentication, and message obfuscation: In which all messages are encrypted and authenticated using different keys (device, network, and application). Also, messages are obfuscated to make it difficult to trace the nodes and network.
- Protection against Replay and Trashcan attacks: in which secure sequence number for packets and a safe node removal from the network mechanisms are provided.

In addition to the above, the protocol has a series of parameters that allow a greater degree of fine tuning and also control the balance between the level of network saturation and the robustness and reliability of the network, since they affect the number and times at which a message is broadcast. These parameters are:

- Publish Retransmit Count (PRC) and Publish Retransmit Interval Steps (PRIS): which indicates the number of times that an application message is retransmitted and the time instants in which they are performed. A message is retransmitted PRC + 1 times, i.e., a PRC value of 0 indicates that the message is only transmitted once. The repetition interval in milliseconds is set as *(1 + PRIS) x 50*, the default PRIS value is 1.
- Network Transmit Count (NTC) and Network Transmit Interval Steps (NTIS): which indicates the number of retransmissions that a packet is broadcast and the time instants in which they are made. A packet is broadcast NTC + 1 times, i.e., an NTC value of 0 indicates that the message is only transmitted once. On the other hand, the repetition interval in milliseconds is set as *(1 + NTIS) x 10*, the default NTIS value is 1.
- Relay Retransmit Count (RRC) and Relay Retransmit Interval Steps (RRIS): which indicates the number of retransmissions that a packet is retransmitted by a Relay and the time instants in which they are performed. A message is retransmitted RRC + 1 times, i.e., an RRC value of 0 indicates that the packet is only retransmitted once. On the other hand, the repetition interval in

milliseconds is set as *(1 + RRIS) x 10*, the default RRIS value is 0.

The effect of these parameters can be seen in Fig. 5, using (NTC=2; PRC=1; RRC=1). When an event occurs that involves the transmission of a packet, it is generated as fast as possible (ASAP) and transmitted on channels 37, 38 and 39. This is retransmitted twice at network level (in blue in the figure). In addition, at application level and after the time defined by PRIS has elapsed, it is retransmitted again (PRC=1, in green in the figure), and will be retransmitted three times as NTC=2 with the updated seq value. The nodes close to the sender that act as relays, when receiving the message, will retransmit it, discounting a TTL value, and will do it twice if RRC=1. The other network nodes within range of the transmitting node, if they have also correctly received the first message, will receive, and discard the other 5 messages sent by the node, and the 2 sent by the relay in the figure.

At the application level, and depending on the definition of the models themselves, there are the roles of Client (make requests and receive their responses) and Server (respond to requests or broadcast their status). Each individual model defines the message types, data content and interaction for each role. This allows the interoperability between nodes that implement and integrate the same models. Finally, there are additional advanced features (which do not form part of the objectives of this article) focused on interoperability and/or reduction of energy and resource consumption, such as the ability to define nodes of type Proxy (allows interaction with devices that not support Bluetooth Low Energy), Friendship and Low Power (that allow Friendship to resent messages from Low Power Nodes when they wake-up).

III. TESTBED, RESULTS AND CONCLUSIONS

BM performance is evaluated through simulations developed in a proprietary software developed in Python (as in [17]), as well as in real testbed (although with a more limited number of nodes). This simulator implements the basic operation of the transport, network and bearer layers (relay control, TTL, timers, configuration values, etc.) allowing us to keep track of the packets that are sent and generated through the simulation. To simulate the range of the signal it is specified through a delivery probability value to the neighboring nodes according to their distance from the sender. A working area of 90x90 meters was chosen, and (NTC=PRC=RRC=0) to avoid retransmissions and to see only the impact of having several routes available. In the simulation the Packet Delivery Rate per link used is ($PDR_{link}i$) as shown in Table II. The distribution of the nodes in a grid topology will allows us to characterize the problem without losing its generality. The parameter that will determine the number of nodes will be the density of sources per square meter. The number of relays will depend on the BLE physical layer used. To illustrate this, and using 1M coding, we consider a first scenario with a source density of one source per 1000 m^2 approximately, as shown in Fig. 6.a (1MLD: LE_1M, Low Density). In this case, 40 relays would be needed to provide connectivity to all nodes. In a randomly distributed environment, the problem of node deployment would arise [18] [19] [20]. In a second scenario, with a density of one source

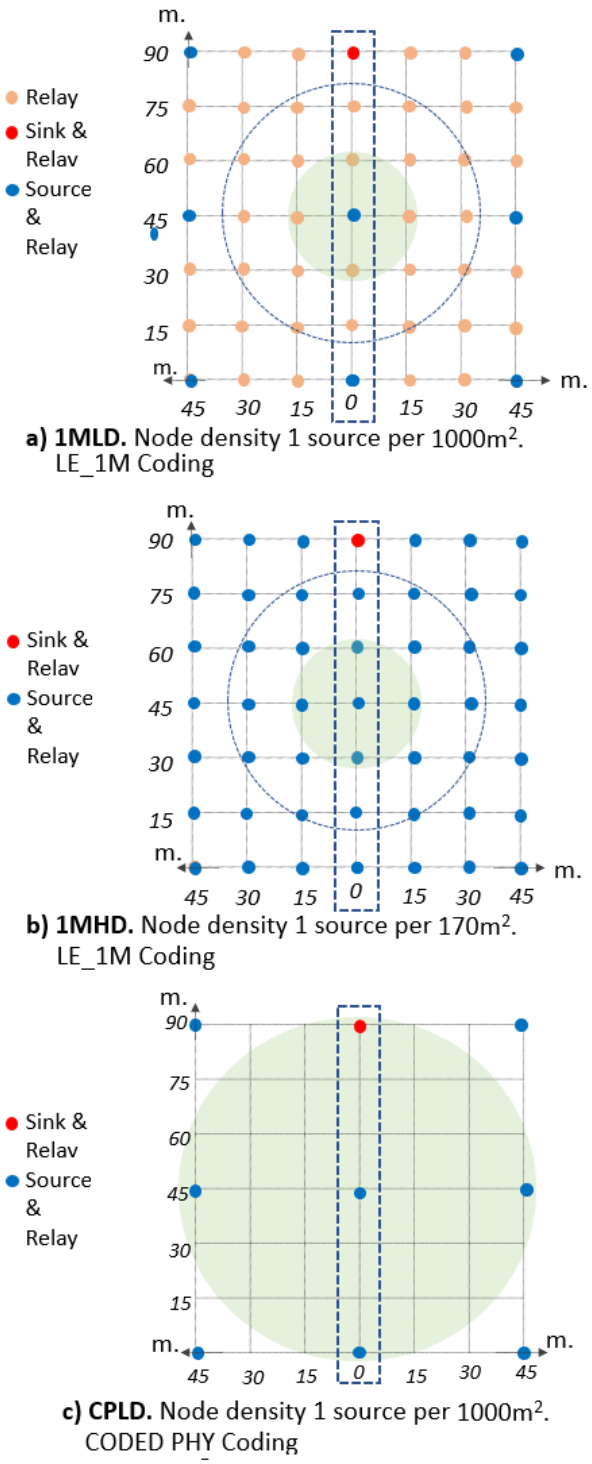


Fig. 27. Scenarios

every 170 per m^2, no relay-only nodes would be needed, as can be seen in Fig. 6.b (1MHD: LE_1M, High Density). In a randomly distributed environment, the problem in this case would be the relay selection [11] [21] [22]. By using CODED PHY coding, the range is greatly increased, so that in the first case, no relay is necessary, as shown in Fig. 6.c (CPLD: CODED PHY, Low Density). This coding is based on the use of robust

coded radio coding introduced from Bluetooth Low Energy 5.0,

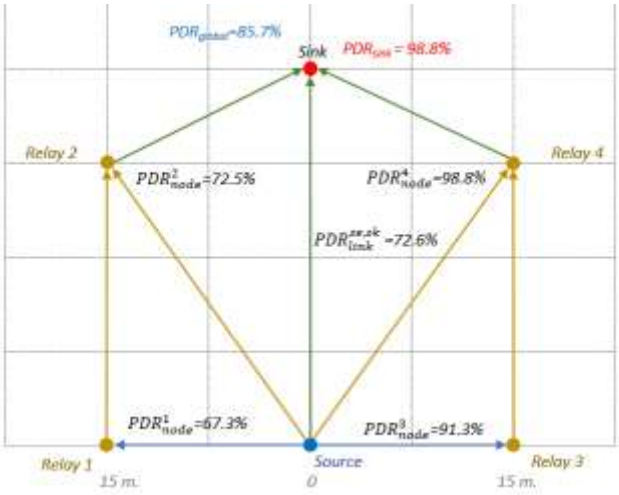


Fig. 7. Wireless network configurations in industry

as previously stated. In all three cases simulations have been made only with the central part to facilitate comparison with the real test. (Dotted line in Fig. 6).

To illustrate these concepts, a test was performed with the real testbed as shown in Fig. 7. Only one source and one sink were used, plus 4 relays. The $PDR_{link}^{se,sk}$, to the separately measured direct source-sink connection is shown. We call the PDR_{node}^{i}, to the PDR obtained at each relay i based on the number of messages forwarded compared to the theoretical number. It does not measure the link connection, so it is not known if the forwarded packets come from the source or from another relay. We also have the final PDR obtained in the sink, named PDR_{sink}, and the PDR_{global} taken as the average of the PDR_{node}^{i} (including that of the sink). The theoretical total number of packets sent (TTN) will be:

$$TTN=(nrl \times nsr \times np + nsr \times (nsr-1) \times np)$$

$$=(nrl + nsr - 1) \times nsr \times np$$

where *nrl* is the no. of nodes that are only relays, *nsr* the no. of sources (which are also relays), and *np* the no. of packets sent per source. This theoretical value (TTN) will be higher than the real value (RTN) as packet losses occurs. Therefore, we have RTN = TTN × PDR_{global}. In the real test in Fig. 7, the PDR_{global} value is much lower than the one obtained in the sink, PDR_{sink}. This difference is due to the difference in the quality of the links that make up this testbed, with the links through Relay 3 and 4 having a very good quality, while in Relay 1 and 2 this value is significantly lower.

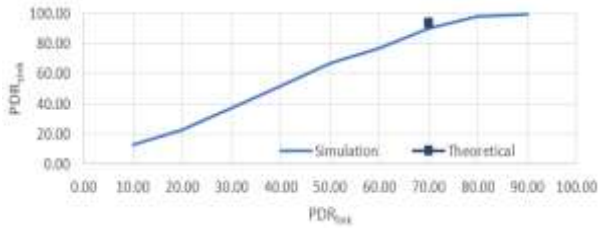


Fig. 8. PDR_{sink} depending on PDR_{link}

The value of PDR_{sink} can be estimated from the PDRs of the remaining links. The largest contribution to the PDR_{sink} is from the direct link between source and sink $PDR_{link}^{se,sk}$, plus the packets arriving by alternative routes. Specifically:

$$PDR_{sink} = PDR_{link}^{se,sk} + \left(\prod_{i=1}^{n} PDR_{node}^{i}\right)^{2}$$

TABLE II. PDR OBTAINED FOR DIFFERENT SCENARIOS

	Simulated			
scenario	**PDR_{link}1**		**PDR_{link}2**	
	PDR_{global}	***PDR_{sink}***	***PDR_{global}***	***PDR_{sink}***
1MLD	99.5	99.0	92.1	87.9
1MHD	99.28	99.5	87.9	91.7
1MLDc	82.6	89.6	36	32
1MHDc	94.2	93.5	43	26
scenario	**PDR_{link}3**		**PDR_{link}4**	
CPLD	98.9	99.1	77.0	80.6
CPLDc	86	83.1	50	46
scenario	**Real testbed**		**PDR_{link}1**: 15m. 90%. 30m. 65% **PDR_{link}2**: 15m. 50%. 30m. 25% **PDR_{link}3**: 45m. 90%. 60m. 55% **PDR_{link}4**: 45m. 50%. 90m. 25%	
	PDR_{global}	***PDR_{sink}***		
1 MLDc	57.9	50.5		
CPLDc	97.4	99.8		
CPLD	75.6	96.3		

In Fig. 8 shows the evolution of the PDR_{sink} depending on the same PDR_{link} value for all links, while the PDR_{sink} obtained from the previous equation and the testbed values of Fig. 7. The results obtained with the simulations and tests in Fig. 6 are analyzed below. Regarding the overload in the medium, which are the packets received by each node regardless of who they are addressed to with respect to those broadcasts by the network, in the grid testbed an increase in the number of packets received of 2.5 was obtained, while in the simulations an increase of 6.97 was obtained in 1MLD for $PDR_{link}1$ and 3.5 for $PDR_{link}2$, while for CPLD a value of 2.4 was obtained. The high value in the 1MLD scenario for $PDR_{link}1$ is due to the high probability that several nodes receive each packet and proceed to rebroadcast it, due to the density of relay nodes in the grid to provide the necessary range using LE1M coding. This demonstrates the reduction in media overload if CODED PHY encoding is used, which would also be reflected in lower energy consumption. Concerning the quality of communications, we can highlight the high value of PDR_{sink} for the 1MLD and 1MHD scenarios with $PDR_{link}1$, and with somewhat lower values for $PDR_{link}2$. This is due to the high probability that several nodes receive the packets and proceed to rebroadcast them, given the high density of relays (in 1MLD) and nodes/relay (in 1MHD) needed to provide the necessary reach using LE_1M.

In the case of CPLD, high values are also obtained despite using a much lower number of nodes compared to 1MLD as it does not need relays, and despite having fewer paths and alternatives to reach the destination. This is logical given the greater robustness and reach of CODED PHY. The use of only

the central part of the scenarios shows a degradation in the PDR_{link} compared to the grid scenario, which is more pronounced the worse the PDRs of the links are. This is due to the reduction in the number of paths to reach the sink. In the testbed, the 1MLD scenario can only be tested in the central part, using 7 nodes. An intermediate result to the one calculated in the simulations using $PDR_{link}1$ and $PDR_{link}2$ was obtained. This behavior is logical in a real scenario, since it is impossible for all of them to be at exactly the same distance, with the same tilt and gain in the antennas, supplying the same power, etc. Thus the value of 90% at 15 meters does not occur in reality in all links. In the CPLD scenario, very similar values are obtained in the central part and with the full scenario. Despite the tripling of the distance between nodes compared to 1MLD, and the fact that there are fewer alternative routes, the greater robustness of this coding means that, starting from acceptable PDR_{link}, a high PDR_{sink} is achieved in the end. The value of PDR_{global} is also significantly lower than PDR_{link} in this other test, which indicates the non-uniformity in the deployment of the links.

In conclusion, the use of CODEC PHY as the coding method in mesh networks instead of 1M used to maintain compatibility allows:

- More reliable communication and better PDR
- Reduction of the number of retransmissions ((NTC=PRC=PRC=0) by taking advantage of the alternative routes of the mesh network, which has a significant impact on reducing the energy consumption of the network.
- Reduction of the number of nodes required in low density environments from 47 nodes in 1MLD to 7 nodes in CPLD, which is a decrease of 85%. This implies a reduction in the cost of deployment, and in the energy consumption of the network.

As future work, we aim to work on obtaining an analytical method to obtain the PDR_{sink} from the different PDR_{link}, routes and parameters (PRC; NTC; RRC) to facilitate the deployment and configuration of the system, and to design the most efficient topology with the best balance between the number of transmitter nodes and relays, the system overload, and the reliability in the reception of messages in the sink. For this, the simulator has to be improved and some levels of the deployment have to be adapted in order to extract more information about the links and relays of the nodes.

ACKNOWLEDGMENT

This work was supported by Grant PGC2018-094151-B-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe.

REFERENCES

[1] Stefano Scancio, Lukasz Wisniewski, Piotr Gaj. “Heterogeneous and dependable networks in Industry – A Survery”, Computers in Industry, vol. 125, p 103388, 2021.

[2] Celia Garrido-Hidalgo, Diego Hortelano, Luis Roda-Sanchez, Teresa Olivares, M. Carmen Ruiz, Vicente Lopez.“IoT Heterogeneous Mesh Network Deployment for Human-in-the-Loop Challenges Towards aSocial and Sustainable Industry 4.0” IEEE Access vol. 6, 2018.

[3] Saleem Raza, Muhammad Faheem, Mesut Guenes.“Industrial Wireles Sensor and Actuator Networks in Industry 4.0: Exploring requirements, protocols and challenges - a MAC survey”, Int. Journal Communication Systems, 32:e4072, 2019J.

[4] Luan H.S. Alves, Elida Antunes, Ricardo Ferreira, Jos´e A. M. Nacif. “A Mesh Sensor Network based on Bluetooth: Comparing Topologies to Crop Monitoring,” in IX Simp´osio Brasileiro de Engenharia de Sistemas Computacionais, November, 2019.

[5] Adnan Aijaz. “Infrastructure-less wireless connectivity for mobile robotic systems in logistics: why Bluetooth mesh networking is important,” in 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), September, 2021.

[6] N. Xia, C. Hsiao-Hwa, C.S. Tang. “Emerging Technologies for Machinetype communication networks”, IEEE Network, vol. January/February, 2020

[7] D. Pérez-Díaz-de-Cerio, A. Hernández-Solana, M. García-Lozano, A. Valdovinos Bardaji, J.L. Valenzuela. “Speeding up Bluetooh Mesh”, IEEE Access, vol. 9, 2021.

[8] Usman Raza, Aleksander Stanoev, Charles Khoury, Alexandru-Loan Pop, Mahesh Sooriyabandara. “Demo: Synchronous Transmissions Based Flooding over Bluetooth 5.0 for Industrial Wireless Applications” Int. Conf. on Embedded Wireless Systems and Networks, 2019.

[9] Luca Leonardi, Gaetano Patti, Lucia Lo Bello. “Multihop Real-Time Communications Over Bluetooth Low Energy Industrial Wireless Mesh Networks” IEEE Access vol. 6, 2018.

[10] Juan Carlos García Ortiz, Javier Silvestre-Blanes, Víctor M. Sempere-Payá, David Cuesta Frau. “Evaluation of improvements in BLE Mesh through CODED PHY,” in 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), September, 2021.

[11] Marco Reno, Raúl Rondón, Lucia Lo Bello, Gaetano Patti, Aamir Mahmood, Alfio Lombardo, Mikael Gidlund. “Relay node selection in bluetooth mesh networks,” in 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), June, 2020.

[12] Stanislav Safaric, Kresimir Malaric. “Zigbee wireless standard” Proceedings of ELMAR, Zadar, Croatia, 2006.

[13] Ishaq Unwala, Zafar Taqvi, Jiang Lu. “Thread: An IoT Protocol” IEEE Green Technologies Conference, April, 2018

[14] Muneer Bani Yassein, Wail Mardini, Ashwaq Khalil. “Smart homes automation using z-wave protocol”, Int. Conf. on Engineering & MIS (ICEMIS), September, 2016

[15] P. Arun Mozhi Devan, Fawnizu Asmadin Hussin, Rosdiazli Ibrahim, Kishore Bingi, Farooq Ahmad Khanday. “A Survey on the Application of WirelessHART for Industrial Process Monitoring and Control”, Sensors 2021, 21(15), 4951

[16] Angela Hernández-Solana, David Pérez-Días-de-Cerio, Mario García-Lozano, Antonio Valdovinos Bardaji, José-Luis Valenzuela. “Bluetooth Mesh Analysis, Issues, and Challenges” IEEE Access vol. 8, 2020.

[17] Lars Almon, Flor Alvarez, Laurenz Kamp, Matthias Hollick. “TheKing is Dead Long Life the King!. Towards Systematic Performance Evaluation of Heterogeneous Bluetooth Mesh Networks in Real World Environments”. IEEE 44th Conf. on Local Computer Networks, 2019.

[18] Changsu Jung, Kyungjun Kim, Jihun Seo, Bhagya Nathali Silva, Kijun Han. “Topology Configuration in Multihop Routing protocol for Bluetooth Low Energy Networks” IEEE Access vol. 5, 2017.

[19] Jia Mao, Xiaoxi Jiang, Xiuzhi Zhang. “Analysis of node deployment in wireless sensor nwtworks in warehouse environment monitoring ystems”.EURASIP Journal of Wireless Communications and Networking. 2019:288.

[20] Hashim. A. Hashim, B. O. Ayinde, M.A. Abido. “Optimal placement of relay nodes in wireless sensor network using artificial bee colony algorithm”. Journal of Network and Computer Applications 64, pp. 239-248, 2016.

[21] Emil A.J. Hansen, Martin H. Nielsen. Daniel E. Serup. Robin J. Williams. “On Relay Selection Approaches in Bluetooth Mesh Networks”. 10th Int. Cong. on Ultra Modern Telecommunications and Control Systems and Workshops. 2018.

[22] Uyoata Uyoata, Joyce Mwangama, Ramoni Adegun. “Relaying in the Internet of Things (IoT): A Survey” IEEE Access vol. 9, 2011.

Application of Hydrogen Fuel-Cell for Auxiliary Power Unit (APU) on aircraft

Truc-Quan Ngo
Faculty of Aviation Engineering
Vietnam Aviation Academy
Ho Chi Minh city, Vietnam
1951200039@vaa.edu.vn

Nhu Tran
Faculty of Aviation Engineering
Vietnam Aviation Academy
Ho Chi Minh city, Vietnam
nhuttq@vaa.edu.vn

***Abstract*—Hydrogen fuel cell is an electrochemical device that converts potential energy from hydrogen fuel into electricity through an electrochemical reaction. Fuel cells produce electricity uninterruptedly with constant intensity when fuel is continuously loaded. The hydrogen fuel cell provides high conversion efficiency and does not create pollution to the environment. The Auxiliary Power Unit (APU) of an aircraft provides indirect power and compressed air to operate aircraft equipment when the main engines are idle. This paper analyzes the APU emission index to show the potential use of hydrogen fuel cells as an alternative green energy.**

***Keywords*—**Hydrogen fuel cell, APU, APU emissions, green energy.

I. INTRODUCTION

A. *Hydrogen fuel cell*

Hydrogen fuel cell is an electrochemical device which is capable of converting chemical energy directly into electrical energy utilizing a redox process. The fuel is usually hydrogen gas and oxygen gas or air [1].

Hydrogen fuel cell is a static electricity generator, running quietly, making no noise, creating no pollution to the environment, and has a very high efficiency if a hydrogen fuel cell is used to generate both electricity and heat, it can reach 90% efficiency [2].

Batteries are the most common type of electrified device we use every day. When the chemicals stored in the battery that used for electricity conversion run out, the battery will have to be discarded. However, hydrogen fuel cells convert energy directly from the chemical reaction that fuses hydrogen and oxygen into water to generate electricity in the external circuit and heat energy for the engine to work. This is an open system, requiring a constant supply of fuel during operation, so the hydrogen fuel cell will last forever [3].

B. *Working principle of hydrogen fuel cell*

The working principle of hydrogen fuel cells is simply [4] [1] a reverse process of the electrolysis of water. Hydrogen and oxygen are combined to form water, which generates electricity and heat without releasing other substances that pollute the environment.

Fig. 1 is drawn based on the principle [5] that illustrate the working process of a hydrogen fuel cell.

Fuel (hydrogen gas) is continuously fed to the anode while the oxygen is introduced into the cathode.

At the anode: hydrogen gas passes through the catalytic membrane under the action of pressure. When a hydrogen molecule comes into contact with Pt, it splits into $2H^+$ and $2e^-$ [3].

$$H_2 \rightarrow 2H^+ + 2e^-$$

At the cathode: The electrons are released from the anode through the external circuit to the cathode, and H^+ ions diffuse through the electrolyte solution to the cathode surface and combine with the oxygen gas to create water. As the result, a current is generated in the external circuit [3].

$$O_2 + 4e^- + 4H^+ \rightarrow 2H_2O$$

General reaction: $H_2 + 0.5\ O_2 \rightarrow H_2O$

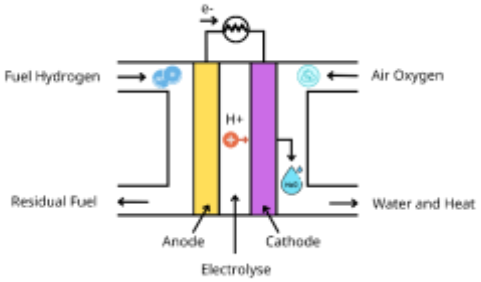


Fig. 1. Principle of a hydrogen-fuelled fuel cell

C. Basic structure

Fuel cells have three layers side by side:

- The first layer is the fuel electrode
- The second layer is the proton conduction electrolyte
- The third layer is the oxygen gas electrode

The two electrodes are made of conductive material. The surface of the electrodes is coated with a very thin layer of platinum catalyst.

When the hydrogen fuel cell is operated under load, the actual generated potential difference between the anode and cathode electrodes is about 0.7V. Therefore, to be able to provide the desired higher voltage or current, It is necessary to place multiple batteries in series or parallel to form a fuel cell stack [6]. In addition to the fuel cell assembly, the hydrogen fuel cell system also requires another auxiliary system: the compressor, the pump to supply the inlet gases, the heat exchanger, the requirements testing system, the reliability of engine operation, fuel storage and handling systems [3]. Figure II illustrates the hydrogen fuel cell system which is used on an aircraft [7].

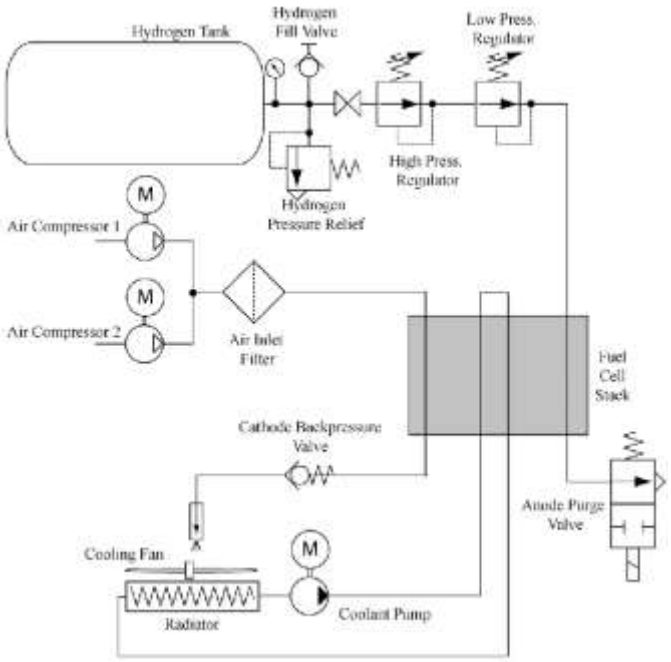


Fig. 2. Structural Diagram of An Aircraft Fuel Cell System

D. Potential applications of hydrogen fuel cells in aviation industry

By the 1960s of the twentieth century, General Electric Company has produced hydrogen fuel cell power supply systems for NASA's Apollo spacecraft, later used for Apollo-Soyuz, Skylab and space shuttles.

At AIRBUS, with the ambition to develop the world's first zero-emissions commercial aircraft powered by hydrogen propulsion, all three ZEROe ideas are hybrid-hydrogen aircraft. They are powered by the combustion of hydrogen through improved gas turbine engines. Liquid hydrogen is used as a fuel for combustion with oxygen. In addition, the hydrogen fuel cell generates additional electricity for the gas turbine, creating a more efficient hybrid-electric propulsion system [8].

Some applications of hydrogen fuel cells in aviation and more specifically airports and aircraft:

- Aircraft Auxiliary Power Unit (APU)
- Airside road equipment (personal cars, vans, and pick-ups, shuttle buses, trucks, and tractors)
- Ground Power Unit (GPU)
- Gate handling piece of equipment, fuel trucks, catering, and water trucks.

This paper studies the possibility of using hydrogen fuel cell for APU. There has been some researches on the development of APU in the world [1] [9]. However, the study on the use of hydrogen fuel cells to supply power for aircraft is still limited.

II. APU OPERATIONS AND EMISSIONS

A. Auxiliary Power Unit

The APU is essentially a relatively small turbine engine in the form of a centrifugal compressor, which is carried by large aircraft, both commercial and military. The APU supplies electricity and compressed air indirectly to operate aircraft equipment when the main engines are switched off. At startup, the APU runs at a constant speed. When there is a problem, the APU turns off automatically [10].

The APU can act as an internal backup source during flight and primary power while the aircraft is on the ground to provide cabin air conditioning [10]. The APU is typically used to start an aircraft's main engines. It ignites jet fuel, and thus produces noise and emissions like that of an aircraft's main engine [9].

When the aircraft parks, the Ground Power Unit (GPU) provides electricity to the aircraft while the Air Climate Unit (ACU) is used to provide air conditioning to the cabin and compressed air for main engine start up. However, the aircraft that parks at a remote stand usually have no access to ground power facilities, APU will operate to provide power and air conditioning on board. Both APU, GPU and ACU cause air pollution and emit quite a lot of emissions.

In fact, at Vietnam's airports, most boarding takes longer than expected, it usually takes more than an hour while the expected time is only. Furthermore, the ramp is also quite inconvenient, APU is used as the ground power source for the aircraft. Therefore, it takes about two hours for an APU to operate on the ground, which requires much fuel supply and generates significant emissions.

B. APU emissions

Unlike aircraft main engines, APUs are not certified for emissions, and manufacturers generally do not provide information on APU of emission rates. As a result, there is little publicly available data to serve as a basis for calculating APU emissions.

There are two approaches to APU emissions:

- The data regarding the APU emission factor are obtained from the emission dispersal models at the airport, the EDMS (Emissions and Dispersion Modeling System (U.S. FAA) and LASPORT (Lagrangian Simulation of particle transPORT (Europe)) [9].
- APU emissions come from the fuel burned in a gas turbine engine. Combustion products from engines include greenhouse gases and other pollutants. Fuel use and emissions will depend on fuel type, aircraft type, and APU engine type.

This paper uses both approaches based on public data to calculate APU emissions.

III. CASE STUDY

A. APU emissions based on uptime

With an operating time of 120 minutes for APU on each flight, in Vietnam, at Tan Son Nhat International Airport (TIA), there are about 300 domestic flights and 50 international flights a day.

Emissions = 2 Emissions Factor 350

Table I presents the amount of APU emissions per day in Tan Son Nhat International Airport (Vietnam).

TABLE I. THE AMOUNT OF APU EMISSIONS PER DAY IN TIA

	HC	CO	NO_x
EDMS APU (kg)	56	924	644
LASPORT APU (kg)	63	1022	917

B. APU emissions based on each type of APU

Emission=∑Fuel Time EF

EF: APU emission factor per amount of fuel (kg/kg of fuel)

With a run time of 120 minutes, the APU emissions are calculated as:

Emission = Fuel 120 60 EF

Table II proves the amount of APU emissions for each type of APU.

TABLE II. THE AMOUNT OF APU EMISSIONS FOR EACH TYPE OF APU

	HC (g)	CO (g)	NO_x (g)
B777	97.2	918.54	5545.26
B747	194.7456	6016.248	3707.1216
A330	94.5576	764.6832	4070.088
A320	30.024	410.328	2021.616

C. Comparison of APU emissions and GPU emissions

Table III indicates comparison of APU emissions and GPU emissions.

TABLE III. COMPARISON OF APU EMISSIONS AND GPU EMISSIONS

	HC	CO	NO_x
EDMS APU (kg)	56	924	644
LASPORT APU (kg)	63	1022	917
GPU (kg)	70	259	469

Based on Table III, although GPU emits less emissions, transporting GPU to aircraft requires specialized vehicles and is still not the best solution.

D. Comparison of APU emissions and bus emissions

Assuming two hours, a 41-60-seat bus running at 50km/h will consume about 13.02 kg of fuel. Table IV compares emissions generated by APU and bus operations over two hours for the purpose of quantifying APU emissions.

Table IV shows comparison of APU emissions and bus emissions.

TABLE IV. COMPARISON OF APU EMISSIONS AND BUS EMISSIONS

	HC (kg)	CO (kg)	NO_x (kg)
APU A320	0.03	0.4103	2.0216
APU B747	0.1947	6.0162	3.7071
Bus	0.02	0.11	0.73

The purpose of the above comparison is only to quantify emissions. Based on the comparison table in Table IV, the emissions of APUs on airplanes are much larger than those of bus engines over the same operating period of two hours.

IV. CONCLUSION

A. Feasibility

The case study shows that an APU releases a large amount of emissions into the environment. Therefore, hydrogen fuel cell is an alternative green energy to replace APU. Nevertheless, it also has certain limitations which need to be considered.

1) *Potential:*

- Hydrogen fuel cell will not pollute the environment, since the only waste generated is purified water, which can be reused as a source of clean water on the aircraft.
- Hydrogen fuel cell operates quietly. It generates no noise, which is especially preferred for an aircraft while operating on the ground [9].

- Hydrogen fuel cells achieve high conversion efficiency (90% efficiency can be achieved if both heat and power are used), with great stability and low emissivity [2].
- Hydrogen fuel cell is a device that produces electricity almost continuously, without interruption. Fuel can be fed and used immediately from the tank, thus refueling time is shortened.
- Other fuels such as jet fuel can also indirectly feed the fuel cell as long as it is broken down into hydrogen before entering the hydrogen fuel cell. This process can produce emissions, but to a lesser extent than the emissions of current engines [9].

2) *Challenge:*

- The biggest problem is the supply of hydrogen fuel: storage, transportation, and distribution are limited.
- The hydrogen fuel tank is large, so it takes up a lot of space on the aircraft. To ensure safety, hydrogen and oxygen are stored in two separate compartments, without direct contact, even in the event of a strong impact. Therefore, the fuel tank must be designed to withstand good, large volume and located near the fuel cell system.
- The cost of hydrogen fuel cells is relatively high due to the use of expensive materials and fabrication technologies.

B. The possibility of the hydrogen fuel cell replacing APU

The APU usage is part of the aircraft emissions. Thus, for environmental considerations, it is important to improve efficiency by reducing fuel burn and emissions. Hydrogen fuel cells provide a good solution for both economical and environmental improvements.

When using APU hydrogen fuel cells, the aircraft's structure will have to change significantly to accommodate the large hydrogen fuel tanks. New aircraft designs are required and allowed ideas like blended wing body (BWB), also known as blended body or hybrid wing body (HWB), which is a fixed-wing aircraft having no clear dividing line between the wings and the main body of the craft. The main advantage of the BWB is to reduce wetted area and the accompanying form drag associated with a conventional wing-body junction. However, the downside is the time involved in certifying new aircraft, along with the significant costs of structural redesign and reconstruction of fuel distribution infrastructure.

With all the environmental benefits that hydrogen fuel cells offer, along with the development of fuel cell technology and the growth of the energy industry, hydrogen is the key to the environmental solution. The application of hydrogen fuel cells in the aviation context is possible and definitely brings many positives. Replacing the gas turbine APU with the hydrogen fuel cell APU is one of the first steps in the hydrogen energy plan as AIRBUS and BOEING develop hybrid-hydrogen commercial aircraft in the next few years.

REFERENCES

[1] A. A. Omkar Yarguddi, "Fuel cell technology: A review. International Journal of Innovative Research in Science, Engineering and Technology.," 2014, pp. 3(7): p. 14668 - 14673.

[2] H. S. Works, "How Fuel Cells Work," 10 August 2015. [Online]. Available: http://auto.howstuffworks.com/fuel-efficiency/alternative-fuels/fuel-cell.htm. [Accessed 07 07 2022].

[3] Trung tâm Thông tin Khoa học và Công nghệ TP. HCM và PGS.TS. Nguyễn Mạnh Tuấn, "NGHIÊN CỨU CHẾ TẠO PIN NHIÊN LIỆU – TRIỂN VỌNG XU HƯỚNG NHIÊN LIỆU SẠCH VÀ XANH," SỞ KHOA HỌC VÀ CÔNG NGHỆ TP-HCM TRUNG TÂM THÔNG TIN KHOA HỌC VÀ CÔNG NGHỆ, TP. Hồ Chí Minh, 12/2011.

[4] E. T. S. Inc, Fuel cell handbook (7th edition), 2016.

[5] Fuel cell handbook (7thedition), EG&G Technical Services Inc., 2016.

[6] N. T. L. Hiền, Pin nhiên liệu - Nguồn năng lượng tương lai, Tạp chí Dầu Khí, 2019.

[7] D. N. Mavris, "ResearchGate," 1 2007. [Online]. Available: https://www.researchgate.net/figure/Fuel-cell-system-architecture-for-the-demonstrator-fuel-cell-airplane_fig1_267222294. [Accessed 2021 12 08].

[8] ZEROe, Airbus, 2020.

[9] A. C. N. D. ENVISA: Sandrine Carlier, G. M. QinetiQ: Chris Eyers and E. P. M. F. Jelinek, GAES project: Potential Benefits of Fuel Cell Usage in the Aviation Context, EEC/SEE/2006/004.

[10] EVN Tập Đoàn Điện Lực Việt Nam, 06/24/2014.

[11] Fuel cell, Wikipedia.

Optimization Integrated Rule-Based Operational Algorithms for an Agricultural Microgrid

Mohammad Hossein Mokhtare
Department of Electrical – Electronics Engineering
Middle East Technical University (METU)
Ankara, Turkey
mohammad.mokhtare@metu.edu.tr

Ozan Keysan
Department of Electrical – Electronics Engineering
Middle East Technical University (METU)
Ankara, Turkey
keysan@metu.edu.tr

***Abstract*—This paper presents operational algorithms for an agricultural microgrid. The proposed agricultural microgrid includes photovoltaic (PV) panels, a battery energy storage system (BESS), an electric water pump, and a water reservoir (WR). Water is stored in an elevated reservoir and used for drip irrigation. Four operational algorithms are proposed for the aforementioned agricultural microgrid. Two new, compact, easy to interpret and modify, rule-based algorithms are introduced based on priorities given to either the electric pump or the BESS. Improved versions of these two algorithms lead to optimization integrated operational algorithms that have the advantages of simplicity and solution optimality together. Simulation analyses in the MATLAB platform are performed to validate the performances of the proposed algorithms.**

Keywords—microgrid, agriculture, optimization, solar, water, EMS

I. INTRODUCTION

Agriculture is without a doubt one of the most fundamental components of the food supply chain. The provision of sustainable energy for this sector would have a considerable positive impact on the food security issue. Due to the decrease in costs of renewable energy sources especially solar energy, attention toward microgrids has increased. Applying renewable energy-based microgrids to farms would decrease their dependability on the main utility grid, provide a clean energy source, and expand farmlands to remote areas.

The operational algorithm of a microgrid which contains the energy management system (EMS) has a significant role in the design and control of these power systems. Two different types of such algorithms are often observed in the literature namely rule-based and optimized algorithms. In [1-10], optimization of microgrid sizing and operation for residential or commercial purposes has been performed using rule-based EMSs. Optimized versions of operational algorithms for similar applications have been adopted in other papers where the power flow of the microgrid along with existing constraints are formulated into a single optimization problem [11, 12].

Fewer studies have explored agricultural microgrids. In [13] a water pumping microgrid has been analyzed using a rule-based EMS. However, water is not considered a separate demand. The electrical load and water demand of a greenhouse supplied by a microgrid have been studied in [14] where the operation of the microgrid is optimized using a comprehensive EMS-based model predictive controller. Optimization of irrigation times and water volume in a hybrid microgrid with pumped hydro storage has been presented in [15].

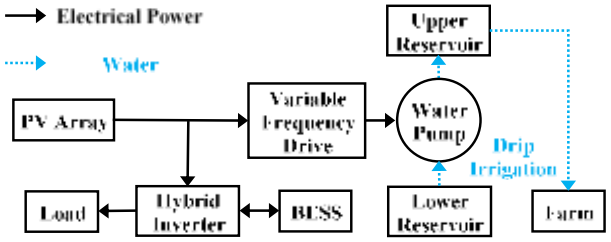


Fig.1. Structure of the considered agricultural microgrid

A grid-connected agricultural microgrid with wind turbines and pumped hydro storage has been defined in [16]. An optimized operational algorithm is utilized to minimize the operation cost of the microgrid in a day ahead market. A similar agricultural microgrid but with different methods has been studied extensively in [17]

In a nutshell, rule-based operational algorithms have the advantages of simplicity, low computational burden, and easier implementation [3]. However, their solutions are not optimized. Furthermore, as can be observed in [8, 13], for more complicated microgrid structures containing more than one energy storage and/or different operation strategies, the simplicity of rule-based algorithms is compromised and any further modifications would require considerable effort. On the other hand, optimized operational algorithms provide more efficient scheduling for a microgrid but their success is dependent on the correct definition of the problem and proper selection of the optimization tool and its parameters. Also, as seen in [14-17], the complexity of the optimization problem increases for systems such as agricultural microgrids with more than a single energy storage element and/or two types of demands because of the new constraints added to the problem. It is clear that rule-based algorithms with true simplicity and ease of implementation that can be readily modified are necessary for agricultural microgrids but have not been addressed in the literature.

This work has been funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK).

The agricultural microgrid structure studied in this paper is shown in Fig. 1 which is constructed of a photovoltaic (PV) array, battery energy storage system (BESS), electric water pump, and water reservoir (WR). PV array is the only source of energy in this islanded microgrid. It is responsible for supplying household appliances in the farmhouse and the excess energy is used to either pump water and/or charge the BESS. WR stores the energy in the form of potential energy which can then be used for drip irrigation of the field without any energy requirement. Based on the priority given to either pump or BESS, two algorithms can be developed in parallel.

This paper proposes four operational algorithms for the given agricultural microgrid. The first two algorithms are fast and straightforward rule-based algorithms that can act as the base for further studies. The other two algorithms are updates on the first two but they benefit from the advantages of both rule-based and optimized algorithms. These are optimization integrated rule-based algorithms and they provide optimal solutions with simple and easy implementations.

The remainder of the paper is organized as follows. Section II provides analytical models of the components in the agricultural microgrid. Proposed operational algorithms are given in section III. The simulation results of a case study are presented in section IV. Section V concludes the paper.

II. ANALYTICAL MODELING

A. Solar PV Array

The output power of a solar PV array can be calculated using the number of modules, solar irradiance, temperature, and datasheet information of the PV module. The approach presented in [18] provides accurate results regarding the maximum power point in any given situation. Therefore, this method is adopted in this paper, as follows:

$$V_{oc,c,stc} = V_{oc,m,stc} / N_c \quad (1)$$

$$I_{sc,c,stc} = I_{sc,m,stc} \quad (2)$$

$$P_{max,c,stc} = P_{max,m,stc} / N_c \quad (3)$$

$V_{oc,c,stc}$, $I_{sc,c,stc}$, and $P_{max,c,stc}$ are open circuit voltage, short circuit current, and maximum power of a single PV cell, respectively. They are calculated from the similar parameters of $V_{oc,m,stc}$, $I_{sc,m,stc}$, and $P_{max,m,stc}$ given in the datasheet for a single PV module. N_c is the number of cells in a module. Under solar irradiances and temperatures other than standard test condition (stc), parameters are calculated as:

$$V_{oc,c} = V_{oc,c,stc} + (K_v \times (T_c - 25) / 100) \quad (4)$$

$$I_{sc,c} = (I_{sc,c,stc} + (K_i \times (T_c - 25) / 100)) \times (G_T / 1000) \quad (5)$$

$V_{oc,c}$ and $I_{sc,c}$ are the open circuit voltage and short circuit current under the given operating conditions. T_c is the ambient temperature in °C. G_T is the solar irradiance in W/m². K_v and K_i are the temperature coefficients of open circuit voltage and short circuit current found in the PV datasheet.

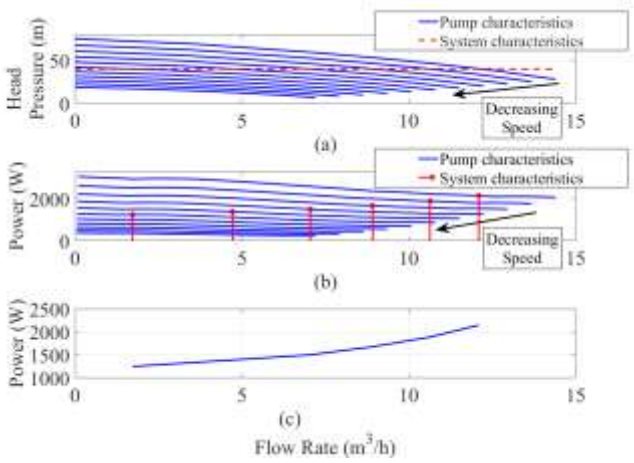


Fig.2. Characteristic curves of Alarko 4SDM12 / 12 submersible pump, (a) Head pressure vs. flow rate, (b) Power vs. flow rate, (c) Power vs. flow rate

Finally, the output power of the solar array is found from:

$$P_{PV} = V_{oc,c} \times I_{sc,c} \times FF \times N_c \times N_m \quad (6)$$

FF is the fill factor and N_m is the number of modules in the array.

B. Electric Water Pump

Three key parameters define the behavior of an electric water pump. These parameters include head pressure H in m, flow rate Q in m³/h, and efficiency η. These are usually presented in the forms of two curves namely head pressure versus flow rate (H-Q curve), and efficiency vs. flow rate (η-Q curve). Some manufacturers may provide additional curves regarding the input pump power (P_{Pump} (t)) and various operating speeds which will result in more accurate calculations. However, most often the two abovementioned curves are the only given information. Hence, the pump model considered in this paper assumes this scenario with minimum information provided.

In the agricultural microgrid structure, head pressure is constant and equal to the height of the elevated water reservoir. For this purpose, it is required to extract the power vs. flow rate (P_{Pump}-Q) curve of the pump under a constant head. This process is shown for a locally available pump manufactured by the Alarko company. The selected water pump is a 4-inch, 12-step, single-phase submersible pump with 2.2 kW rated power. First, the provided H-Q and η-Q curves for the rated speed of 2850 rpm are numerically identified and put through curve fittings. Corresponding H-Q and P_{Pump}-Q curves for different speeds can be obtained using (7) and the affinity laws (8)-(10):

$$P_{Pump} = (\rho Q g H / 3600) / (\eta / 100) \quad (7)$$

$$Q_1 / Q_2 = n_1 / n_2 \quad (8)$$

$$H_1 / H_2 = (n_1 / n_2)^2 \quad (9)$$

$$P_{Pump,1} / P_{Pump,2} = (n_1 / n_2)^3 \quad (10)$$

where ρ is the water density equal to 1000 kg/m³, g is the Earth gravity equal to 9.81 m/s². In addition, n is the operating speed of the pump. Results for the selected pump are shown in Figures 2(a) and 2(b). All that remains is to intersect the H-Q curves with the constant head of the system as seen in Fig. 2(a). Interception points are extracted and their corresponding powers are obtained as illustrated in Fig. 2(b). Finally, the P_{Pump}-Q curve of the selected pump driven under variable frequency drive for the given constant head is extracted and presented in Fig. 2(c). Hence, the flow rate of the pump under the available pumping power can be achieved as in Fig. 2(c). Moreover, the minimum amount of power that the pump can work with to provide a reasonable flow rate to the designated head is obtained as 1250 W. Water level in the water reservoir (W) can be calculated using the equation below:

$$W(t) = W(t-1) + (Q \times \Delta t) - (W_{Dem}(t)) \quad (11)$$

where Δt is the time step and W_{Dem} is the water demand in m³. A modified version of this equation is used in the proposed operational algorithms.

C. Battery Energy Storage System

The battery state of charge (SOC) can be obtained from:

$$SOC(t) = SOC(t-1) + (P_{BESS}(t) \times \Delta t \times 100 / C_{BESS}) \quad (12)$$

where $P_{BESS}(t)$ is the power of the battery which can be either positive or negative depending on the direction of power flow. C_{BESS} represents the capacity of the BESS in Wh.

III. PROPOSED OPERATIONAL ALGORITHMS

Operational algorithms are discussed in this section which can be divided into two groups. Fundamental rule-based algorithms are described first. Then, optimization integrated rule-based algorithms are presented.

A. Rule-Based Algorithms

In the agricultural microgrid structure of Fig. 1, the power balance equation in the system can be written as:

$$P_{PV}(t) = P_{Pump}(t) + P_{BESS}(t) + P_{Load}(t) + P_{Dump}(t) \quad (13)$$

where P_{Load} (t) is the load power, and P_{Dump} (t) is the power that could neither be used nor stored so it has to be dumped by shifting the solar array from working at its maximum power point. Equation (13) leads to defining the effective power:

$$P_{Net}(t) = P_{PV}(t) - P_{Load}(t) \quad (14)$$

P_{Net} is the power left for the pump and the BESS. It is important to understand that P_{Net} can be either positive, zero, or even negative. Positive value translates to extra power that can be used to either pump water and/or charge the BESS given that the water reservoir and/or batteries have free space. Zero means the PV power is exactly equal to the power demand. Negative values show that PV power alone is not sufficient to supply the loads and discharging of the BESS is required if batteries are not empty. Hence, P_{Net} is an appropriate input for the operational algorithms.

Two algorithms can be programmed based on the given priorities. Algorithms with priorities given to pump and BESS are presented in algorithms 1 and 2, respectively. In these algorithms, upper and lower limits on the SOC of batteries are represented by SOC^U and SOC^L, respectively. Similarly, W^U and W^L show upper and lower boundaries of the water reservoir, respectively. $P_{Pump\text{-}rated}$ and $P_{Pump\text{-}min}$ are the rated and minimum pump power, respectively. Furthermore, quantities of deficiencies in supplying power and water demands are represented by P_{Def} and W_{Def}, respectively. The efficiency of the agricultural microgrid in supplying the demands can be analyzed by using two reliability measures namely loss of power supply probability (LPSP) and loss of water supply probability (LWSP) as follows:

$$LPSP = \Sigma P_{Def} / \Sigma P_{Load} \quad (15)$$

$$LWSP = \Sigma W_{Def} / \Sigma W_{Dem} \quad (16)$$

LPSP and LWSP are parameters between 0 and 1. Considering the same level of importance for both electrical load and water needs, loss of supply probability (LSP) can be defined as the mean of the two abovementioned indices:

$$LSP = (LPSP + LWSP) / 2 \quad (17)$$

Assurance of keeping the water level in the reservoir within the given boundaries is integrated into the algorithms by introducing the K_W factor. In contrast, two separate factors are required to maintain the batteries SOC within limits. These include the K_{SOC1} factor to keep the SOC below SOC^U and K_{SOC2} for holding the SOC above SOC^L. Utilization of these parameters enables the operational algorithms to be fast, easy to implement, free of any loop functions, and robust against mistakes. Furthermore, calculation of supply deficiencies that are essential for the analysis of the agricultural microgrid operation, are included in the algorithms.

B. Optimization Integrated Rule-Based Algorithms

Commonly used optimized algorithms formulate the power balance equation and constraints of existing components into the optimization problem. Furthermore, typically allocated power of some of the components and/or states of the energy storage units are taken as decision variables. These facts complicate and hinder the energy management procedure and make the application of further improvements a difficult task.

On the contrary, the power flow balance of the microgrid and relevant constraints are naturally programmed into rule-based algorithms. This feature can be utilized to construct optimization integrated algorithms that are fast, accurate, and easy to implement. These algorithms are developed by making small modifications to the previous algorithms.

In the rule-based algorithms 1 and 2, the priority of power allocation has been given to pump and BESS, respectively. This absolute priority could adversely affect the optimality of

Algorithm 1: Pump First Operational Algorithm

input $P_{Net}(t)$;
$K_{SOC1}(t) = (SOC^U - SOC(t-1)) / (\Delta t \times 100)$;
$K_{SOC2}(t) = (SOC^L - SOC(t-1)) / (\Delta t \times 100)$;
if $P_{Net}(t) \geq 0$
  if $P_{Net}(t) \geq P_{Pump\text{-}min}$
    Calculate Q with min $(P_{Pump\text{-}rated}, P_{Net}(t))$;
    $K_W(t) = (W^U - W(t-1) + W_{Dem}(t)) / (Q \times \Delta t)$;
    $P_{Pump}(t) = \min(1, K_W) \times \min(P_{Pump\text{-}rated}, P_{Net}(t))$;
  else
    $Q = 0$;
    $K_W(t) = 0$;
    $P_{Pump}(t) = 0$;
  end
  $P_{BESS}(t) = \min(P_{Net}(t) - P_{Pump}(t), K_{SOC1}(t) \times C_{BESS})$;
else
$Q = 0$;
$K_W(t) = 0$;
$P_{Pump}(t) = 0$;
$P_{BESS}(t) = \max(P_{Net}(t), K_{SOC2}(t) \times C_{BESS})$;
end
$P_{Dump}(t) = \max(P_{Net}(t) - P_{Pump}(t) - P_{BESS}(t), 0)$;
$P_{Def}(t) = \max(0, (K_{SOC2} \times C_{BESS}) - P_{Net}(t))$;
$W(t) = W(t-1) + (Q \times \Delta t \times \min(1, K_W(t))) - (W_{Dem}(t))$;
$W_{Def}(t) = \max(0, W^L - W(t))$;
Calculate $SOC(t)$ from (12);

TABLE I. AGRICULTURAL MICROGRID COMPONENTS

CSUN 255-60 M Solar PV Array					
$V_{oc,m,stc}$	37.4 V	$I_{sc,m,stc}$	8.85 A	$P_{max,m,stc}$	254.9 W
N_c	60	K_v	-0.34 % / °C	K_i	0.05 % / °C
Alarko 4SDM12 / 12 Submersible Water Pump					
$P_{Pump\text{-}rated}$	2200 W	$P_{Pump\text{-}min}$	1250 W	Rated Speed	2850 rpm
Water Reservoir					
Capacity	50 m³	W^U	50 m³	W^L	10 m³
Battery Energy Storage System					
C_{BESS}	20 kWh	SOC^U	80%	SOC^L	20%

the solution. Hence, in each of these algorithms, after the power of the previously prioritized component is calculated, it is multiplied by a factor between 0 and 1. This creates an optimization problem with a single decision variable. So, the optimization integrated pump first rule-based algorithm can be constructed from algorithm 1 simply by updating the pump power after it has been obtained, as follows:

$$P'_{Pump}(t) = \alpha(t) \times P_{Pump}(t) \quad (18)$$

It should be mentioned that for $\alpha(t)<1$, the reduction in pump power is applied as a lesser time specified for the pump

Algorithm 2: BESS First Operational Algorithm

input $P_{Net}(t)$;
$K_{SOC1}(t) = (SOC^U - SOC(t-1)) / (\Delta t \times 100)$;
$K_{SOC2}(t) = (SOC^L - SOC(t-1)) / (\Delta t \times 100)$;
if $P_{Net}(t) \geq 0$
  $P_{BESS}(t) = \min(P_{Net}(t), K_{SOC1}(t) \times C_{BESS})$;
  if $P_{Net}(t) - P_{BESS}(t) \geq P_{Pump\text{-}min}$
    Calculate Q with min $(P_{Pump\text{-}rated}, P_{Net}(t) - P_{BESS}(t))$;
    $K_W(t) = (W^U - W(t-1) + W_{Dem}(t)) / (Q \times \Delta t)$;
    $P_{Pump}(t) = \min(1, K_W) \times$
      $\min(P_{Pump\text{-}rated}, P_{Net}(t) - P_{BESS}(t))$;
  else
    $Q = 0$;
    $K_W(t) = 0$;
    $P_{Pump}(t) = 0$;
  end
else
$Q = 0$;
$K_W(t) = 0$;
$P_{Pump}(t) = 0$;
$P_{BESS}(t) = \max(P_{Net}(t), K_{SOC2}(t) \times C_{BESS})$;
end
$P_{Dump}(t) = \max(P_{Net}(t) - P_{Pump}(t) - P_{BESS}(t), 0)$;
$P_{Def}(t) = \max(0, (K_{SOC2} \times C_{BESS}) - P_{Net}(t))$;
$W(t) = W(t-1) + (Q \times \Delta t \times \min(1, K_W(t))) - (W_{Dem}(t))$;
$W_{Def}(t) = \max(0, W^L - W(t))$;
Calculate $SOC(t)$ from (12);

TABLE II. PERFORMANCE INDICES OF SIMULATED CASE STUDY

Algorithm & Window \ Index		LPSP	LWSP	LSP
Pump First Rule Based	–	0.0908	0.0066	0.0487
Optimization Integrated Pump First Rule Based	1 Day	0.0758	0.0155	0.0456
	3 Days	0.0728	0.0144	0.0436
	7 Days	0.0724	0.0066	0.0395
BESS First Rule Based	–	0.0537	0.1166	0.0851
Optimization Integrated BESS First Rule Based	1 Day	0.1215	0.0301	0.0758
	3 Days	0.0807	0.0221	0.0514
	7 Days	0.0771	0.0097	0.0434

operation in the one-hour period. So, the previously calculated Q corresponding to $P_{Pump}(t)$ is still correct but needs to be adjusted for the reduced time duration as below:

$$Q'(t) = \alpha(t) \times Q(t) \quad (19)$$

For the BESS first version, a small update in the battery power is sufficient as shown below:

$$P'_{BESS}(t) = \beta(t) \times P_{BESS}(t) \quad (20)$$

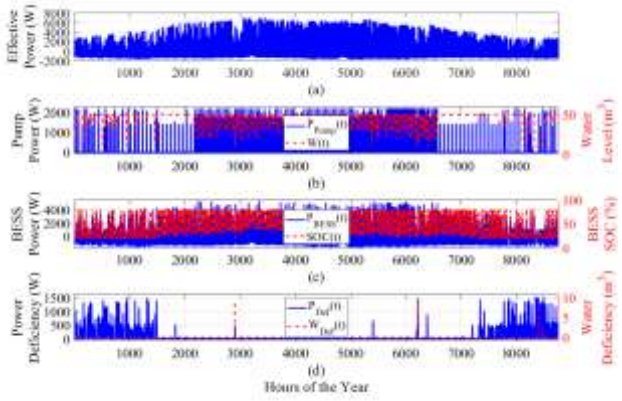


Fig. 3. Simulation results for the pump first rule-based operational algorithm

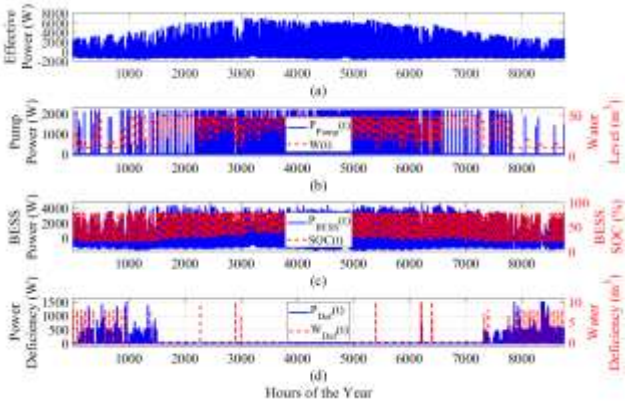


Fig. 4. Simulation results for the BESS first rule-based operational algorithm

In this paper, genetic algorithm is used to find α and β arrays with minimum *LSP*. Depending on the available predictions, different optimization windows can be adopted to calculate elements of these arrays in groups.

IV. SIMULATION RESULTS

The proposed algorithms have been simulated in the MATLAB platform and run for a case of a potential agricultural microgrid near the city of Ankara, Turkey. Environmental data for the previous year (2021) have been collected using the virtual meteorological station created in the FieldClimate software. Also, the hourly load profile of an average household in the area during the same year has been obtained from Energy Exchange Istanbul (EXIST) transparency platform. The selected water pump has already been analyzed in section II. B. Characteristics of other components in the microgrid are listed in Table 1.

Water demand in the form of a pre-planned irrigation pattern is defined as follows. From October to March, 40 m³ of water is sent to the field between 10:00 and 15:00 once every three days. In April, May, and June, irrigation is done once every two days where 30 m³ of water is released from 10:00 to 13:00 and another 30 m³ is supplied from 17:00 to 20:00. In July, August, and September, irrigation is done once

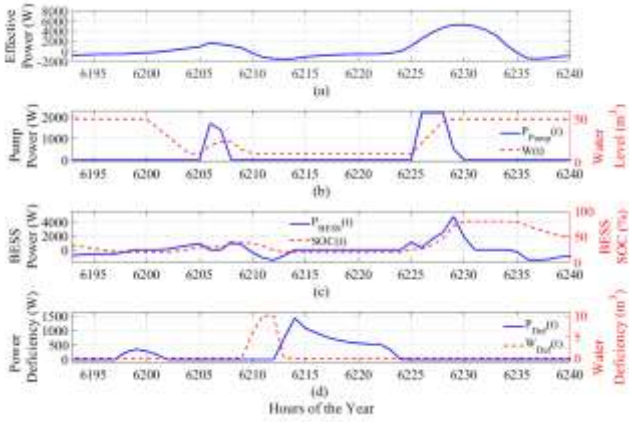


Fig. 5. Zoomed in results of 259th and 260th days of the year for the pump first rule-based operational algorithm

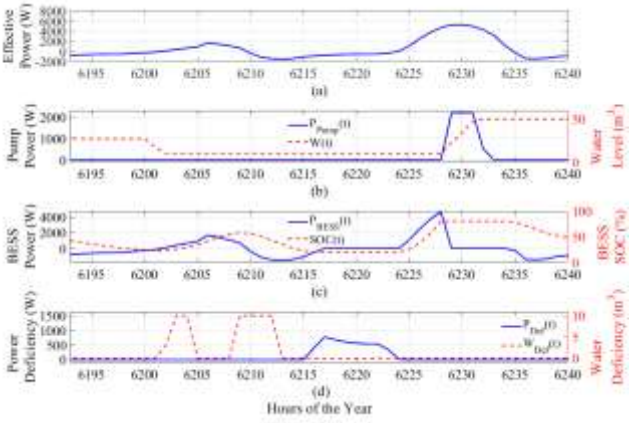


Fig. 6. Zoomed in results of 259th and 260th days of the year for the BESS first operational algorithm

every two days but with 40 m³ of water from 8:00 to 12:00 and another 40 m³ from 16:00 to 20:00.

Simulation results of operational algorithms 1 and 2 for the entire year with hourly steps are presented in Figures 3 and 4, respectively. It can be seen from Fig. 3(a) (or Fig 4 (a)) that P_{Net} is much lower in value during cold months when PV-generated power has decreased considerably. This causes deficiencies to be concentrated around this period as shown in Figures 3(d) and 4(d). Nonetheless, in both of the algorithms, W(t) in Figures 3(b) and 4(b), and SOC(t) shown in Figures 3(c) and 4(c) have stayed within their allowed limits. As expected, more power deficiency is present in the case of the pump first algorithm. On the other hand, the BESS first algorithm experiences more water shortage in comparison.

To present more details on the results of the proposed algorithms, zoomed-in waveforms for the 259th and 260th days are shown in Figures 5 and 6 for both algorithms. These days are chosen in a way to include both types of deficiencies. It can be observed that based on the available effective power in Fig. 5(a) (or Fig. 6(a)), the first algorithm has pumped water on both days as seen in Fig. 5(b) but the second algorithm has

directed the entire extra power in the first day towards charging of the BESS as shown in Fig. 6(b). Consequent different waveforms in the allocated powers for BESS in the two algorithms are presented in Figures 5(c) and 6(c). As a result, under the pump first algorithm, less water shortage has occurred but the power deficiency started earlier since batteries were not as full as they were in BESS first algorithm. Ultimately, they have experienced different waveforms and deficiencies (see Figures 5(d) and 6(d)). Next, optimization integrated algorithms are simulated with windows of 1, 3, and 7 days for the same year. The resulting performance indices are shown in Table 2. It can be observed that improvements can be made in the overall microgrid operation by using optimization integrated algorithms. Moreover, a comparison of results for different optimization windows shows the effect of optimization window and data forecast on this process. Results, numerically validate the expectations and previous analysis.

V. CONCLUSION

This paper presented two operational algorithms specifically designed for the defined agricultural microgrid which includes two types of energy storage namely water reservoir and BESS. These algorithms are highly accurate, easy to implement and they eliminate conventional tedious rule-based algorithms. Power and water demands have been treated within algorithms separately. This simplifies the path for more complicated studies on the application of similar microgrids in the agriculture sector.

Two variations of the operational algorithm have been presented namely the pump first and the BESS first algorithms. A sample microgrid has been simulated and run with both algorithms. The results validated the expected performance of the system. Moreover, optimization integrated versions of these algorithms have been proposed which provide optimal solutions to the scheduling of the agricultural microgrid while keeping the benefits of using rule-based management systems. Performance indices of modified algorithms are compared with basic ones which show improvements in the microgrid operation. Further studies include sizing optimization of the microgrid components, the addition of flexible irrigation patterns and demand response by having deferrable loads, and agricultural microgrids with multiple generation sources.

REFERENCES

[1] M. Alramlawi and P. Li, "Design Optimization of a Residential PV-Battery Microgrid With a Detailed Battery Lifetime Estimation Model," in IEEE Transactions on Industry Applications, vol. 56, no. 2, pp. 2020-2030, March-April 2020, DOI: 10.1109/TIA.2020.2965894.

[2] A. Fathy, K. Kaaniche and T. M. Alanazi, "Recent Approach Based Social Spider Optimizer for Optimal Sizing of Hybrid PV/Wind/Battery/Diesel Integrated Microgrid in Aljouf Region," in IEEE Access, vol. 8, pp. 57630-57645, 2020, DOI: 10.1109/ACCESS.2020.2982805.

[3] S. Bandyopadhyay, G. R. C. Mouli, Z. Qin, L. R. Elizondo, and P. Bauer, "Techno-Economical Model Based Optimal Sizing of PV-Battery Systems for Microgrids," in IEEE Transactions on Sustainable Energy, vol. 11, no. 3, pp. 1657-1668, July 2020, DOI: 10.1109/TSTE.2019.2936129.

[4] A. A. Z. Diab, H. M. Sultan, I. S. Mohamed, O. N. Kuznetsov, and T. D. Do, "Application of Different Optimization Algorithms for Optimal Sizing of PV/Wind/Diesel/Battery Storage Stand-Alone Hybrid Microgrid," in IEEE Access, vol. 7, pp. 119223-119245, 2019, DOI: 10.1109/ACCESS.2019.2936656.

[5] M. Kharrich et al., "Developed Approach Based on Equilibrium Optimizer for Optimal Design of Hybrid PV/Wind/Diesel/Battery Microgrid in Dakhla, Morocco," in IEEE Access, vol. 9, pp. 13655-13670, 2021, DOI: 10.1109/ACCESS.2021.3051573.

[6] H. M. Sultan, O. N. Kuznetsov, A. S. Menesy and S. Kamel, "Optimal Configuration of a Grid-Connected Hybrid PV/Wind/Hydro-Pumped Storage Power System Based on a Novel Optimization Algorithm," 2020 International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), 2020, pp. 1-7, DOI: 10.1109/REEPE49198.2020.9059189.

[7] M. Das, M. A. K. Singh, and A. Biswas, "Techno-economic optimization of an off-grid hybrid renewable energy system using metaheuristic optimization approaches - Case of a radio transmitter station in India," in ELSEVIER Energy Conversion and Management, vol. 185, pp. 339-352, 2019, DOI: 10.1016/j.enconman.2019.01.107.

[8] R. Shi, W. Wang, Z. Yuan, X. Fan and E. Ramezani, "A novel optimum arrangement for a hybrid renewable energy system using developed student psychology based optimizer – A case study," in ELSEVIER Energy Reports, vol. 7, pp. 70-80, 2021, DOI: 10.1016/j.egyr.2020.11.168.

[9] N. S. Attemene, K. S. Agbli, S. Fofana and D. Hissel, "Optimal sizing of a wind, fuel cell, electrolyzer, battery and supercapacitor system for off-grid applications," in ELSEVIER Int. Journal of Hydrogen Energy, vol. 45, no. 8, pp. 5512-5525, 2020, DOI: 10.1016/j.ijhydene.2019.05.212.

[10] C. Xu, Y. Ke, Y. Li, H. Chu, and Y. Wu, "Data-driven configuration optimization of an off-grid wind/PV/hydrogen system based on modified NSGA-II and CRITIC-TOPSIS," in ELSEVIER Energy Conversion and Management, vol. 215, pp. 112892, 2020, DOI: 10.1016/j.enconman.2020.112892.

[11] A. Ghasemi, "Coordination of pumped-storage unit and irrigation system with intermittent wind generation for intelligent energy management of an agricultural microgrid," in ELSEVIER Energy, vol. 142, pp. 1-13, 2018, DOI: 10.1016/j.energy.2017.09.146.

[12] M. Y. Zhang, J. J. Chen, Z. J. Yang, K. Peng, Y. L. Zhao and X. H. Zhang, "Stochastic day-ahead scheduling of irrigation system integrated agricultural microgrid with pumped storage and uncertain wind power," in ELSEVIER Energy, vol. 237, pp. 121638, 2021, DOI: 10.1016/j.energy.2021.121638.

[13] A. Ouammi, Y. Achour, H. Dagdougui, and D. Zejli, "Optimal operation scheduling for a smart greenhouse integrated microgrid," in ELSEVIER Energy for Sustainable Development, vol. 58, pp. 129-137, 2020, DOI: 10.1016/j.esd.2020.08.001.

[14] N. Mousavi, G. Kothapalli, D. Habibi, C. K. Das, and A. Baniasadi, "A novel photovoltaic-pumped hydro storage microgrid applicable to rural areas," in ELSEVIER Applied Energy, vol. 262, pp. 114284, 2020, DOI: 10.1016/j.apenergy.2019.114284.

[15] A. L. Bukar, C. W. Tan, K. Y. Lau, "Optimal sizing of an autonomous photovoltaic/wind/battery/diesel generator microgrid using grasshopper optimization algorithm," in ELSEVIER Solar Energy, vol. 188, pp. 685-696, 2019, DOI: 10.1016/j.solener.2019.06.050.

[16] A. L. Bukar, C. W. Tan, L. K. Yiew, R. Ayop, and W. Tan, "A rule-based energy management scheme for long-term optimal capacity planning of grid-independent microgrid optimized by multi-objective grasshopper optimization algorithm," in ELSEVIER Energy Conversion and Management, vol. 221, pp. 113161, 2020, DOI: 10.1016/j.enconman.2020.113161.

[17] Y. Li, S. Q. Mohammed, G. S. Nariman, N. Aljojo, A. Rezvani, and S. Dadfar, "Energy Management of Microgrid Considering Renewable Energy Sources and Electric Vehicles Using the Backtracking Search Optimization Algorithm," in Journal of Energy Resources Technology, vol. 142, no. 5, pp. 052103, 2020, DOI: 10.1115/1.4046098.

[18] M. Alramlawi, E. Mohagheghi, and P. Li, "Predictive active-reactive optimal power dispatch in PV-battery-diesel microgrid considering reactive power and battery lifetime costs," in ELSEVIER Solar Energy, vol. 193, pp. 529-544, 2019, DOI: 10.1016/j.solener.2019.09.034

Hands-on Detection for Steering Wheels with Neural Networks

Michael Hollmer
Faculty of Computer Science
Deggendorf Institute of Technology
Dieter-Görlitz-Platz 1
94469 Deggendorf
E-Mail: michael.hollmer@stud.th-deg.de

Andreas Fischer
Faculty of Computer Science
Deggendorf Institute of Technology
Dieter-Görlitz-Platz 1
94469 Deggendorf
E-Mail: andreas.fischer@th-deg.de (Corresponding author)

***Abstract*—In this paper the concept of a machine learning based hands-on detection algorithm is proposed. The hand detection is implemented on the hardware side using a capacitive method. A sensor mat in the steering wheel detects a change in capacity as soon as the driver's hands come closer. The evaluation and final decision about hands-on or hands-off situations is done using machine learning. In order to find a suitable machine learning model, different models are implemented and evaluated. Based on accuracy, memory consumption and computational effort the most promising one is selected and ported on a micro controller. The entire system is then evaluated in terms of reliability and response time.**

***Index Terms*—machine learning, hands-on detection, driving assistance**

I. INTRODUCTION

The development of advanced driver assistance systems is an essential goal for car manufacturers. As can be seen from a survey, driver assistance systems are by now an important purchase criterion for over 60% of potential buyers [1]. In addition, a unique selling point over the competition and thus a competitive advantage can be gained through the further automation of vehicles. An example is the system from Mercedes-Benz, which was the first to receive approval for autonomous driving at level 3 in December 2021.

Autonomous driving at level 3 enables the driver to divert his attention from what is happening on the road in certain situations. The vehicle takes over the lateral and longitudinal guidance and independently recognizes errors or departure from system limits. In such a case, the system would prompt the driver to take back control of the vehicle. This transfer of vehicle control is a crucial challenge. An autonomous system must be able to recognize whether the driver is ready to take over control of the vehicle again. To ensure this, some form of driver monitoring is required. One way of detecting the driver's condition is a hands-on detection (HOD). This is a system that detects whether the driver's hands are on the steering wheel and therefore control over the vehicle can safely be transferred. A HOD can be implemented inexpensively by measuring steering angle and torque acting on the steering wheel. The necessary sensors are required for the servo-assistance, anyway. However, there is the disadvantage that false hands-off messages often occur in situations where the driver does not exert any significant force for lateral guidance. In such a case, the driver would be asked to put his hands back on the steering wheel, even though he has not let go of the steering wheel.
A better HOD variant, also used in this paper, uses a capacitance sensor. This allows to detect the driver's contact with the steering wheel, without relying on any exerted force to the steering wheel. However, the evaluation of capacitance values is more complex, since these are dependent on the driver and his environment.

In this paper a machine learning algorithm is implemented, which is able to distinguish between a hands-on and a handsoff situation based on the capacitance values. The AI model is then ported to a micro controller and the reliability and response time of the HOD is evaluated. A maximum response time of 200ms is assumed to be appropriate for timely HOD. This paper aims to answer the question: Can neural networks increase reliability of HOD within a response time of 200ms?

II. BACKGROUND

Two techniques are combined in this paper to realize HOD: Capacity measurement and machine learning.

A. Capacity measurement

One option to realize HOD is detection of a contact between the driver and the steering wheel by measuring the change in capacitance. There are different methods to measure the capacitance of the steering wheel. In this paper, a frequencybased measurement method is used. Touching the steering wheel is detected by a change in capacitance in a sensor element, with the capacitance being calculated indirectly from the measured frequency. The sensor element represents a measuring capacitor which forms a resonant circuit together with another capacitor and a coil. The frequency of the resonant circuit can be calculated using equation 1, which describes an ideal resonant circuit.

$$f_0 = \frac{1}{2\pi\sqrt{L(C_k + C_s)}} \tag{1}$$

The equation depends on the capacitance of the capacitor C_k, the capacitance of the sensor element C_s and the inductance of the coil L. As long as the steering wheel is in an untouched state, the resonant circuit oscillates with its maximum frequency f_0. If the driver puts his hands on the steering wheel, the capacity of the sensor element is increased, leading to a reduction in the frequency of the resonant circuit. The sensor element is a capacitive mat that is wrapped around the core of the steering wheel and represents the active part of the measuring capacitor. Since there is no opposite side, a stray electric field forms between the active capacitor side and the environment. An approaching object causes a change of the capacitance value of the sensor element.

To illustrate, the measuring capacitor can be seen as a plate capacitor, which can be described by the equation $C = \varepsilon_0 \cdot \varepsilon_r \cdot \frac{A}{d}$. Here, both electrically conductive and nonconductive objects cause a change in capacitance for different reasons. A nearby conductive object causes the distance d between the active capacitor side and its surroundings to decrease, increasing the capacitance. On the other hand, nonconductive objects lead to an increase in capacity via a change in relative permittivity $_r$.

B. Machine learning approaches

To classify the capacitance values four different machine learning models are trained. In the following a brief overview of the different approaches is given:

1) Time Delay Neural Network: One machine learning approach is the Time Delay Neural Network (TDNN). The TDNN is structured as a standard multiperceptron with a delay buffer connected in front. New values in a time series are buffered until a certain amount is reached. Subsequently, these buffered values are passed in a final input into the multiperceptron, which then carries out the classification [2].

2) Long Short Term Memory: As a second approach to classify the capacitance values a Long Short Term Memory (LSTM) net, a variant of a recurrent neural network is used. In difference to feed-forward networks like the TDNN, the neurons of the LSTM can have connections to neurons in the previous layer, to the same layer or to themselves, in addition to the standard forward-pointing connections. The feedback loops implement a memory, which allows the network to remember previous events [2]. This is an advantage in timedependent series of measurements, since each measured value is dependent on its predecessor in a certain way. In contrast to TDNN, which assumes independent measured values, recurrent neural networks can use this memory to take account of the temporal dependency [3].

3) Random Forest: The last approach is the random forest which combines the prediction results of multiple decision trees using the bootstrap aggregating (bagging) method. The idea behind bagging is to train several decision trees with a subset of the training data. The subsets are created by randomly selecting samples from the entire training data. This process is also called bootstrapping [4]. The result are multiple decision trees that are structured differently and ideally even out in their classification errors. The output of the random forest is the class chosen by most of the decision trees.

III. RELATED WORK

Other work also used machine learning to develop HOD, differing in sensors, algorithms, and response times. Johansson and Linder [5] used a camera system and the torque acting on the steering wheel to implement HOD. For the camera, two CNN approaches were compared in classifying the most recently acquired image. For evaluating the torque measurement a one-dimensional convolutional neural network and an LSTM network were used. According to the authors, the evaluation of the torque requires a few seconds to detect a hands-off situation and up to two seconds to detect a handson. The camera approach reacted to a situation change within 5.4 seconds. Both solutions are thus well above the response time of 200ms we aim for in this work.

Hoang Ngan Le et al. [6] have also developed a machine learning based HOD with a camera system. In their paper the image evaluation is performed by a Region Based Convolutional Neural Network (RCNN) which has been improved for the specific purpose. The improved RCNN achieved 0.09 frames per second, which roughly corresponds to the evaluation of one frame every eleven seconds. As such, the time required for detection is also well above the 200ms limit.
A solution not based on machine learning was published as a patent by Volkswagen AG. This connects two possible approaches for HOD. On the one hand, the values of the steering angle or torque sensor are used and on the other hand, the capacitance values of the steering wheel are considered to distinguish between a hands-on and hands-off situation. The idea behind the combined approach is to use the torque sensor to detect hands-on situations with high confidence. During these situations, the corresponding capacitance values are recorded. With the data a function is set up with which it is possible to quickly decide for each new capacitance value whether it corresponds to a hands-on or hands-off situation [7].

Another non machine learning option for evaluating capacitance sensors was published by Analog Devices [8] and relies on dynamic threshold values. An algorithm continuously monitors the values of the capacitance sensor and measures the ambient level if no touch is detected. In addition, the average maximum sensor value is measured with each touch. The threshold from which a capacitance increase is counted as a touch is a certain percentage of the average measured maximum sensor value.

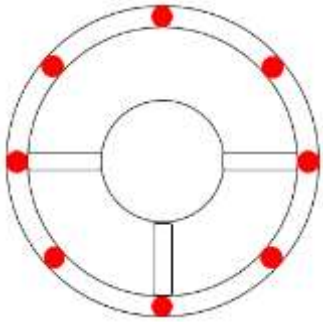


Fig. 1. Contact points on the steering wheel during the training phase

These approaches bear a potential problem: If the driver only touches the steering wheel very lightly, the measured average maximum sensor value decreases. The dynamic threshold adapts to the small capacitance values. Thus, at some point a

slight increase in capacitance values is erroneously recognized as a touch. If the driver brings both hands close to the steering wheel without touching it, this could trigger a similar increase in capacity as previous two-finger touch. The HOD would then recognize a hands-on situation even though the driver is not touching the steering wheel.

IV. EXPERIMENTAL DESCRIPTION

The implementation of the different machine learning models is divided into several steps. First, training data is recorded, classified and processed, which is then used to train the machine learning models. Based on the model results, the most promising model is selected and transferred to a micro controller. Finally, the system is evaluated in terms of reliability and response time.

A. Generating training data

For generating the training data, the steering wheel is alternately touched and released at defined points for five seconds. This process is repeated 30 minutes each for a twofinger, four-finger and two-hand touch. Figure 1 shows the points of contact. It should be noted that the points in the figure do not only refer to the front of the steering wheel. Alternately also the outside, the back and the inside were touched. Regarding the sampling rate a new capacitance value was recorded every 2 ms.

B. Preprocessing data for learning

After recording the training data two preprocessing steps were implemented. In the first step, every sample was assigned a "hands-on" or "hands-off' label. This was automated by following the change in capacitance when touching or releasing the steering wheel. The corresponding edge was used to separate and label all samples. Once the difference between two measured capacity values is above noise level, it is interpreted as an edge. The required change in capacitance to trigger an edge was set separately and fine tuned for each of the three data sets. A rising edge triggers the "hands-on" label, while a falling edge triggers the "hands-off" label.

In the second preprocessing step, every sample in the dataset was normalized in its length. This was done because the machine learning model should learn to classify a hands-on or hands-off situation based on capacitance values of just a few hundred milliseconds to speed up the reaction time of the HOD. Therefore a window with a fixed length of 100 values is placed over every sample. The values in the window form the input for the machine learning models. In each step, this window is moved one value, dropping an old value and adding a new value. Thus, all models have a fixed length input of 100 capacitance values, corresponding to 200 ms of recorded time.

C. Preparation of gradient data

In order for the machine learning models to deliver optimal results, capacitance values have to be normalized. For this, it is necessary to obtain minimum and maximum capacitance value during execution. In the training phase this is not a problem because all data is known a-priori. In a real world application this is not the case, which means that minimum and maximum values have to be determined dynamically. An estimate of the minimum value can be obtained by measuring the ambient level when the steering wheel is untouched. The maximum value, however, is a greater challenge. It would require the driver to place both of his hands on the steering wheel, which the system can never be sure is the case. Additionally, estimating the maximum value from the minimum value is not possible, as the change in capacitance caused by a driver heavily depends on his body weight. To eliminate this issue, the absolute capacitance values were converted into gradient values, focussing on change in capacity over time instead. This makes it easier to normalize the values, since only the maximum capacitive rate of change need to be known. Figure 2 shows gradient values when the steering wheel is touched with one hand.

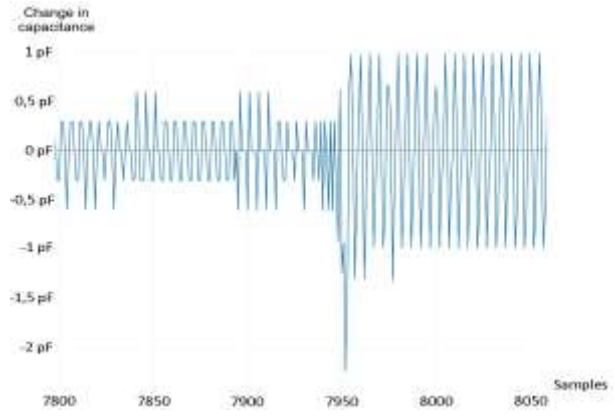


Fig. 2. Change in capacity when the hand is approached.

V. EVALUATION

For the evaluation all machine learning approaches are trained with the created datasets. The most promising model is then selected and ported to the STM32F769 micro controller where the final reliability and reaction time testing is done.

1) *A. Training*

In order to decide which machine learning model is best suited for the classification task, all models are trained with five different combinations of parameters. The resulting machine learning models are examined based on memory consumption, execution time and reliability.

1) Without gradient data: First, the models are trained with absolute capacitance values. All models achieve a very high level of accuracy as well as precision and recall, with differences visible mainly in memory usage and execution time (cf. Tab. I). With no major difference in accuracy, the random forest requires far more memory than other models, which is particularly disadvantageous for embedded systems. Therefore, measuring the execution time was neglected.

Looking at the neural networks, the biggest difference is the execution time. While the TDNN only need a few microseconds for a forward pass, the LSTM networks need several milliseconds due to their more complex structure. This is relevant, because data is sampled with a rate of 2ms in the experiments. Thus, when used on the micro controller, LSTM networks with more than one hidden neuron lose data because the measurement is faster than the processing.

2) With gradient data: Looking at the models trained with the gradient data, the previously observed disadvantages regarding the memory consumption of the random forest remain (cf. Tab. II). The processing time of the LSTM is still inferior to that of the TDNN. However, the LSTM with a hidden neuron performs slightly better in accuracy, precision and recall compared to the TDNN with 50 hidden neurons and occupies

with 27 kB of memory almost just a third of the memory. The TDNN on the other hand, offers a significantly shorter execution time. The delay between input and output for the TDNN with 50 hidden neurons is only 150 μs compared to the 0.6 ms of the smallest LSTM. Training the TDNN takes 1:08 minutes which is only a fraction of the training time for the LSTM, which takes 20:48 minutes. For this reason, the TDNN with 50 hidden neurons was selected and ported to the micro controller.

TABLE I. TRAINING DATA

Model	Hidden Neurons	Accuracy	Precision	Recall	F0,5-Score	Memory	Exec. time
TDNN	1	99,28%	98,78%	99,65%	98,95%	25kB	12µs
TDNN	5	99,62%	99,53%	99,64%	99,56%	31kB	21µs
TDNN	10	99,62%	99,53%	99,64%	99,56%	37kB	34µs
TDNN	20	99,62%	99,54%	99,63%	99,56%	49kB	59µs
TDNN	50	99,63%	99,56%	99,62%	99,57%	84kB	150µs
LSTM	1	99,07%	98,35%	99,64%	98,61%	27kB	0,6ms
LSTM	5	99,29%	98,87%	99,59%	99,01%	27kB	3ms
LSTM	10	99,57%	99,48%	99,58%	99,50%	31kB	7ms
LSTM	20	99,60%	99,48%	99,64%	99,51%	48kB	15ms
LSTM	50	99,60%	99,49%	99,63%	99,52%	150kB	59ms
Model	Estimators	Accuracy	Precision	Recall	F0,5-Score	Memory	Features
RF	1	99,39%	99,45%	99,43%	99,44%	146kB	10
RF	5	99,62%	99,66%	99,64%	99,65%	723kB	10
RF	10	99,63%	99,68%	99,64%	99,67%	1.459kB	10
RF	100	99,64%	99,70%	99,63%	99,69%	14.444kB	10
RF	5	99,62%	99,67%	99,64%	99,66%	668kB	15

2) *B. Practical reliability test:* To test if the system recognizes touches by the driver reliably, the steering wheel was touched with two fingers, four fingers one hand and two hands at the points shown in figure 1. In the runs in which the steering wheel was touched with two and four fingers, a distinction was made between touching the front, back, inside and outside for each point. In each of the four runs, all points were touched ten times to see whether touches were only recognized sporadically in some places.

In these experiments, the recognition of two fingers proved to be the most difficult. Especially on the inside, where there is a seam, the distance to the sensor mat is particularly large. This decreases sensitivity and a touch triggers only a small increase in capacitance, resulting in no touch detection at all for the two finger experiments and a maximum of 7 out of 10 correctly identified events in the four finger experiment.

Regarding the position, the 6 o'clock position proved to be difficult, both with two and with four fingers. Somewhat less (but still noticeably) impacted positions were the 3 o'clock and 9 o'clock positions. These three positions are located, where the steering wheel spokes connect to the wheel—likely the root causes of the problem.

In the 10 and 2 positions typical for driving a car, all events were recognized reliably irrespective of finger count, as long as any area apart from the inside of the wheel was touched. Also, when the steering wheel was not just touched, but gripped with one or two hands, the success rate rose to 100%.

3) *C. Reaction time*

Next, the reaction time of the system was tested with two fingers which represents the hardest challenge as shown in the previous section. Fig. 3 shows the capacitance values over time

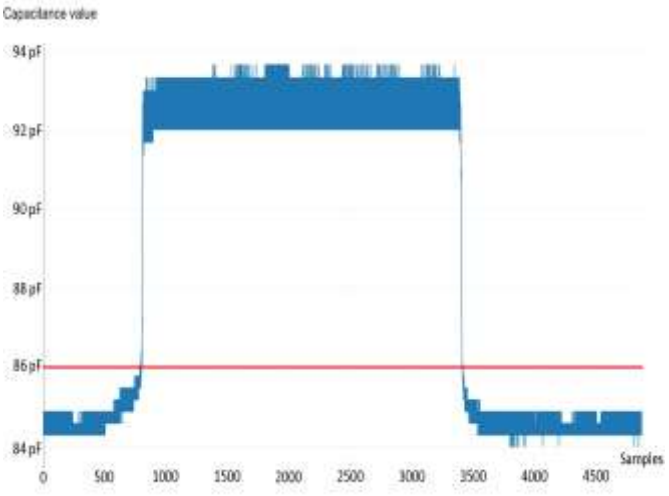


Fig. 3. Touching the steering wheel with two fingers for the reaction time measurement. The red line represents the threshold from which the time measurement is started and stopped.

when the steering wheel was touched with two fingers. The red line represents the threshold from which the steering wheel was actually touched and released. The increase in capacitance below the red line is caused by approaching the fingers but not having made contact yet. Reaction time measurement is started, when the capacitance values first exceed the threshold and stopped when the values drop below it. In ten experimental

TABLE II. RESULTS OF THE TDNN, LSTM AND THE RANDOM FOREST (RF) AFTER TRAINING WITH THE CALCULATED GRADIENT DATA

Model	Hidden Neurons	Accuracy	Precision	Recall	F0,5-Score	Memory	Exec. time
TDNN	1	93,19%	98,81%	92,69%	97,52%	25kB	12µs
TDNN	5	93,18%	99,50%	96,14%	98,81%	31kB	21µs
TDNN	10	93,68%	99,87%	97,10%	99,30%	37kB	34µs
TDNN	20	95,64%	99,84%	97,74%	99,41%	49kB	59µs
TDNN	50	99,08%	99,64%	98,63%	99,44%	84kB	150µs
LSTM	1	99,86%	99,79%	100%	99,83%	27kB	0,6ms
LSTM	5	99,84%	99,86%	99,80%	99,85%	27kB	3ms
LSTM	10	99,66%	99,80%	99,66%	99,78%	31kB	7ms
LSTM	20	98,23%	99,89%	99,72%	99,86%	48kB	15ms
LSTM	50	99,82%	99,88%	99,63%	99,83%	150kB	59ms
Modell	Estimators	Accuracy	Precision	Recall	F0,5-Score	Memory	Features
RF	1	95,51%	94,88%	94,59%	94,82%	165kB	10
RF	5	98,96%	98,41%	99,13%	98,55%	872kB	10
RF	5	99,00%	98,50%	99,34%	98,67%	760kB	15
RF	10	99,48%	99,29%	99,57%	99,35%	1.733kB	10
RF	100	99,79%	99,61%	99,89%	99,67%	17.240kB	10

VI. runs, results ranged between 108ms and 294ms for "hands-on" events and a significantly faster reaction time of 30–60ms for "hands-off" events, if two fingers were used. With four fingers, reaction times could be reduced to 74–94ms (hands-on) and 38–58ms (hands-off), respectively.

VI. CONCLUSION

The results show that it is possible to use a machine learning algorithm to evaluate capacitance values for HOD and achieve fast reaction times. By using the change in capacitance instead of the absolute values in the machine learning model, the problem of normalizing the input values was solved and the HOD worked without external calibration, independent of the driver and environment.

REFERENCES

[1] VuMA. (2020, November) Wichtigste kriterien beim autokauf in deutschland in den jahren 2017 bis 2020. [Online]. Available:https://de.statista.com/statistik/daten/studie/171605/umfrage/wichtige-kriterien-beim-autokauf/

[2] E. Alpaydin, *Maschinelles Lernen*. De Gruyter Oldenbourg, Mai 2019, publication Title: Maschinelles Lernen. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/9783110617894/html

[3] M.-P. Hosseini, S. Lu, K. Kamaraj, A. Slowikowski, and H. C. Venkatesh, "Deep learning architectures," in *Deep Learning: Concepts and Architectures*, ser. Studies in Computational Intelligence, W. Pedrycz and S.-M. Chen, Eds. Springer International Publishing, 2020, pp. 1–24. [Online]. Available: https://doi.org/10.1007/978-3-030-31756-0_1

[4] M. Kubat, *An Introduction to Machine Learning*. Springer International Publishing, 2021. [Online]. Available: https://link.springer.com/10.1007/978-3-030-81935-4

[5] E. Johansson and R. Linder, "System for hands-on steering wheel detection using machine learning," 2021. [Online]. Available: https://odr.chalmers.se/handle/20.500.12380/302723

[6] T. Hoang Ngan Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides, "Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[7] D. Musial, Marek and D. To, Thanh-Binh, "Hands-on-erkennung im kraftfahrzeug," patentde 102011109711, Februar, 2013.

[8] *AD7143 - Programmable Controller for Capacitance Touch Sensors*, Analog Devices, Januar 2007. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/AD7143.pdf

Management and Detection System for Medical Surgical Equipment

Alexandra Hadar
Faculty of Industrial Engineering and Technology Management
Holon Institute of Technology - HIT
Holon, Israel
msalexandrahadar@gmail.com

Natan Levy
School of Computer Science & Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel
Natan.levy1@mail.huji.ac.il

Michael Winokur
Faculty of Industrial Engineering and Technology Management
Holon Institute of Technology - HIT
Holon, Israel
Michaelw@hit.ac.il

***Abstract*—Retained surgical bodies (RSB) are any foreign bodies left inside the patient after a medical procedure. RSB is often caused by human mistakes or miscommunication between medical staff during the procedure. Infection, medical complications, and even death are possible consequences of RSB, and it is a significant risk for patients, hospitals, and surgical staff. In this paper, we describe the engineering process we have done to explore the design space, define a feasible solution, simulate, verify, and validate a state-of-the-art Cyber-Physical System that can significantly decrease the incidence of RSB and thus increase patients' survivability rate. This system might save patients' suffering and lives and reduce medical staff negligence lawsuits while improving the hospital's reputation. The paper illustrates each step of the process with examples and describes the chosen solution in detail.**

Keywords—surgical equipment, monitoring system, IoT and Sensors networks, Systems Engineering, RSB, FFB, RFID

I. INTRODUCTION

Today the complexity of the modern healthcare system is growing alongside the population growth in such a way that requires the provision of treatments to more patients with less staff. The workload on the medical staff in surgeries is increasing. Medical equipment is reconciled manually in most hospitals during surgeries. Retained Surgical Bodies (RSB) can cause infections, severe medical complications, and even death [1]. This event is also described in the literature as Forgotten Foreign Bodies (FFB) [2]. The incidence of this condition is between 0.3 and 1.0 per 1,000 abdominal operations. In the United States, for example, there are about 1,500 RSB cases per year [1]. According to a recent national survey in the United States, retained sharp instruments (needle, blade, guidewire, metal fragment) are more prevalent than reported in the current literature [3]. Standardizing reports and implementing new technologies is the most effective way to improve the management and prevention of these events [4].

Internet of Things (IoT) technology can provide solutions that allow medical staff to track medical equipment and prevent RSB. In this work, we describe the design of a Cyber-Physical System, centered around Radio Frequency Identification (RFID) based IoT system that will highly decrease the incidence of RSB. This technology can reduce the need for unnecessary secondary surgical procedures and increase patients' survivability rate.

There is evidence that technologies like RFID may help to reduce the occurrence of RSB [5]. The use of RFID sponge detection technology reduced the percentage of procedures in which a search for a sponge was performed, the number of unreconciled sponge counts, the amount of time spent searching for sponges and obtaining radiographs, and costs [6].

There is medical equipment marked with RFID and barcode, as well scanning and tracking systems with RFID, like Xarefy [7] and ORLocate [8], and barcode like Stemato [9].

In those systems, there is no automatic detection of the medical equipment at the room entrance. There is a need to bring the detector near to the medical equipment. Moreover, in the case of a barcode, there is a need for a line of sight. There is no registration and automatic alerts for all cases where medical equipment leaves or enters the operating room during surgery. Except for ORLocate, the above systems do not have a dedicated component for locating equipment in the patient's cavity.

This paper manuscript focuses on the design challenges we faced to achieve an optimal feasible solution. Further development is needed to implement a full prototype to perform controlled experiments to rate the success of the proposed design in the field. Section II presents an overarching view of the process, Section III describes in detail the design space exploration performed to reach the desired solution, and section IV describes the selected architecture including steps taken during conceptual design to augment the system's robustness. Section V describes the solution validation. The paper's concluding summary and directions for future work are presented in Section V.

II. WORK PROCESS

The authors started by looking at existing technologies. Next, the stakeholder disclosure process was done, and a collection of their needs and requirements, including detailed interviews with clinical staff, reviews of news media, and professional literature.

In all the interviews, the interviewees expressed their concerns about the current situation. The majority of the interviewees believe that a system is needed to solve those concerns. The main topics that were raised:

- The process of counting medical equipment in most hospitals is currently done manually - exposing the process to human error.
- It is necessary to reduce the load on the medical staff.
- The proposed solution must include a way of human intervention.
- Concerns have been raised about the level of disruption the system may cause to the medical staff (physical, visual, vocal), concerns about unnecessary system movement that could interfere with the proper course of surgery and damage the sterile field.
- Medical staff may have difficulty using a cumbersome solution.
- To avoid unnecessary interference, the solution should be operated "hands-free" via voice commands and gestures, rather than only by buttons.

In conclusion, there is a need to facilitate the work processes of the medical staff and help prevent human error. There is a need for a solution with minimal intervention in the existing process, without unnecessary movements that can impair the surgical process and the sterile field.

Next, we sorted the needs according to the KANO method [10] and rated them according to the Nominal Group Technique (NGT) method [11]. As a result, the six top fundamental requirements are: (1) documenting and reporting the presence, (2) counting, (3) monitoring and identifying the medical equipment, (4) identifying the room where the medical equipment is present, (5) monitoring the medical equipment in the operating room, and (6) locating medical equipment in the operating space.

The following stages included the definition of system use cases (Fig. 1 presents an example of a central use case of locating medical equipment within the patient's cavity once medical staff announce patient closing), system requirements, engineering characteristics of the possible solutions, and building a system model, including its dynamics (Fig. 2).

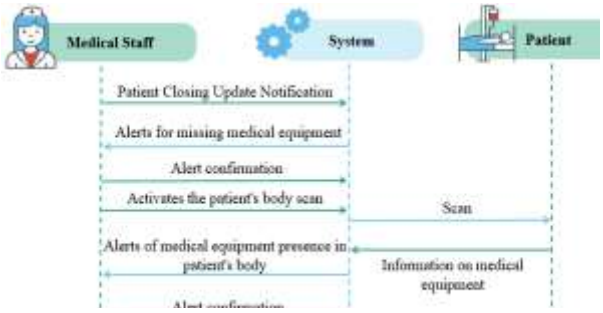


Fig. 1. Use Case - Locating medical equipment in the patient's cavity.

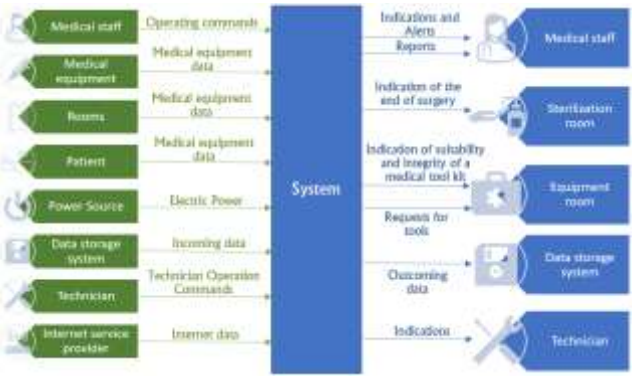


Fig. 2. System model with its dynamics.

Seven different solution concepts were proposed, to systematically select the most appropriate solution. We performed a comprehensive failure modes analysis for the selected solution and suggested new features to improve system robustness. Next, we defined the requirements for the subsystem components and established a detailed system Verification and Validation (V&V) process, including a detailed simulation of the system dynamics with MathWorks Simulink [12] to check the design validity and robustness.

III. DESIGN SPACE EXPLORATION AND SYSTEM SELECTION

The design space exploration performed started with the established principles of concept definition of possible solutions followed by the generation of design variants and their evaluation to reach a selected "best alternative" [13]. This exploration included a combination of different technologies to monitor the medical equipment and locate it in the operating space via, for example, RFID, Bluetooth, Ultra-Sound, and Cameras. Some solutions included robots to handle the equipment while other solutions included a mobile cart with portable detectors for this task.

A. Solution Variants

A morphological matrix as presented in Table I was applied to generate five solutions variants and perform a risk assessment for components in each solution. We proposed two more combinations of solutions using mix & match principles [13]. Concepts short description:

1) *Dr. Tool:* RFID-based equipment tracking system. Built-in RFID sensors into all medical equipment. Scattered sensors in the rooms' entrances (Room Sensors) enable monitoring of the location of the equipment in the room area. Mobile Tool Cart (MTC) RFID-based medical tool cart with monitor, computer and communication system, RFID tool tray, and RFID trash bin. Medical Equipment Detector (MED) RFID detector for manual scanning or mounted to surgical lights pivoting arm to locate misplaced equipment. Cloud-based Central Management System (CMS) for database management of the medical equipment, receiving an indication from the various sensors, and generating reports.

TABLE I. THE MORPHOLOGICAL MATRIX

Main Functions	**Concepts**						
	Dr. Tool	***Blue Tool***	***Ultra Tool***	***Robi Tool***	***BB Tool***	***Dr. Robi Tool Mix & Match***	***Dr. RoBBi Tool Mix & Match***
Monitoring of medical equipment	RFID-based system: RFID sensors built into medical equipment, RFID Room Sensors, MTC, and MED	Integrated Bluetooth and RFID system: Bluetooth sensors built into medical tools, RFID sensors built into consumable equipment, Bluetooth and RFID-based medical equipment vending machine, RFID Room Sensors, MTC, and MED	RFID-based system: RFID sensors built into medical equipment, RFID Room Sensors, MTC, and MED	RFID-based system with a robot: RFID sensors built into medical equipment, RFID Room Sensors, Robot with built-in RFID readers	Integrated system based on RFID and cameras: cameras with lighting and medical equipment detection algorithm, RFID sensors built into medical equipment, RFID Rooms Sensors, MTC, and MED with cameras	RFID-based integrated system with a robot: A system of RFID sensors built into medical equipment, RFID Rooms Sensors, Robot with built-in RFID readers	RFID-based integrated system with a robot and cameras: Camera system with lighting and medical equipment detection algorithm, A system of RFID sensors built into medical equipment, RFID Rooms Sensors, Robot with built-in RFID readers
Identifying equipment in the patient cavity	MED	Bluetooth for the medical tools, RFID sensors for the consumable medical equipment with MED	Dedicated ultrasound monitoring system in the patient's bed	Surgery bed with RFID readers	IR cameras system with medical equipment detection algorithm, thermography	MED	MED
Providing alerts and indications	Alerts and indication system with monitor and speaker on the MTC's local computer, sounds and indication lights on MED	Alerts and indication system on smartphone connected to the MTC's monitor	Alerts and indication system with monitor and speaker on the MTC's local computer, Ultrasound system with sound and indication lights	Alerts and indication system in the robot's tablet, sounds and indication lights on the surgery bed	Alerts and indication system on the MTC's tablet	Alerts and indication system in the robot's tablet, sounds and indication lights on MED	Alerts and indication system in the robot's tablet, sounds and indication lights on MED
Communicating with medical staff	Communication system[1] Installed on MTC's local computer	Communication system[1] Installed on smartphone connected to the MTC's monitor	Communication system[1] Installed on MTC's local computer	Communication system[1] Installed on robot	Communication system[1] Installed on MTC's tablet	Communication system[1] Installed on robot	Communication system[1] Installed on robot
Task management and generating reports	CMS with SQL DB + Python	CMS with SQL DB + Matlab	CMS with Cassandra DB + Python	CMS with Oracle DB + Python	CMS with Oracle DB + Matlab	CMS with SQL DB + Python	CMS with Oracle DB + Matlab
Saving data and history	Cloud	Cloud	External backup drive	Internal backup drive	Company servers	Cloud	Cloud

[1]. Communication system with command decoding - voice decoding software with microphone, speakers, modem, camera, and hand gesture decoding software

2) *Blue Tool:* Bluetooth and RFID-based integrated system. Bluetooth sensors built into tools and RFID built into consumable medical equipment, RFID Room Sensors. Medical equipment vending machine based on Bluetooth and RFID, issuing medical equipment according to a command from MTC or CMS, smartphone-based MTC, and MED for the consumable misplaced medical equipment, Cloud-based CMS.

3) *Ultra Tool:* Ultrasound and RFID-based system with built RFID sensors inside all the medical equipment. RFID Room Sensors, MTC, and MED for locating medical equipment in the operating room space, a dedicated ultrasound system for identifying equipment in the patient cavity, and CMS with an external backup drive.

4) *Robi Tool:* RFID-based system equipment tracking system with a robot. RFID sensors built into all medical equipment, RFID Room Sensors, a robot with a built-in tablet, communication system, RFID tool tray, RFID trash bin, and scattered sensors in the robot frame for locating equipment. The

robot can receive tools from the equipment room, move them to the operating room and deliver them to medical staff. A dedicated surgical bed equipped with RFID readers for identifying equipment in the patient cavity and CMS with an internal backup drive.

5) *BB Tool:* Integrated RFID, cameras with a lighting system, and Infra-Red (IR) cameras. Cameras and RFID Room Sensors enable monitoring of the location of tools in the room. Cameras enable proactive detection of missing tools in the room space. Tablet-based MTC and MED, in addition to the MED containing cameras for detection in a wider visible range in the operating room space. IR cameras for identifying equipment in the patient cavity using thermography, and CMS with storage on company servers.

6) *Dr. Robi Tool (Mix&Match):* Combination of Dr. Tool & Robi Tool. RFID Room Sensors, Robi Tool Robot, MED, and Cloud-based CMS.

7) *Dr. RoBBi Tool (Mix&Match):* Combination of Dr. Tool & Robi Tool & BB Tool. Cameras and RFID Room Sensors, Robi Tool Robot, MED, and Cloud-based CMS.

B. Evaluating Main Solution Variants

To find the best solution, we used the PUGH method [14]. In this section, we describe the process to find the three most suitable candidates using Engineering and Qualitative Characteristics and refined the choice to three possible solutions. The preferred solution is then described in Section IV below.

1) *Engineering characteristics*

The essential mechanism for solution selection is the engineering characteristics that should be common to all of them [13]. Table II presents the technical engineering characteristics that were selected for solutions comparison.

We calculated the relative importance of each engineering characteristic using the "Voice of the Customer" aspects of the Quality Function Deployment (QFD) method [15] by the quantitative analysis of the correlation between characteristics and stakeholders' needs. The top five highest scoring characteristics were used for an initial evaluation of all the solutions (Availability, Detection Range, Reliability – MTBF, Charging Time, Screen Size).

2) *Qualitative characteristics*

We apply a slight variation of classical QFD for solution ranking for the sake of clarity, using a two-dimensional diagram where the engineering characteristics are used for technical ranking on the Y axis, while on the X axis we display the ranking of "qualitative" characteristics as applied in [16]. It is shown in Fig 3. The qualitative characteristics for solutions comparison that were selected are Time to Market, Life Cycle Cost, End User, Dependence on Suppliers, Reliability, Availability, Maintainability, and Safety (RAMS).

TABLE II. TABLE II. ENGINEERING CHARACTERISTICS TABLE

Characteristic Name	**Characteristic Data**		
	Type	***Range***	***Target***
Height	Quantitative	(0.85 to 1.20) m	(0.85 to 1.15) m adjustable
Diameter	Quantitative	(0.60 to 0.75) m	(0.67 to 0.71) m adjustable
Weight Carrying	Quantitative	(15 to 18) kg	18 kg
Availability	Quantitative	(98 to 99) %	98 %
Maximum Humidity	Quantitative	(80 to 90) %	90 %
Minimum Humidity	Quantitative	(0 to 10) %	5 %
Maximum Temperature	Quantitative	(40 to 50) ℃	40 ℃
Minimum Temperature	Quantitative	(-5 to 1) ℃	1 ℃
Noise Level	Quantitative	(65 to 85) dB	80 dB
Charging Time	Quantitative	(9000 to 18000) s	9000 s
Screen Size (diagonal)	Quantitative	(0.5461 to 0.6858) m	0.6858 m
Detection Range	Quantitative	(0.2 to 1) m	0.9 m
Reliability - Mean time between failures (MTBF)	Quantitative	(4320000 to 5400000) s	5184000 s

3) *Top Three concepts*

After performing an initial evaluation table (screening) we found that the attempt to use robots to handle the medical equipment in cooperation with surgical personnel did not pass the initial evaluation stage in the PUGH method. Robots are a more complex alternative compared to the medical cart with scattered sensors and detectors with RFID and are still considered a less reliable option by skilled personnel. The top three alternatives are Dr. Tool, Ultra Tool, and Blue Tool.

4) *The best concept*

After conducting a comprehensive scoring evaluation for the top three variants in a relation to all the technical engineering characteristics, and the summary of the weighted scores for each of the alternatives, we found that Dr. Tool's alternative is the top concept. We performed a qualitative characteristics rating for the top three solutions. The results of the technical characteristics scoring and the qualitative characteristics comparison appear in Fig. 3.

From the comparison of solution alternatives, using both technical and qualitative ratings, Dr. Tool is the best concept.

IV. SELECTED CONCEPT

A. Solution Robustness

Before we can target a final state-of-the-art design concept, we applied techniques to check and enhance the robustness of the solution since this is a key to its success. For example, see [13, 17]. We used NASA's Risk Management Matrix [18] to identify weaknesses and potential failures in the selected option in order to reduce risks. Risk reduction is achieved thru redundant sensors and additional batteries, reinforcements to the tray, lowering the center of mass, and improving cyber security. Consequently, the robust Dr. Tool concept presented in Section B below got the highest score with PUGH reevaluation. Reevaluation of the results of the technical characteristics scoring and the qualitative characteristics comparison with the robust solution appears in Fig. 4.

B. Final Concept

The design of the final concept is presented schematically in Fig. 5. The architecture includes the following components:

1) *Built-in RFID sensors in medical equipment.*
2) *Room Sensors:*
   Rooms are equipped with an RFID-based tracking system to monitor medical equipment location within the equipment room, Sterile Processing Department (SPD), and operating room.

3) *Mobile Tool Cart (MTC):*
   The central part of the innovative design. Portable RFID-based medical cart with a monitor and computer. The MTC can communicate with medical staff using voice commands, gestures, and a touch screen. It is equipped with an RFID tool tray and an RFID trash bin for tracking medical equipment.

4) *Medical Equipment Detector (MED):*
   RFID detector to locate misplaced medical equipment in the patient cavity or operating room. The detector can be used for manual scanning or mounted to surgical lights pivoting arm.

5) *Central Management System (CMS):*
   Central Cyber component of the system. Cloud-based service system responsible for communicating with all room sensors and MTC, creating alerts to the medical staff, managing the medical equipment database, and generating reports.

This system can identify each medical equipment item unambiguously, including needles. The MED can be installed above the operating space. The MTC supports voice commands and hand gestures in addition to a touch screen. The strengths of the system are the ability to locate medical equipment automatically at the entrance to the rooms, and it will automatically notify the presence of the medical equipment during the operation. When the patient's cavity is about to close the system will notify the medical staff for RSB.

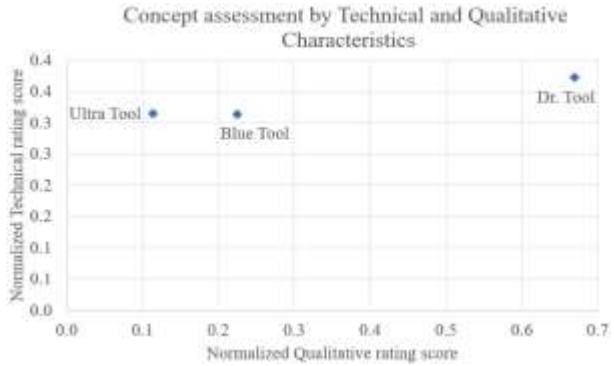


Fig. 3. Comparison of the technical and qualitative ratings of the solution concepts

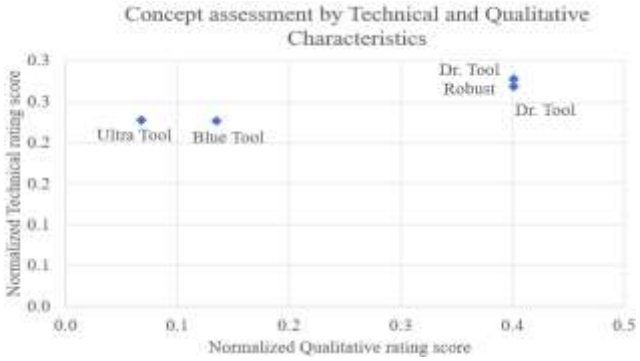


Fig. 4. Comparison of the technical and economic ratings of the solution concepts with robust Dr. Tool.

V. SYSTEM SIMULATION AND VERIFICATION

An early-stage Verification and Validation (V&V) process was applied to check the system design. We developed a model of the system and its dynamics, and a state-flow simulation was performed on the main processes using Simulink [3]. This includes monitoring the medical equipment in the operating room during surgery, including patient cavity, medical equipment room locator, medical staff reporting in case of a damaged item, and generating reports. Part of the simulation for monitoring medical equipment during surgery appears in Fig. 6.

During simulation, new inputs were added, such as acknowledgment by the SPD that they are ready to receive contaminated instruments at the end of the operation. The simulation highlighted the relevance of communication between the CMS and MTC in reporting the status of the medical equipment presence in the Operation Room (OR). In case new medical equipment is brought into the OR without placing it on the MTC, the CMS will notify the MTC about the presence of new equipment, and the MTC will add it to the monitoring list and alert the medical staff. All medical equipment is being monitored constantly during operation, even if the medical staff accidentally forgot a tool in his pocket and left the OR, CMS will notify the MTC, and MTC will remove this tool from the monitoring checklist for this surgery and will inform the medical staff. After performing the simulation multiple times, with

different sequences and inputs, we got a high level of trust that the design meets the stakeholders' requirements.

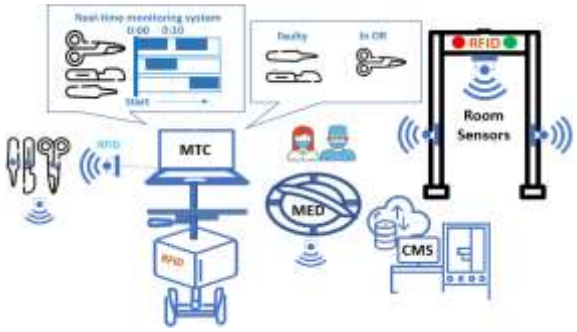


Fig. 5. System Design

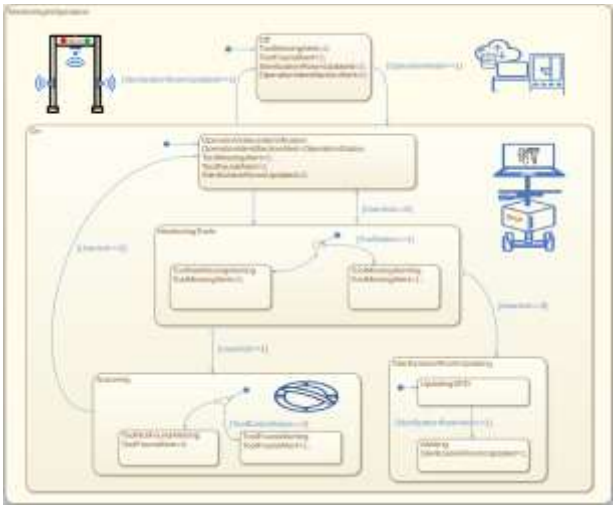


Fig. 6. Part of State-Flow simulation in Simulink

VI. CONCLUSIONS

Technology is evolving and advanced IoT systems today allow us to improve existing processes and increase automation. In our work, we have taken a step towards automation in the field of monitoring medical equipment - in a hospital that will save lives. However, the most suitable design we have chosen is not completely autonomous, as it still requires a final scanning performed by trained medical personnel to have full confidence of non-residual RSB. We argue that at this stage the best way to enter the market is with a less complex and more reliable system that the medical staff can incorporate into the standard procedure in the hospital. Future development envisages the construction of a working prototype and the performance of controlled experiments to gather data supporting its validity. Future versions of the system should incorporate robotic handling of the surgical equipment and AI. In a collaborative environment with medical personnel, this will further streamline and fully automate the process of monitoring the medical equipment to minimize the risk of RSB.

ACKNOWLEDGMENT

We would like to thank Mr. Yossi Hen for his originality, intense work investment, and partnership in the design phase of the project.

REFERENCES

[1] V. A. Zejnullahu, B. X. Bicaj, V. A. Zejnullahu, and A. R. Hamza, "Retained surgical foreign bodies after surgery," *Open Access Macedonian Journal of Medical Sciences*, vol. 5, no. 1, pp. 97–100, Feb. 2017, doi: https://doi.org/10.3889/oamjms.2017.005.

[2] F.R. Polat, E. Saka, and Y. Duran, "Forgotten Foreign Bodies (FFB) during Surgery and Malpractice," *EC Gastroenterology and Digestive System*, vol. 6, no. 3, 180-181, Mar. 2019.

[3] S. A. Weprin, D. Meyer, R. Li, U. Carbonara, F. Crocerossa, F. J. Kim, R. Autorino, J. E. Speich, and A. P. Klausner, "Incidence and or team awareness of 'near-miss' and retained Surgical Sharps: A national survey on United States Operating Rooms," *Patient Safety in Surgery*, vol. 15, no. 1, Apr. 2021, doi: https://doi.org/10.1186/s13037-021-00287-5.

[4] S. Weprin, F. Crocerossa, D. Meyer, K. Maddra, D. Valancy, R. Osardu, H. S. Kang, R. H. Moore, U. Carbonara, F. J. Kim, and R. Autorino, "Risk factors and preventive strategies for unintentionally retained Surgical Sharps: A systematic review," *Patient Safety in Surgery*, vol. 15, no. 1, Jul. 2021. doi: https://doi.org/10.1186/s13037-021-00297-3.

[5] T. Sebastian, M. Dhandapani, L. Gopichandran, and S. Dhandapani, "Retained surgical items: A review on Preventive Strategies," *Asian Journal of Nursing Education and Research*, vol. 10, no. 3, p. 375, 2020, doi: 10.5958/2349-2996.2020.00080.4.

[6] V. M. Steelman, A. G. Schaapveld, H. E. Storm, Y. Perkhounkova, and D. M. Shane, "The effect of radiofrequency technology on time spent searching for surgical sponges and associated costs," *AORN Journal*, vol. 109, no. 6, pp. 718–727, May. 2019, doi: https://doi.org/10.1002/aorn.12698.

[7] Xerafy. "Surgical instrument tracking systems for hospitals in SPD and the OR." Xerafy.com. https://www.xerafy.com/post/medical-device-and-surgical-instrument-tracking (accessed Jul. 26, 2022).

[8] Haldor. "ORLocate Solution." Haldor-Tech.com. https://www.haldor-tech.com/products/the-orlocate-solution (accessed Jul. 26, 2022).

[9] Besco. "SteMaTo®: Unique Software for your Sterile Processing Department (SPD)." Besco. https://besco.be/en/solutions/stemato (accessed Jul. 26, 2022).

[10] F.-H. Lin, S.-B. Tsai, Y.-C. Lee, C.-F. Hsiao, J. Zhou, J. Wang, and Z. Shang, "Empirical research on Kano's model and customer satisfaction," *PLOS ONE*, vol. 12, no. 9, 2017, doi: https://doi.org/10.1371/journal.pone.0183888.

[11] S. I. Harb, L. Tao, S. Peláez, J. Boruff, D. B. Rice, and I. Shrier, "Methodological options of the nominal group Technique for Survey Item Elicitation in Health Research: A scoping review," Journal of Clinical Epidemiology, vol. 139, pp. 140–148, Nov. 2021, doi: https://doi.org/10.1016/j.jclinepi.2021.08.008.

[12] M. Winokur and A. Zaguri, "Recent Advances in System Modelling and Simulation", presented at the MODPROD, Linköping, Sweden, Feb. 3-4, 2021, vol. 26, no. 15, p. 16.

[13] D. M. Buede and W. D. Miller, *The engineering design of systems: Models and methods*, 3rd ed. Hoboken, NJ: Wiley, 2016.

[14] S. Pugh, *Total design: Integrated methods for successful product engineering*, 1st ed. Boston, MA, USA: Addison-Wesley, 1991.

[15] A. R. Suhardi, "Quality Function Deployment to Improve Quality of Service" presented at the IBSM 2 Trisakti University Conference, Chiang Mai, Thailand, Oct. 2013.

[16] G. Pahl, W. Beitz, J. Feldhusen, and K. H. Grote, *Engineering design: a systematic approach*, 3rd ed. London, United Kingdom: Springer, 2007, pp. 119.

[17] B. Bergman, J. D. Mare, S. Loren, and T. Svensson, *Robust design methodology for reliability exploring the effects of variation and uncertainty*. Chichester, West Sussex, U.K.: Wiley, 2009.

[18] L. Gipson, "NASA Systems Engineering Award 2013 guidelines and rules." *NASA*.gov. https://www.nasa.gov/aeroresearch/resources/design-competitions/sae-guidelines (accessed Jul. 26, 2022).

A Robot That is Always Ready for Safe Physical Interactions

Huthaifa Ahmad
RIKEN Information R&D and Strategy Headquarters
RIKEN
Kyoto, Japan
huthaifa.ahmad@riken.jp

Yutaka Nakamura
RIKEN Information R&D and Strategy Headquarters
RIKEN
Kyoto, Japan
yutaka.nakamura@riken.jp

***Abstract*—For robots to operate in real environments where humans live, they should be able to interact safely and adequately with humans and their surroundings due to the inevitability of physical contact in such unstructured environments. However, although this is a crucial and essential skill, it is still an unsolved problem in the robotics field. Modeling all the events in a real situation is impossible, and the necessity to have explicit models of the robot's body and the environment does not suit continuously changing conditions. Therefore, robots should have flexible bodies capable of producing proper spontaneous reactions to unexpected stimuli instead of having active control of every joint angle of the robot's body at all times. In this regard, this paper introduces the development of a prototype robot that utilizes the mechanical structure of its body to realize adaptable passive dynamics. The implemented mechanisms allow safe interactions with the robot, whether it is passively or actively actuated. Here, we demonstrate the mechanical design of the robot, validate its performance through two experiments of physical human-robot interaction, and discuss its potential advantages for future research.**

Keywords—physical interaction, human-robot interaction, adaptive morphology, passive dynamics, mechanical design

I. INTRODUCTION

Humans live in unstructured environments where unpredictable scenarios are happening all the time. Physical contact in such environments is crucial and inevitable; it could be planned in advance to benefit from the surrounding objects, or even unplanned reactions to adapt to abrupt stimuli. It is also essential for communicating social cues and conveying information about the emotional state of a person [1], [2], [3]. Therefore, the ability of physical interaction, whether it's with humans or surroundings, whether it's planned in advance or not, is a very important skill for humans. Thus it is an important skill for robots too.

Developing a fully autonomous robot capable of physical interaction is a tough task and still an unsolved problem in the robotics field. Trying to model all the events in a real situation is nearly impossible, and relying on the active control approach solely won't be a practical solution for continuously changing environments. Yet, most of the work in this direction focuses on the model-based approach [4], [5], [6]. Although these robots demonstrated impressive motion behaviors, they are usually task-specific and operate in prescribed workspaces with distinct workflows. Therefore, they lack adaptability and avoid direct contact unless it was pre-modeled in advance [7].

In order to realize a versatile adaptation, and expand the robots' application domains from static environments with well-defined geometries to unstructured anthropic environments, robots must be augmented with adaptive bodies. Many studies addressed this issue by carefully considering the robot's embodiment [8], [9], [10]. Such robots exploit their natural dynamics by implementing adjustable impedance actuators [11], [12], soft materials [13], adaptive mechanisms [14], [15], and so on.

By following the embodiment-based approach, here in this paper, we propose an attempt to overcome this problem by giving extra attention to the robot's mechanical design and harnessing morphological computation-based mechanisms. The goal is to develop robust robots capable of exploiting their bodies to coexist with humans and acclimate to their surroundings, robots with adaptive passive dynamics always ready for safe, proper, and fast physical interactions. Even in the case of no power input supplied to the system, the robot should perform safe reactions in response to external influences. From this perspective, mechanisms such as gravity compensation (GC) and mass dampers (MD) will be used.

The remainder of this paper is organized as follows. Section II shows the followed approach to developing the robot. Sections III and IV present the conducted experiments, clarify the experimental setups, and explain the results. Section V discusses the results and addresses the robot's potential advantages for future applications. Section VI concludes the paper with general remarks.

II. THE ROBOT'S ADAPTIVE BODY

Figure 1 shows a concept sketch of a robot with adaptive passive dynamics. Its simple mechanical structure enables it to be always ready for physical interactions. For example, if someone poked the robot's head, the robot would comply with the applied force. And because of the torso's mass which behaves as a counterweight, the robot will then restore its upright posture. Similarly, if the robot was pushed sideways, the ball-shaped, backdrivable mobile base will prevent the robot from falling. This compliant behavior of the robot leads to a safe and more natural physical interaction experience.

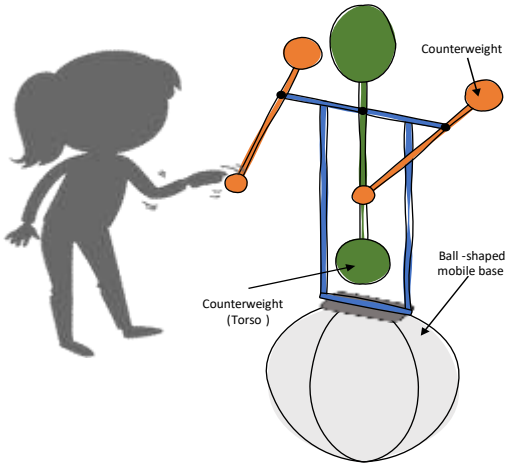


Fig. 1. Sketch of a simple robot with adaptive passive dynamics.

Following the same strategy, the robot Yajirobe (named after a traditional Japanese balance toy), shown in figure 2, was built as a prototype attempt toward developing robust robots that can share the same habitat with humans. It is 130 cm tall and weighs 12 kg, excluding the mobile base.

The mechanical structure of Yajiroboe is demonstrated in figure 3. The robot employs techniques such as GC and MD mechanisms in order to operate both passively and actively at the same time. The specifications of the robot's degrees of freedom (DoF) are presented in table I. It has eight DoF as illustrated in figure 4. The active actuation of the robot's joints is realized through electric actuators that are soundless and smoothly backdrivable [16]. More details about the robot's design are explained below.

A. *Upper body*

The robot needs to have an interactive body that is always ready for physical interactions. Whether the robot is executing active movements or being fully passive, once faces an external force it should comply with it and show safe and natural behaviors.

Fig. 2. The developed robot, Yajirobe.

TABLE I. SPECIFICATIONS OF THE ROBOT'S JOINTS

Body part	DoF	Range of motion (deg)	Actuator specification		Minimum torque for active operation (N.m)
			Speed (rpm)	Torque (N.m)	
Head	Pitch	60	720	0.1	0.02
	Yaw	180	720	0.1	0.02
Arm * 2	Shoulder joint	360	260	0.3	0.14
	Elbow joint	110	260	0.3	0.25
Upper body	Passive hinge	14	----		----
	Counterweight	50	260	0.3	0.3

Similar to the schematic diagram of figure 1, we designed the robot's body to oscillate around a passive hinge joint as illustrated in figure 5. The weighty body parts, such as the head's actuators, were placed below the rotational axis to form the counterweight instead of adding extra heavy components.

The body actuator is attached through a belt and hangs loosely as demonstrated in the figure. The role of this actuator is threefold. First, the motor mass itself represents a part of the counterweight. Second, responsible for the active movements of the upper body; the swinging motion of the motor in the sagittal plane changes the body posture as shown in figure 6. Third, behaves as a damper during the passive mode for faster restoration of the upright posture; this is equivalent to the tuned MD mechanism which is often used in buildings to reduce the vibrations caused by earthquakes [17].

B. *Arms*

Each of the robot's arms consists of 2 DoF; shoulder and elbow joints, as shown in figure 7. The elbow joint allows flexion and extension movements of the forearm in the sagittal plane with (110 deg) range of motion as shown in table I.

A rack and pinion gear system is used to rotate the elbow joint. While the rack is fixed as part of the arm's frame, the motor attached to the pinion gear will behave as both actuator and counterweight for the forearm. The motor's mass will counterbalance the movement of the forearm in order to keep the CoM stationary, and thus maintain equilibrium around the shoulder axis throughout the entire range of motion. This GC mechanism reduces the load on the actuators and ensures compliant behaviors.

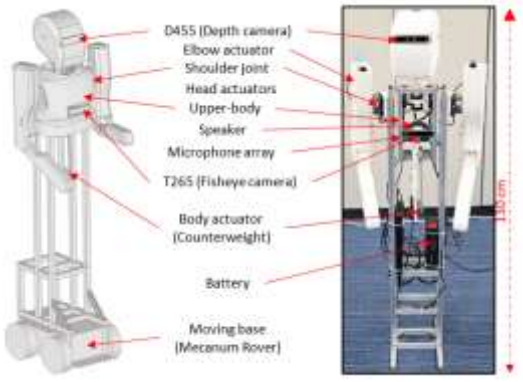


Fig. 3. The robot's mechanical structure and it's main components.

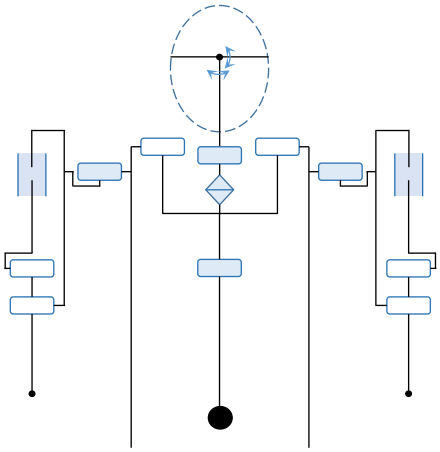


Fig. 4. Joints' structure and distribution. Shaded symbols are the actuators.

The arm mechanism can also be used to exert different force values at the end-effector and rotate the arm at different speeds while the same torque is applied at the shoulder. This can be realized by utilizing the arm's inertia. Adjusting the arm's posture will change the mass distribution around the shoulder axis, which, based on the conservation law of angular momentum, can be used to control the force and angular velocity. For example, if the arm is extended, the masses of the forearm and the elbow actuator will be located further from the shoulder rotational axis. This leads to higher inertia, the arm feels heavier, and the angular speed will be slower, and vice versa when the arm is flexed.

C. Head

The head's 2-DoF mechanism is illustrated in figure 5. It consists of two motor modules placed beneath the body's rotational axis. The first motor, placed at the bottom, is responsible for yaw movement, while the second motor, attached to the head through bevel gears, generates pitch motion. The symmetrical distribution of the head's mass around its rotational axes reduces the power requirements for the actuators. The range of motion of each DoF is demonstrated in table I.

In the following sections, two experiments were conducted to evaluate the robot's performance. The first experiment, in section III, sought to examine the body's mechanism in suppressing the vibration after encountering external stimuli; while the second experiment, in section IV, studies how the arm's mechanism can influence the participants' impressions during physical interaction.

III. EXPERIMENT 1: REACTION TO EXTERNAL FORCE

The robot's upper body was designed to comply with external forces to achieve adaptive behaviors. Once the stimuli disappear, the counterweight will cause the robot's body to oscillate back and forth until returns to its equilibrium position. To test how fast the robot returns to equilibrium (upright body posture), a simple form of touch (poking) was performed on the robot as shown in figure 8. During the experiment, reflective markers were placed on the robot, and a motion capture device (OptiTrack-V120: TRIO) was used to track the robot's movement.

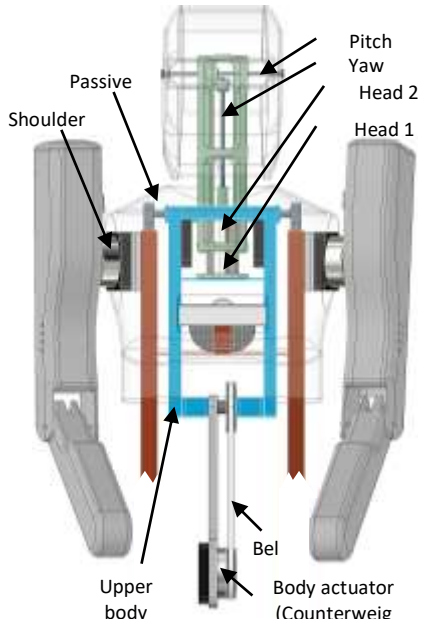


Fig. 5. Illustration of the body's mechanisms.

A. Experimental setups and procedures

For a better understanding of the results, the experiment was conducted under two conditions:

- Locked counterweight: the counterweight is mechanically fixed to oscillate in phase with the upper body, as shown in figure 9-b.
- Free counterweight: the counterweight is free to swing out of phase with respect to the upper body, as shown in figure 9-c.

The following procedures were implemented for each condition:

1. Poke (push) the robot's body to its mechanical limits, as shown in figures 8 and 9.
2. Release the robot.
3. Let the robot passively return to its equilibrium position without intervention.

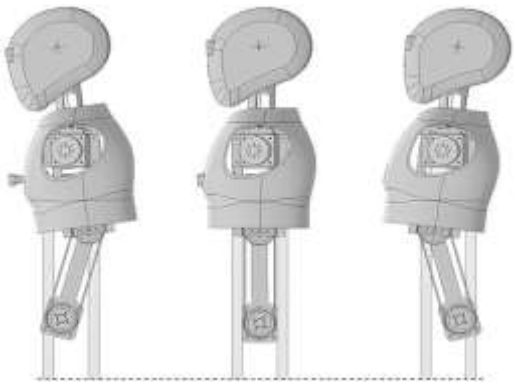


Fig. 6. Side view of the upper body's orientation during active actuation.

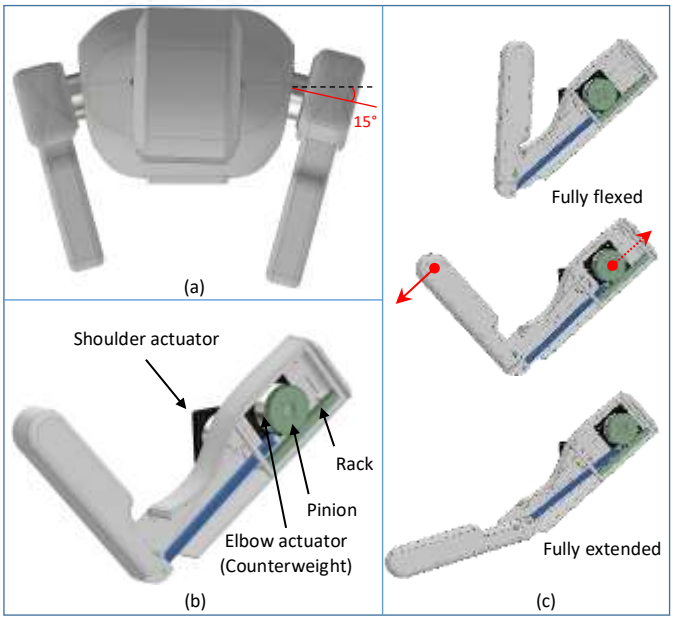


Fig. 7. The robot's arm. (a) Top view. (b) The GC mechanism of the arm. (c) Illustration of the arm movement.

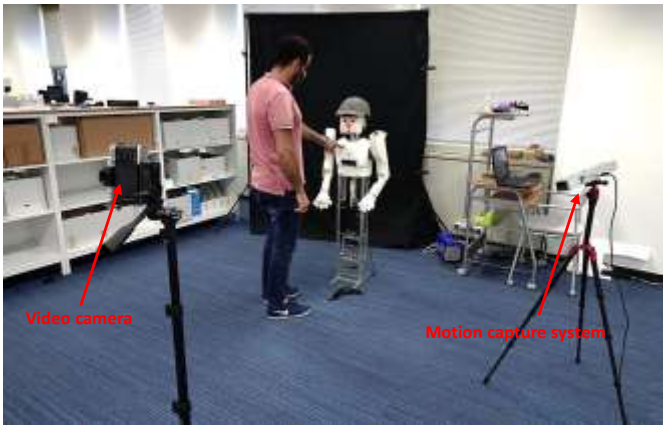


Fig. 8. Experimental settings of the poking experiment.

B. Results

The graphs in figure 10 compares the oscillatory behavior of the robot's body for the tested conditions. As can be deduced from the figure, under the first condition of locked counterweight, the natural damping of the robot's body caused the robot to eventually restore its equilibrium posture. However, the free counterweight under the second condition reduced the oscillation amplitude by absorbing the system's kinetic energy. As a result, the robot only needed around 20% of the time required in the first condition to return to its equilibrium position.

IV. EXPERIMENT 2: PHYSICAL HUMAN-ROBOT INTERACTION THROUGH ARM CONTACT

This experiment aims to investigate how changing the arm's inertia will affect the humans' impression of the robot during their physical interaction. For this purpose, participants were asked to perform a high five with the robot.

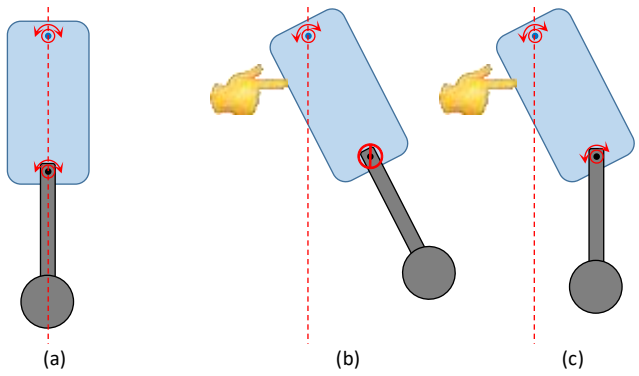


Fig. 9. Illustration diagram of the tested conditions showing the side view of the robot. (a) The robot's upright body posture at equilibrium. (b) Condition 1 of locked counterweight. (c) Condition 2 of free counterweight.

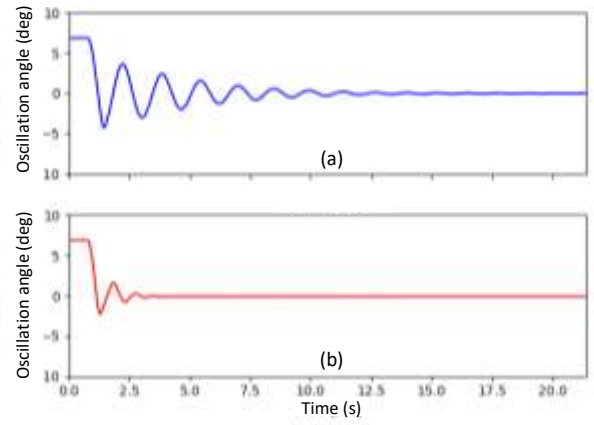


Fig. 10. The damping behavior of the robot's upper body under both conditions, (a) locked counterweight and (b) free counterweight.

A. Experimental setups and procedures

The high five in real life has various types and occurs in different contexts. It could be a high/ low five, gentle/ strong touch, slow/ fast movement, flexed/ extended arms, etc. And it happens between two people who both could actively move their hands to perform it, or sometimes one person is actively giving the high five while the other passively receives it.

During the experiment, participants were asked to high-five the robot's arm in various scenarios while comparing the difference between two arm conditions:

- C1 : Low arm inertia.
- C2 : High arm inertia.

Although knowing that adjusting the arm's inertia is realized by changing the arm's posture, to prevent any psychological effect on the participants during their interaction with the robot and to guarantee that their impression is evaluated solely based on their physical contact, the following setups were conducted:

- Instead of comparing two different arm postures (flexed and extended), we used a single posture (extended) and inserted inside the arm some weights to get the same effect of changing the arm's inertia without being noticed by the participants. This is demonstrated in figure 11.

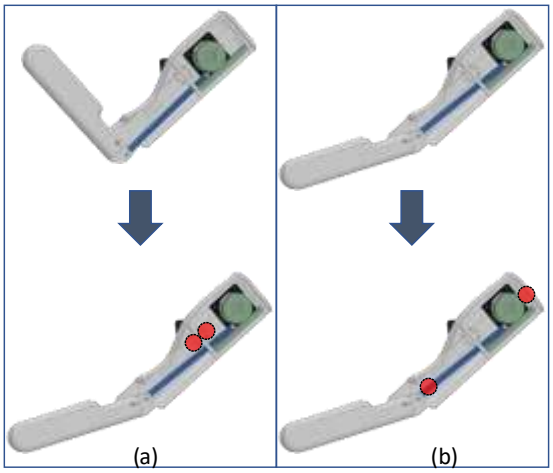


Fig. 11. Illustration diagram of the arm conditions tested during the high five experiment. (a) Condition C1 of low arm inertia. Two mass units are inserted near the shoulder's axis of rotation to represent the arm flexion. (b) Condition C2 of high arm inertia. The mass units are placed away from the central axis of rotation to represent the arm extension.

- The arms were detached from the robot, and the experiment was conducted away from it to exclude any psychological influence from the robot's appearance.
- Two arms with an identical appearance and different inertia, representing the two conditions, were presented to the participants as shown in figure 12.

The participants compared the difference between the two arm conditions in four different scenarios:

1. Passive High Five (PH): The robot's arm is passive. The participant actively high-fives the robot.
2. Passive Low Five (PL): The robot's arm is passive. The participant actively low-fives the robot.
3. Active Low Five 1 (AL1): The robot's arm is active. The participant places his/her hand on a designated spot. The robot's arm then performs a low five at a constant speed.
4. Active Low Five 2 (AL2): the settings here are similar to AL1's; however, instead of performing the low five at a constant speed, the robot's arm accelerates freely before touching the participant.

The robot's arm performs the high five with a single DoF by rotating around the shoulder axis. The applied torque value at the shoulder joint is the same for both conditions during every scenario, 0.3 Nm during the active cases and 0.0 Nm for the passive ones. The followed procedures during each of the scenarios mentioned above are described below:

1. The participant high/low-fives the robot's arm under condition C1 three times.
2. The participant high/low-fives the robot's arm under condition C2 three times.
3. Repeat the first two steps.
4. The participant then answers a questionnaire that compares the two conditions.

Fig. 12. Experimental setups of the high five experiment.

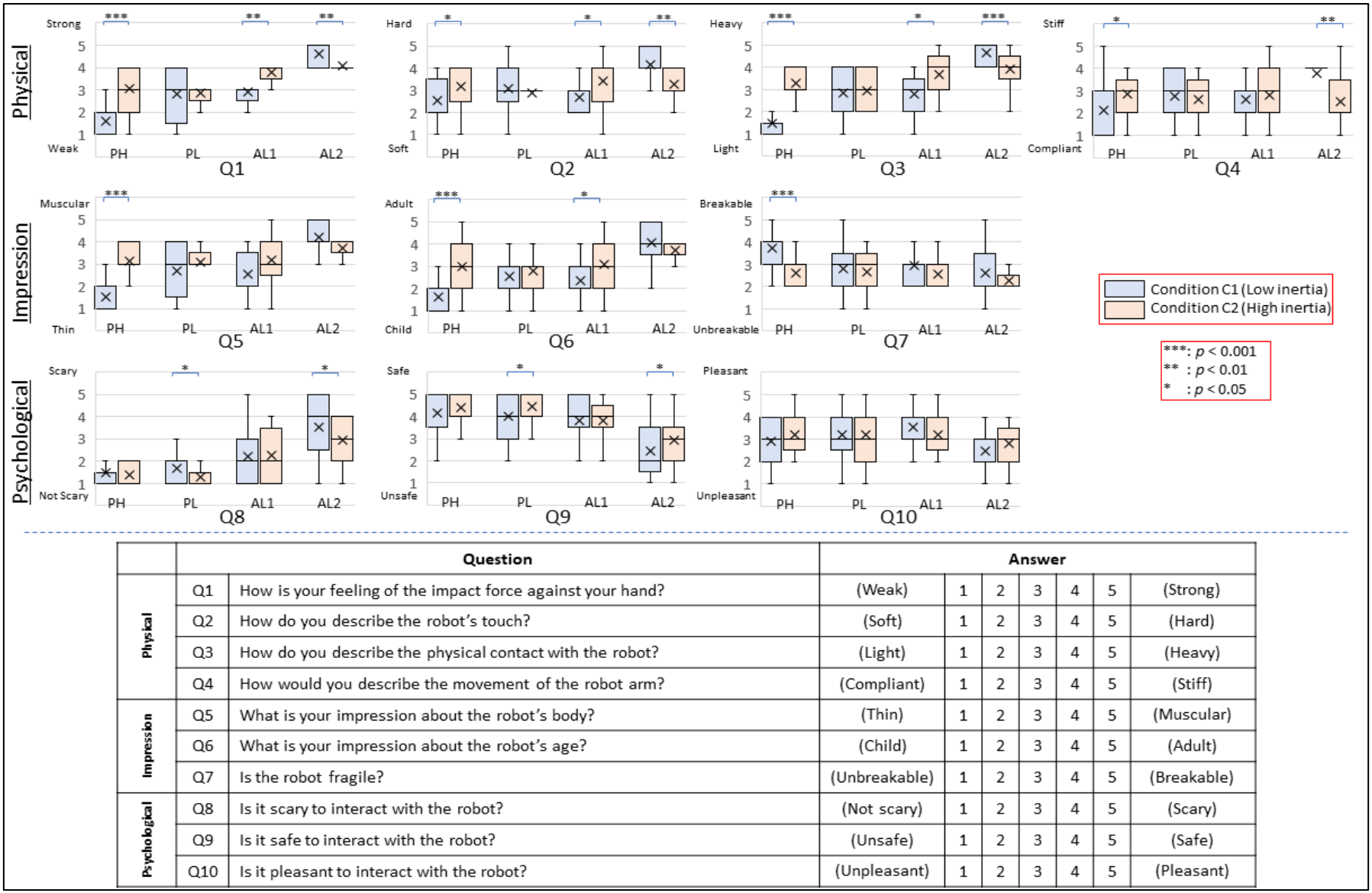
B. *Results*

Fifteen participants, 13 males and 2 females with an average age of (31.9 ± 7.4), compared the difference between the two arm conditions in each scenario by answering a questionnaire of 10 questions. The answers were rated on a 5- point semantic differential scale as demonstrated in figure 13. The questions were selected to address three categories, the physical aspects, the impression of the robot, and the psychological effects.

As depicted in the charts of figure 13, the results clearly demonstrate how changing the arm's inertia can be utilized to perceive different aspects of feelings. Although the same torque value was applied at the shoulder joint for both arm conditions in every scenario, different physical feelings, psychological impacts, and impressions were obtained.

Regarding the physical aspects, the high five with the robot arm under condition C1 was perceived to be more compliant with lighter and softer touch compared to condition C2 during the first scenario (PH). These results were expected since the arm under condition C1 had lower inertia compared to the arm under condition C2. Similarly, relatively the same difference between the two arm conditions can be seen during the third scenario (AL1), where the robot arm actively performs a low five at a constant speed. However, as the arms freely accelerated prior to the physical contact during the fourth scenario (AL2), opposite physical feelings were sensed. The reason behind this is that even though the same torque value was applied to the shoulder joint, the low inertia of the robot's arm under condition C1 caused the arm to rotate with higher angular velocity as shown in figure 14. This caused the sensation of a stronger impact force against the participants' hands.

These differences in perceiving the physical touch led to influence the participants' impressions of the robot and affected their psychological feelings. The robot arm under condition C2 (high inertia) was perceived during the PH and AL1 scenarios to belong to a robot that is more likely to be an adult with a



		Question	Answer						
Physical	Q1	How is your feeling of the impact force against your hand?	(Weak)	1	2	3	4	5	(Strong)
	Q2	How do you describe the robot's touch?	(Soft)	1	2	3	4	5	(Hard)
	Q3	How do you describe the physical contact with the robot?	(Light)	1	2	3	4	5	(Heavy)
	Q4	How would you describe the movement of the robot arm?	(Compliant)	1	2	3	4	5	(Stiff)
Impression	Q5	What is your impression about the robot's body?	(Thin)	1	2	3	4	5	(Muscular)
	Q6	What is your impression about the robot's age?	(Child)	1	2	3	4	5	(Adult)
	Q7	Is the robot fragile?	(Unbreakable)	1	2	3	4	5	(Breakable)
Psychological	Q8	Is it scary to interact with the robot?	(Not scary)	1	2	3	4	5	(Scary)
	Q9	Is it safe to interact with the robot?	(Unsafe)	1	2	3	4	5	(Safe)
	Q10	Is it pleasant to interact with the robot?	(Unpleasant)	1	2	3	4	5	(Pleasant)

Fig. 13. 5-point Semantic Differential Scale questionnaire. The charts show the participants' answers that compare the arm conditions, C1 and C2, during the four scenarios: PH, PL, AL1, and AL2.

muscular body that is unlikely to break compared to the other condition. On the other hand, relatively opposite results were noticed during the AL2 scenario

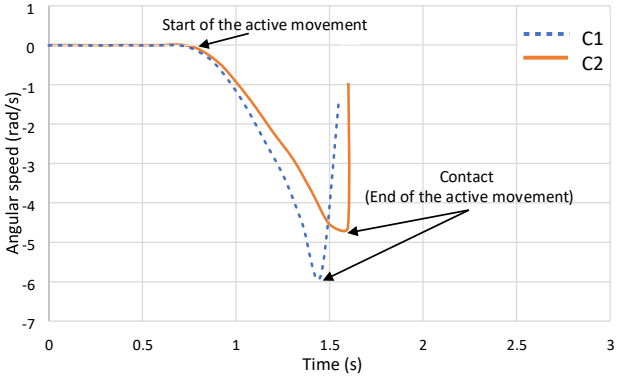


Fig. 14. The angular speed of the robot's arm under the conditions, C1 and C2, during the AL2 scenario of high five.

The last three questions of the questionnaire address the psychological impact of the conducted experiment. During the passive arm scenarios (PH and PL) and active arm with constant speed (AL1), the interactions were always perceived to be safe and not scary regardless of the arm condition. However, the interactions during the AL2 scenario were rated to be less safe and scarier. Especially under condition C1, which had an average score of less than 2.5 on the (safety scale) and more than 3.5 on the (scary scale).

V. DISCUSSION

The coexistence of robots and humans in a shared workspace requires the robotic systems to have adaptive bodies to handle the various forms of physical interaction. These robots must perform diverse touch behaviors to convey information, navigate through unstructured environments where physical contact is inevitable, utilize the surrounding objects through planned movements, or even produce spontaneous, unplanned reactions to adapt to sudden stimuli. Thus, it's impossible to model every type of physical interaction. In this regard, we gave extra attention to the robot's embodiment and employed passive mechanisms to achieve adaptive and safe interactions.

The selection of backdrivable and noiseless motors for the robot's development enriched the quality of the human-robot interaction experience. Although these motors had low mechanical power, the implemented mechanisms made it possible for the robot to operate and produce enough force at the end effector to influence the participants' feelings.

In the first experiment, we only focused on showing the mechanical advantage of the robot's body in restoring the equilibrium position passively. However, suppressing the vibration can be further enhanced by the active actuation of the body's actuator. Additionally, knowing that the performed poking gesture is a form of physical touch usually used to convey information [18], [19], this experiment can be extended to investigate the psychological impact on the participant under different scenarios.

The second experiment was conducted away from the robot's body to exclude any psychological influence from the robot's appearance. However, since the experiment came to support our expectations, the arms will be returned to the robot, and the effects of simultaneous engaging of the different body parts during the interaction will be studied.

The current version of the robot has limited DoF that confine its capabilities. Therefore, the next step in our research is to upgrade the robot's design to include more DoF. Furthermore, the body's actuators will be directly linked with each other to form a network of mutually interconnected parts. This will generate coordinated body movements and enhance the robot's adaptability [10]. Additionally, the robot's appearance needs to be carefully considered to encourage interaction [20], [21].

The experiments carried out in this paper were demonstration samples to show the potential benefits of this robot. Although both experiments focused on evaluating the employed mechanism for each body part separately, the coordination between the different body parts can be utilized for a wide range of experiments. For example, walking hand-in-hand with the robot, adaptability to various collisions, and navigation between objects without precise knowledge of the environment can be tested after upgrading the robot.

VI. CONCLUSION

In this paper, we attempted to build a prototype robot that is always ready for safe physical interactions. We gave extra attention to the robot's embodiment to take advantage of the passive dynamics. We used GC mechanisms for building the robot's arms, an MD mechanism for the body, and implemented soundless motor modules with smooth backdrivability.

To validate the robot's adaptive behavior, we conducted two experiments that examined the employed mechanisms. The first experiment investigated the role of the body mechanism in suppressing the vibration after encountering an external force. The applied method significantly reduced the oscillation amplitude by absorbing the system's kinetic energy. The second experiment, on the other hand, demonstrated how the arm's mechanism would influence the participants' impressions of the robot during their physical interaction. By asking the participants to high-five the robot in different scenarios, the results clearly showed how varying the arm conditions could be utilized to perceive various aspects of feelings.

Although the robot's structure is simple and confined to limited DoF, the obtained results illustrated the potential advantages of following this approach for various fields of applications.

REFERENCES

[1] M. J. Hertenstein, J. M. Verkamp, A. M. Kerestes, and R. M. Holmes, "The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research," Genetic, social, and general psychology monographs, 132(1), 5-94, (2006).

[2] S. Yohanan, and K. E. MacLean, "The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature," International Journal of Social Robotics, 4(2), 163-180, (2012).

[3] E. M. Kerruish, "Affective touch in social robots," Transformations, 29, 116-134, (2017).

[4] J. Park, and O. Khatib, "Robot multiple contact control," Robotica, 26(5), 667-677, (2008).

[5] J. Carpentier, A. Del Prete, S.Tonneau, T. Flayols, F. Forget, A. Mifsud, ... and N. Mansard, “Multi-contact locomotion of legged robots in complex environments–the loco3d project,” In RSS workshop on challenges in dynamic legged locomotion (p. 3p), (2017, July).

[6] P. Seiwald, S. C. Wu, F. Sygulla, T. F. Berninger, N. S. Staufenberg, M. F. Sattler, ... and F. Tombari, “LOLA v1. 1–An Upgrade in Hardware and Software Design for Dynamic Multi-Contact Locomotion,” In 2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids) (pp. 9-16). IEEE, (2021, July).

[7] A. De Santis, B. Siciliano, A. De Luca, and A. Bicchi, “An atlas of physical human–robot interaction,” Mechanism and Machine Theory, 43(3), 253-270, (2008).

[8] R. Pfeifer, and J. Bongard, “How the body shapes the way we think: a new view of intelligence,” MIT press, (2006).

[9] H. Ahmad, Y. Nakata, Y. Nakamura, and H. Ishiguro, “PedestriANS: a bipedal robot with adaptive morphology,” Adaptive Behavior, 29(4), 369-382, (2021).

[10] H. Ahmad, Y. Nakata, Y. Nakamura, and H. Ishiguro, “Adjustable whole-body dynamics for adaptive locomotion: the influence of upper body movements and its interactions with the lower body parts on the stable locomotion of a simple bipedal robot,” Robotica, 40(9), 3340-3354, (2022).

[11] R. Van, T. Sugar, B. Vanderborght, K. Hollander, and D. Lefeber, “Compliant actuator designs. review of actuators with passive adjustable compliance/controllable stiffness for robotic applications,” IEEE Robotics Automation Magazine, 16(3), 81-94, (2009).

[12] B. Vanderborght, A. Albu-Schäffer, A. Bicchi, E. Burdet, D. G. Caldwell, R. Carloni, ... and S. Wolf, “Variable impedance actuators: A review,” Robotics and autonomous systems, 61(12), 1601-1614, (2013).

[13] N. Elango, and A. A. M. Faudzi, “A review article: investigations on soft materials for soft robot manipulations,” The International Journal of Advanced Manufacturing Technology, 80(5), 1027-1037, (2015).

[14] A. J. Ijspeert, “Biorobotics: Using robots to emulate and investigate agile locomotion,” science, 346(6206), 196-203, (2014).

[15] S. Mintchev, and D. Floreano, “Adaptive morphology: A design principle for multimodal and multifunctional robots,” IEEE Robotics & Automation Magazine, 23(3), 42-54, (2016).

[16] Keigan Motor. https://en.keigan-motor.com/

[17] M. Gutierrez Soto, and H. Adeli, “Tuned mass dampers,” Archives of Computational Methods in Engineering, 20(4), 419-431, (2013).

[18] M. J. Hertenstein, R. Holmes, M. McCullough, and D. Keltner, “The communication of emotion via touch,” Emotion, 9(4), 566-573, (2009).

[19] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, and A. R. Jaskolka, “Touch communicates distinct emotions,” Emotion, 6(3), 528-533, (2006).

[20] M. Mori, “The uncanny valley,” Energy, 7(4), 33-35, (1970).

[21] F. Hegel, M. Lohse, and B. Wrede, “Effects of visual appearance on the attribution of applications in social robotics,” In RO-MAN 2009-The 18th IEEE International symposium on robot and human interactive communication (pp. 64-71). IEEE, (2009).

Flow Direction Control Using a Circular Cylinder with a Single/Double Slot

Daiki Yaguchi
Mechanical Engineering Program
Kogakuin University
Tokyo, Japan
am21064@ns.kogakuin.ac.jp

Kohei Okuma
Mechanical Engineering Program
Kogakuin University
Tokyo, Japan
a219019@ns.kogakuin.ac.jp

Kotara Sato
Department of Mechanical System Engineering
Kogakuin University
Tokyo, Japan
at12164@ns.kogakuin.ac.jp

***Abstract*—A circulation control wing (CCW) is sometimes used as a high-lift device for aircraft. A circular cylinder with tangential blowing is a type of CCW, and a method that utilizes the hydrodynamic forces generated in these circular cylinders has been put to practical use in helicopters. However, most previous studies have focused on a single slot. In this study, circular cylinders with single- and double-slotted tangential blowing were used to control the direction of the primary jets. The flow characteristics around the single- and double-slotted circular cylinders were compared, and the commonality/differences were analyzed.**

Keywords—Coanda Effect, Circulation Control Wing, Jet Vectoring

I. INTRODUCTION

A circulation control wing (CCW) is sometimes used as a high-lift device for aircraft. Flaps are still widely used as high-lift devices, but they require a change in geometry, such as the flap angle, and many moving parts to adjust the lift. In contrast, a CCW including circular cylinders with tangential blowing does not require moving parts because the lift is adjusted by the momentum of the jet sheets (tangential blowing) ejected from the slots.[1]-[8]Circular cylinders with tangential blowing are CCWs and have already been put to practical use to suppress helicopter self-rotation. Studies on the flow around a circular cylinder with tangential blowing have been reported in past confirmation coordination[9]–[11], and the fluid force and oscillation characteristics have been clarified. Recently, there have been reports on the directional control of jets for applications in fields other than aeronautics engineering[9]–[11].However, most previous reports have focused on a single-slotted circular cylinder with tangential blowing, and there have been only a few studies on double-slotted cylinders, so the knowledge obtained is probably insufficient.In this study, an attempt was made to control the direction of the primary jet using two types of a circular cylinders with a single or double slot for jet sheets. In particular, the flow characteristics around circular cylinders were compared, and the commonalities and differences were analyzed.

II. NOMENCLATURE

W	:	Wind tunnel width [mm]
D	:	Cylinder diameter [mm]
$C_{\mu 1}$	:	Single-slot momentum ratio [-]
$C_{\mu 2}$	:	Double-slot momentum ratio [-]
U_p	:	Main stream velocity [m/s]
V_j	:	Cylinder blowout jet velocity [m/s]
θ	:	The angle from the front edge of the cylinder [°]
b	:	Slot width [mm]
θ_j	:	Slot angle [°]

III. EXPERIMENTAL DEVICE AND METHOD

Figure 1 shows the test section and coordinate system of the experimental apparatus. A wind tunnel is used to generate the primary jet, and circular cylinders are placed 200 mm from the wind tunnel outlet center (y = 0) in the x-direction. The outlet width of the wind tunnel is W = 200 mm, and the height is 200 mm. Figure 2 shows the geometric shape of the test circular cylinder with slots and an enlarged view of the slot. The circular cylinder has a diameter of D = 50 mm, spanwise length of 200 mm, and slot width b = 1 mm for the jet sheets. The working fluid is air, and the flow generated by the blower leads to a plenum tank, from which it flows out as tangential blowing from the slots through the cavity in the central part of the cylinder. The slot angle θj is variable. The circular cylinder center is defined as the origin, and the angle is defined as positive clockwise.The velocity profiles were measured using hot-wire anemometers. The velocity distribution was measured at intervals of 5° in the range θ = 90°–300° on a reference circular arc of radius r = 300 mm from the origin. For flow visualization, a smoke generator was installed at the suction port of the wind tunnel, and the behavior of the smoke emitted

from the wind tunnel outlet was captured using a digital camera with a frame rate of 480 fps.

1. Numerical Models and Boundary Conditions

ANSYS Fluent (ANSYS, Inc.), a general-purpose thermal fluid analysis software with an unstructured grid, was used for numerical simulations. The standard k–ε model was applied to the turbulence model, assuming a two-dimensional flow of incompressible viscous fluid. The number of elements in the mesh was approximately 200,000. As boundary conditions, flow velocity regulation was applied to the primary jet outlet and slot outlet of the jet sheets. Static and total pressure regulations were imposed on the outlet boundary of the calculation area and upper and lower boundaries, respectively. Nonslip conditions were also applied to the cylinder surface and slot sidewalls.

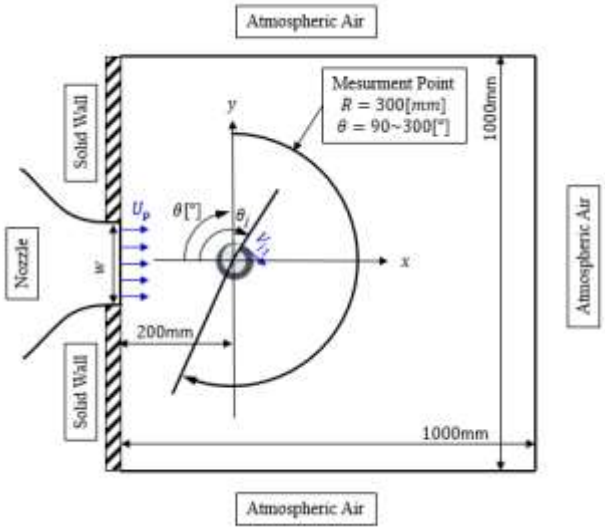


Fig. 1. Experiment summary chart.

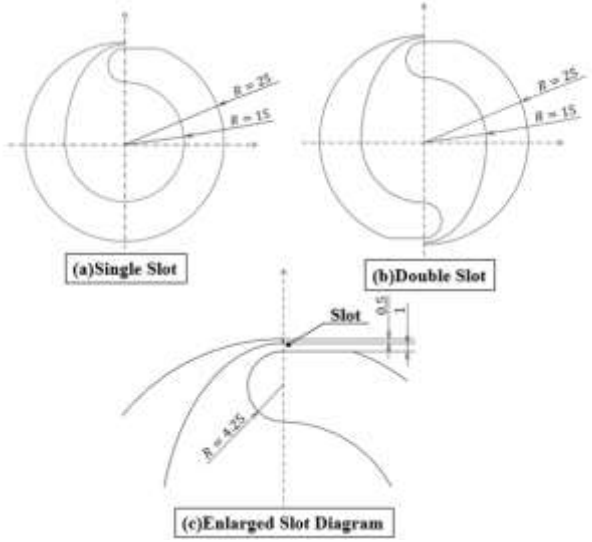


Fig. 2. Test cylinder.

IV. RESULTS AND DISCUSSION

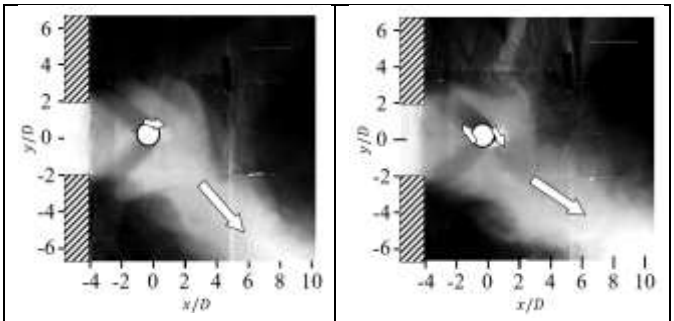
The degree of deflection of the primary jet considered is highly dependent on the momentum coefficient[11]. The momentum coefficient is defined by Equation (1) for the single-slot case and Equation (2) for the double-slot case.

$$C_{\mu 1} = \frac{{V_j}^2 b}{{U_P}^2 W} \quad (1)$$

$$C_{\mu 2} = \frac{{V_j}^2 2b}{{U_P}^2 W} \quad (2)$$

The primary jet velocity was kept constant at Up = 8.3 m/s in this experiment.

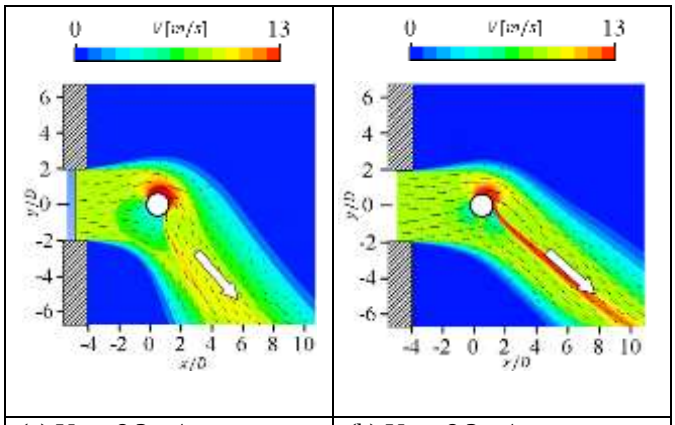
Figure 3 shows the experimental results obtained using flow visualization. Figures 3(a) and (b) show examples of the observed flows for single and double slots, respectively. In this figure, the flow fields for single and double slots are compared under the condition that the slot angles are adjusted to achieve approximately equal jet deflection angles with an equal momentum coefficient Cμ = 0.27. Figure 3(a) shows the behavior for blowing velocity Vj = 60 m/s and θj = 100°, and Figure 3(b) shows the behavior for Vj = 43 m/s and θj = 160° and 340°. In both figures, the jets travel in approximately the same direction (60°) despite the large difference in slot angle θj. However, quantitatively, the jet deflection angle in Figure 3(a) slightly exceeds that in Figure 3(b). Preliminary experiments confirmed that the columnar wake oscillates slightly more in the single-slot case than in the double-slot case.



(1) Up = 8.3 m/s, V_j = 60 m/s, $C_{\mu 1}$ = 0.27, W = 200 mm θ_j = 100°	(2) Up = 8.3 m/s, V_j = 43 m/s, $C_{\mu 2}$ = 0.27, W = 200 mm θ_j = 160,340°

Fig. 3. Experimental results from visualization.

Figure 4 shows the numerical simulation results obtained under the same conditions as those in Figure 3. Figure 4(a) shows the single-slot case, and Figure 4(b) shows the double-slot case. Qualitatively, the flow visualization observation results in Figure 3 and the numerical results in Figure 4 agree well. Comparing the flow direction at the center of the jet downstream of the cylinder in Figures 4(a) and (b) reveals that the deflection angle in Figure 4(a) is slightly larger than that in Figure 4(b), which corresponds to the experimental results.



(a) Up = 8.3 m/s, V_j = 60 m/s, $C_{\mu 1} = 0.27$, W = 200 mm θ_j = 100°	(b) Up = 8.3 m/s, V_j = 43 m/s, $C_{\mu 2} = 0.27$, W = 200 mm θ_j = 160,340°

Fig. 4. CFD analysis results

Figure 5 shows the time-averaged velocity in minutes on an arc of r = 300 mm measured under the same experimental conditions as the visualization experiment in Figure 3. The horizontal axis is angle θ, the vertical axis is the time-averaged velocity $\sqrt{u^2 + v^2}$, and the parameter is the number of slots, where the red and blue markers represent single and double slots, respectively. Both velocity profiles show a maximum velocity of approximately 265°, and there is no significant difference between the single- and double-slot velocity profiles. One difference is that the single slot exhibits a velocity defect of approximately 240°, unlike the double slot. Because the difference in the velocity gradient around 240° is also related to vortex formation in the wake, it may appear as a difference in the unsteady characteristics. The verification of the unsteady characteristics is an issue requiring further research.

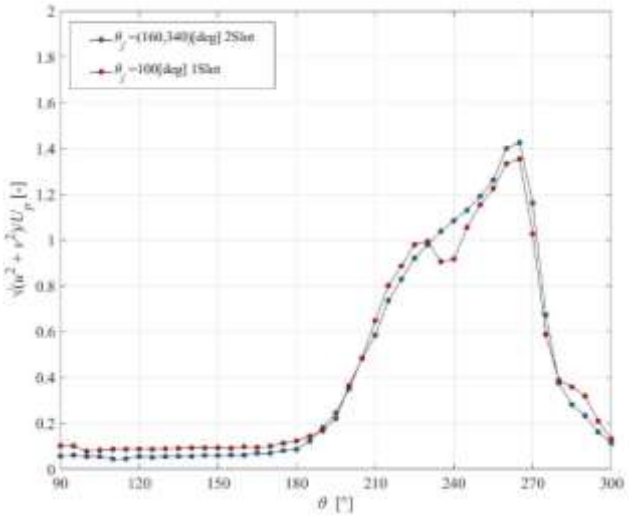


Fig. 5. Velocity distribution

V. CONCLUSION

To compare the flow characteristics of single- and double-slotted circular cylinders, slot positions that exhibited similar deflection angles at the same momentum ratio were examined. The conclusions are as follows.

(1) The slot positions are different for the single and double slots to deflect the jet to the same degree at the same momentum.

(2) There is a slight difference in the velocity distributions of the backward flow between the single and double slots.

REFERENCES

[1] T. Shakouchi," Jet Flow Engineering", Fundamentals and Application, Morikita Publishing, 2004

[2] H. Yamada, "Effects of End Plates on Spanwise Characteristics of Flow around a Circular Cylinder," Trans. JSME B, 58, no. 552, pp. 2368–2373, 1992

[3] F. Yoshino, R. Waka, T. Iwasa, and T. Hayashi, "The Effect of a Side-Wall on Wing Characteristics of a Circular Cylinder with Tangential Blowing," Trans. Jpn. Soc. Mech. Eng. B, 46, pp. 1890–1898, 1980

[4] R. Waka, F. Yoshino, and T. Hayashi, "Distributions of Characteristic Values near a Side-Wall of a Circular Cylinder with Tangential Blowing: Influences of Slot Structure of a Cylinder-Sidewall Juncture and Angular Location of a Blowing Slot," Trans. Jpn. Soc. Mech. Eng., 50, pp. 2307–2315, 1984

[5] R. Waka, F. Yoshino, T. Hayashi, and T. Iwasa, "The Aerodynamic Characteristics at the Mid-Span of a Circular Cylinder with Tangential Blowing," Trans. Jpn. Soc. Mech. Eng. B, 26, no. 215, pp. 755–762, 1983

[6] S. Okayasu, K. Sato, T. Shakochi, and K. Furuya, "Flow around a Tangential Blowing Cylinder Placed near a Wall (1st Report, Consideration of Fluid Power)," Proc. of JSME, 73, no. 733, pp. 1790–1797, 2007

[7] S. Okayasu, K. Sato, T. Shakochi, and K. Furuya, "Flow around a Tangential Blowing Cylinder Placed near a Wall (2nd Report, Consideration of Unsteady Characteristics)," Proc. of JSME, 74, no. 744, pp. 1725–1734, 2008

[8] S. Okayasu, Y. Arakawa, K. Sato, T. Shakochi, and K. Furuya, "Influence of Ground Effect and Ceiling Effect on High Lift Device Using Tangential Blowout Cylinder," Proc. of the Japan Society for Aeronautical and Space Sciences, 56, no. 656, pp. 29–36, 2008

[9] A. Fuji and K. Sato, "Directional Control of Jet Flow Using a Tangential Spray Cylinder," Proc. of the 72nd General Meeting of the Kyushu Branch of the Japan Society of Mechanical Engineers

[10]) F. Takahashi and K. Sato, "Control of Flow Field by Cylindrical Wall Jet," The 57th Conference of Hokkaido Branch, The Japan Society of Mechanical Engineers

[11] Q. Zhang, F. Takahashi, K. Sato, W. Tsuru, and K. Yokota, "Jet Direction Control Using Circular Cylinder with Tangential Blowing," Trans. Japan Soc. Aerospace Sci., 64, no. 3, pp. 181–188, 2021

[12] T. Nakayama, D. Yaguchi, Q. Zhang, K. Sato, and K. Yokota, "Flow around a circular cylinder with double slots for the tangential blowing," The Japan Society of Mechanical Engineers, Kanto Student Association 61st Graduation Research Presentation by Students,2022

[13] S. Fukui, Q. Zhang, and K. Sato, "Fundamental study on jet vectoring by circulation control wing," International Symposium on Advanced Technology (ISAT20),2022

[14] D. Yaguchi, T. Nakayama, and K. Sato, "Flow characteristics around double slotted circulation control wing," International Symposium on Advanced Technology (ISAT20),2022

[15] D. Yaguchi, D. Kang, and K. Sato, "Flow Characteristics of Circular Cylinder with Tangential Blowing under Various Width Conditions of Main Jet Flow," The 32nd International Symposium on Transport Phenomena, March 19–21,2022

[16] R. Kobayashi, Y. Watanabe, Y. Tamanoi, K. Nishibe, D. Kang, and K. Sato, “Jet vectoring using secondary Coanda synthetic jets,” Bulletin of the JSME., 7, no. 5, 2022

[17] Y. Tamanoi, Y. Watanabe, R. Kobayashi, and K. Sato, “Jet direction control using secondary flows,” 18th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery, 2022

[18] Y. Tamanoi, R. Kobayashi, K. Sato, K. Nishibe, and D. Kang, “Flow Control Using Excited Jets with Coanda Surfaces,” The 31st International Symposium on Transport Phenomena, 2022

[19] Y. Tamanoi, and K. Sato, “Structure of Jet Deflected by Secondary Flow,” International Symposium on Advanced Technology (ISAT-19), 2021

[20] R. Kobayashi, H. Terakado, K. Sato, J. Taniguchi, and K. Nishibe, “Behavior of Plane Synthetic Jets Generated by an Asymmetric Stepped Slot,” International Journal of Fluid Machinery and Systems, 1882–9554, 2018

Simulating the Impact of Tractive Power on the Operating Fuel Economy of Light Duty Vehicles on Driving Cycles

Surath Gajanayake
Department of Mechanical Engineering
University of Moratuwa
Moratuwa, Sri Lanka
198099f@uom.lk

Saman Bandara
Department of Civil Engineering
University of Moratuwa
Moratuwa, Sri Lanka
bandara@uom.lk

Thusitha Sugathapala
Department of Mecahnical Engineering
University of Moratuwa
Moratuwa, Sri Lanka
thusitha@uom.lk

***Abstract*— Due to rapid motorization, fuel economy has become a major area of concern in the context of present-day automotive industry. When appraising the factors affecting the fuel usage of a light duty vehicle (LDV), tractive load provides a major contribution whereas other auxiliary engine loads provide a minor contribution. Thus, the study delves into its scope of evaluating the impact of tractive load on the operating fuel economy of LDVs. Tractive load provides the necessary energy for an LDV to propel on a given terrain and comprises of three main contributors viz. rolling resistance, aero-dynamic drag resistance and inertial resistance. A driving cycle can be defined as a speed-time profile of a given vehicle category which is utilized to test the performance of a given LDV. Hence, the said three major factors are analysed during the study using an analytical approach as a single parametric study and the derived relationships are validated using a simulation software, i.e., Future Automotive Systems Technology Simulator (FASTSim™). The study portrays that each of the individual factors of coefficient of rolling resistance, coefficient of aero-dynamic drag resistance and translational inertia have closely linear relationships with operating fuel economy of LDVs when considered as single parameters.**

Keywords*— *fuel economy; light duty vehicles; tractive power; rolling resistance; aero-dynamic drag; inertia

I. INTRODUCTION

The fuel economy is an essential parameter in assessing performance of an LDV. With the increasing number of vehicles manufactured and used on the roads, their fuel consumption has become a vital factor. The study limits its scope to LDVs, which can be defined as road vehicles having gross vehicle weight rating (GVWR) no more than 3.5 tons [1]. The study delves into the driving cycle dynamic variables of tractive force, tractive power, and wheel work respectively. These variables deliver the force, power and the energy required at the wheels during driving. The tractive force which needs at the wheels when driving is required to overcome the resistive forces viz., the rolling resistance, the aerodynamic drag resistance, the grade resistance, and the inertial resistance [2].

The conventional equation to determine the tractive force is stated in (1).

$$F_{TR} = \underbrace{(C_r.M.g)}_{\text{Rolling Resistance}} + \underbrace{\left(\frac{C_D.A_F.\rho.v^2}{2}\right)}_{\text{Drag Resistance}} + \underbrace{(M.g.\sin\alpha)}_{\text{Grade Resistance}} + \underbrace{(M.\delta.a)}_{\text{Inertial Resistance}} \quad (1)$$

Eq. (1) portrays the tractive force by summing the forces required at the wheels due to rolling, aerodynamic drag and grade resistance and inertia. Where, in the rolling resistance term, C_r is the coefficient of rolling resistance of the vehicle, M is the vehicle test mass and g is the gravitational acceleration. In the term for aerodynamic drag, ρ is the density of air (taken as a constant of 1.2 kg/m^3), C_D is the coefficient of drag, A_F is the frontal area of the vehicle, and v is vehicle speed. In the grade resistance term, α is the angle of the road gradient. In the inertia term, δ is a mass correction factor which accounts for the fact that the 4 rotating wheels must be angularly as well as linearly accelerated and is assumed constant at 1.04 [3]. Furthermore, during driving cycle test procedures, the gradient of the terrain remains unchanged; preferably at zero degrees. Thus, the grade resistance component can be approximated to zero and the resulting formula for the tractive force can be depicted as in (2).

$$F_{TR} = (C_r.M.g) + \left(\frac{C_D.A_F.\rho.v^2}{2}\right) + (M.\delta.a) \quad (2)$$

When a vehicle is moving at a constant speed on a flat and smooth surface, which means that the tractive power is equivalent to the power required to overcome drag resistance which is equal to the product of drag force (F_a) and vehicular speed (v) as depicted in (3) and (4) [4].

$$F_a = C_a \frac{1}{2} A\rho v^2 \quad (3)$$

$$P = F_a v \quad (4)$$

$$P_d \ \alpha \ v^3 \quad (5)$$

Hence as mentioned in (5), the power required to overcome the drag is proportional to the cube of the vehicular speed or the flow velocity (v^3) [5]. This critically emphasizes the impact of vehicular speed on the power required to overcome the aerodynamic drag and hence stresses on the fact that, the higher degree of increase in fuel consumption. Moreover, Kelly et al. has experimentally studied the power requirements of a typical passenger car cruising at a speed greater than 80 km/h and consequently found out that the power required to overcome the aerodynamic resistance is greater than that of rolling resistance of tyres and the resistance in the transmission [6]. During the study, a section is allocated to analyze the impact of aero-dynamic drag resistance, due to its significance in the contribution towards the tractive force. Furthermore, the study models the impact of coefficient of rolling resistance and the vehicular mass which represents the inertia on the operating fuel consumption of LDVs.

During the study, the impact of rolling resistance, aerodynamic drag resistance and inertia on fuel consumption of LDVs are analytically evaluated and subsequently the obtained relationships are validated using a simulation tool. The Future Automotive Systems Technology Simulator (FASTSim) provides a way to compare powertrains and estimates the impact of technology improvements on light-, medium-, and heavy-duty vehicle efficiency, performance, cost, and battery life. The said simulation tool has been developed by the National Renewable Energy Laboratory (NREL), United States for the automotive simulation purposes. It accommodates a range of vehicle types, including conventional vehicles, electric-drive vehicles, and fuel cell vehicles. It also simulates driving cycle test procedures and estimates fuel consumption values under varying driving cycles. This attribute is utilized during the study to generate results by simulating the selected types of Light Duty Vehicles (LDVs) under two main test cycles, i.e., US06 driving cycle and JP10 driving cycle.

TABLE I . THE COMPARISON BETWEEN THE CHARACTERISTICS OF US06 AND JP10 DRIVING CYCLES.

CPs	US06	JP 10-15
Duration (s)	596	660
Distance (m)	12894	4165
Average Speed (m/s)	21.64	6.31
Maximum Speed (m/s)	35.81	19.47
Average Running Speed (m/s)	22.12	8.54
Average Acceleration (m/s^2)	0.54	0.37
Average Deceleration (m/s^2)	0.57	-0.39
Idle Time Percentage %	2.18	26.06

The details of the driving cycles used for the simulations during the study are provided as depicted in Table 1. The reason for opting the two specific driving cycles is that US06 cycle reflects a higher average speed with less stop-go traffic behaviour whereas JP10 cycle reflects mainly the stop-go traffic behaviour with less average speed and comparatively higher idle time percentage. Thus, it's able to simulate the performance of a vehicle in terms of fuel economy pertaining to two different driving conditions.

II. RELATIONSHIP BETWEEN FUEL ECONOMY AND THE COEFFICIENT OF ROLLING RESISTANCE

The rolling resistance of the tyres is one of the key factors affecting the operating fuel economy of a vehicle. A pneumatic tyre is a flexible structure of the shape of a toroid filled with compressed air [7]. The most important element of the tyre is the carcass [8]. It is made up of a number of layers of flexible cords of high modulus of elasticity encased in a matrix of low modulus rubber compounds [8]. There are 2 types of tyres w.r.t the type of plies, i.e., bias-ply and radial-ply tyres [8]. The extension of layers of cords within the carcass assist the area of the tyre which touches the terrain to get deformed. Rolling resistance contributes for 10-13% of the overall fuel consumption of a vehicle [9], [10].

Eq. (3) is used when modelling the relationship between fuel consumption and coefficient of rolling resistance. By multiplying Equation (2) by v with reference to $P = F.v$ which is stated in (6), (7) and (8) can be derived.

$$P = F.v \quad (6)$$

$$F_{TR}.v = \{(C_r.M.g) + \left(\frac{C_D.A_F.\rho.v^2}{2}\right) + (M.\delta.a)\}.v \quad (7)$$

$$P_{TR} = \{(C_r.M.g) + \left(\frac{C_D.A_F.\rho.v^2}{2}\right) + (M.\delta.a)\}.v \quad (8)$$

In Eq. (8), P_{TR} denotes the tractive power at the wheels. During the study, it's only considered the LDVs with IC engines. Therefore, the brake power generated by the IC engine can be determined using (9).If the thermal efficiency of an internal combustion engine is $'\eta'$, the amount of energy generated by the heat of combustion: q_c, mass fuel flow rate $\dot{m}_f$ then, the brake power, P_b can be determined as in (9) [11], [12]:

$$P_b = \eta.\dot{m}_f.q_c \quad (9)$$

Assuming the equivalence of brake power and the tractive power (10) and (11) can be determined.

$$P_b = P_{TR} \quad (10)$$

$$\eta . \dot{m}_f . q_c = \{(C_r Mg) + \left(\frac{C_D . A_F . \rho . v^2}{2}\right) + (M\delta a)\} v \quad (11)$$

$$\dot{m}_f = \frac{1}{\eta . q_c}\{(C_r Mg) + \left(\frac{C_D . A_F . \rho . v^2}{2}\right) + (M\delta a)\} v \quad (12)$$

Partially differentiating (12) w.r.t C_r, in order to evaluate the relationship between C_r and $\dot{m}_f$, (13) to (15) can be determined.

$$\frac{\partial}{\partial C_r}\dot{m}_f = \frac{\partial}{\partial C_r}\{\frac{v}{\eta . q_c}\{[(C_r . g) + (. \delta . a)\}. M + \left(\frac{C_D . A_F . \rho . v^2}{2}\right)\}\} \quad (13)$$

$$\frac{\partial}{\partial C_r}\dot{m}_f = \frac{v . g . M}{\eta . q_c} . \frac{\partial C_r}{\partial C_r} + \frac{\partial}{\partial C_r} \delta a M v \eta q_c + \frac{\partial}{\partial C_r}\frac{C_D . A_F . \rho . v^3}{2\eta . q_c} \quad (14)$$

$$\frac{\partial}{\partial C_r}\dot{m}_f = \frac{vgM}{\eta q_c} \quad (15)$$

In Eq. (15), the gravitational acceleration g , engine efficiency η, specific energy of the fuel q_c, mass of the vehicle M and vehicular speed v are estimated to be constant and to be positive values. Hence:

$$\frac{\partial \dot{m}_f}{\partial C_r} = K\ (\in R) \quad (16)$$

Therefore, it can clearly be seen the fact that, the result of (16) yields a linear relationship between $\dot{m}_f$ and C_r.

$$FE = (v\ d / \dot{m}_f)10^{-3} \quad (17)$$

The output of the exercise should be the operating fuel economy of a given vehicle. In (17), d is the density of the respective fuel. The fuel economy (FE) can be determined using (17) which is derived during the study. As per (17), FE does have an inverse proportional relationship with $\dot{m}_f$. Consequently, FE portrays an inverse proportional relationship with C_r. Using 'FAST Sim' Simulation tool, the relationship between FE and C_r is determined. Using the simulation results, the relationship is depicted graphically. The simulation experiment is set up by keeping all the other variables constant. It is a Toyota Corolla 2016, 4-cylinder, 2-wheel drive car which has been opted for simulations. Two driving cycles which have been used for the simulation purposes are EPA US06 and JP10. The US06 is the supplemental federal test procedure adopted for LDVs in the United States whereas JP10 is the latest version of Japanese driving cycle adopted for LDVs. The vehicle GVWR is kept constant at 1046 kg whereas the drag coefficient is kept constant at 0.3. The simulation results which have been obtained for the two driving cycles have then been plotted using MATLAB and the resulting curves can be portrayed as in Fig. 1.

Linear regression is used in MATLAB to interpolate the relationship between FE versus the coefficient of rolling resistance (RR). The equation of the linear graph is derived as:

$$y = -126.7x + 12.2 \quad (18)$$

Here in Eq. (18), y denotes the FE and x denotes the C_r. Statistically the R-squared value can be determined as 0.9516 whereas norm of residuals lie quite low as 1.107 which depicts the goodness of the fit. It could clearly be seen that, the said graph has a negative, constant gradient and hence can be proven that FE and C_r have a linear relationship as mentioned in Equation (16) for the US06 driving cycle related data.

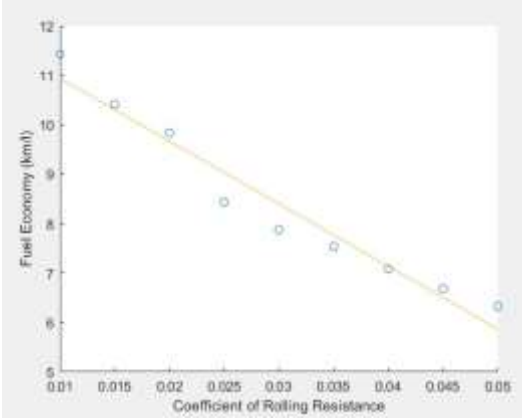


Fig. 1. Fuel Economy versus Coefficient of Rolling Resistance relationship w.r.t US06.

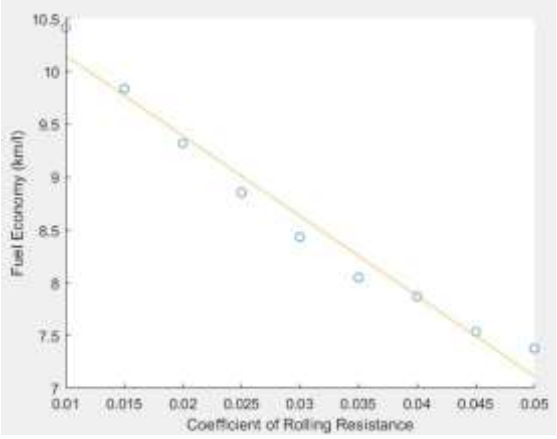


Fig. 2. Fuel Economy versus Coefficient of Rolling Resistance relationship w.r.t JP10.

The respective equation of the graph can be determined as:

$$y = -75.88x + 10.91 \quad (19)$$

During the simulation test under JP10 driving cycle, same initial conditions as tested for the US06 are adopted. Similar MATLAB code is also used. Here also it can clearly be cited that the fuel economy has depicted a linear relationship w.r.t the coefficient of rolling resistance. Moreover, the regression model reflects better statistics since R-squared value lies at 0.971 whereas norm of residuals lies at 0.5082. From the simulation results under both US06 and JP10 driving cycles, it's conspicuous that FE maintains a linear proportional

relationship with coefficient of rolling resistance when a single parametric analysis is performed.

III. RELATIONSHIP BETWEEN FUEL CONSUMPTION AND THE COEFFICIENT OF AERODYNAMIC DRAG RESISTANCE

The tractive force (F_{TR}) is comprised of four main components as previously noted in (1). F_D, the aerodynamic drag force is one of the main components of the said tractive force and the determination of the drag force is depicted in Equation (20).

$$F_D = \frac{1}{2}.C_D.\rho.A_F.v^2 \tag{20}$$

In Eq. (20), C_D denotes the aerodynamic drag coefficient, ρ denotes the density of air, A_F denotes the frontal area of the respective vehicle and v denotes the vehicular speed. When considering the factors affecting the aerodynamic drag force, the drag coefficient has a significant impact. The study analyzes the effect of drag coefficient on the overall operating fuel economy of an LDV on two different driving cycles, i.e., US06 and JP10. The tractive power can be determined as described in Equations (7) and (8). Here the terminology is similar to that of Equation (9). Subsequently, the relationship mentioned in (21) and (22) can be determined.

$$\eta.\dot{m}_f.q_c = \{(C_r.M.g) + \left(\frac{C_D.A_F.\rho.v^2}{2}\right) + (M.\delta.a)\}v \tag{21}$$

$$\dot{m}_f = \frac{1}{\eta.q_c}\left\{(C_r.M.g) + \left(\frac{C_D.A_F.\rho.v^2}{2}\right) + (M.\delta.a)\right\}v \tag{22}$$

Partially differentiating (22) w.r.t C_D, in order to determine the relationship between C_D and $\dot{m}_f$, (23) can be determined.

$$\frac{\partial}{\partial C_D}\dot{m}_f = \frac{\partial}{\partial C_D}\{\frac{v}{\eta.q_c}\{[(C_r.g) + (.\delta.a)]M + \left(\frac{C_D.A_F.\rho.v^2}{2}\right)\}\} \tag{23}$$

Here $\frac{\partial \dot{m}_f}{\partial C_D}$, the rate of change of $\dot{m}_f$ w.r.t C_D can be stated as noted in Equation (24).

$$\frac{\partial}{\partial C_D}\dot{m}_f = \frac{A_F.\rho.v^3}{2\eta.q_c} \tag{24}$$

Here the gravitational acceleration g, engine efficiency η, specific energy of the respective fuel q_c, mass of the vehicle M and vehicular speed v are estimated to be constant. Hence (25) can be derived:

$$\frac{\partial \dot{m}_f}{\partial C_D} = K\ (\in R) \tag{25}$$

Therefore, it can clearly be seen the fact that, (25) yields a linear relationship between $\dot{m}_f$ and C_D. As stated in (17), the output should be conveyed in terms of the operating fuel economy (FE) . As per (17), FE does have an inverse proportional relationship with $\dot{m}_f$. Consequently, FE portrays an inverse linear relationship with C_D.

Using FAST Sim simulation tool, the relationship between the Drag Coefficient and the fuel economy is simulated under US06 and JP10 driving cycles respectively. The test results of the simulation study can be depicted as shown in Fig. 3 and Fig. 4. The constant conditions can be elaborated as the GVWR is 1046 kg, coefficient of rolling resistance is 0.01, frontal area is 2.574 m^2. The vehicle type is opted as 2016/Toyota Corolla/4 cylinder/2-wheel drive.

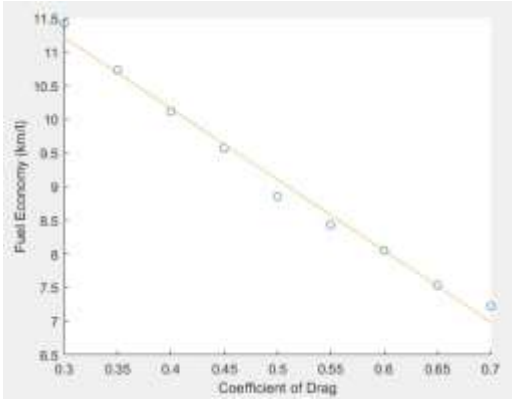


Fig. 3. - FE versus C_D relationship w.r.t US06 driving cycle, curve-fitted linearly.

The line in yellow in Fig. 3 is the curve fitted graph for the fuel economy points under US06 driving cycle. The graph equation can be determined as mentioned in (26)

$$y = -10.55x + 14.38 \tag{26}$$

The linear regression model for US06 driving cycle data reflects R-squared of 0.9888 and norm of residuals at 0. 4355.This depicts the goodness of the fit and it portrays an inverse linear relationship between the operating fuel economy and coefficient of drag.

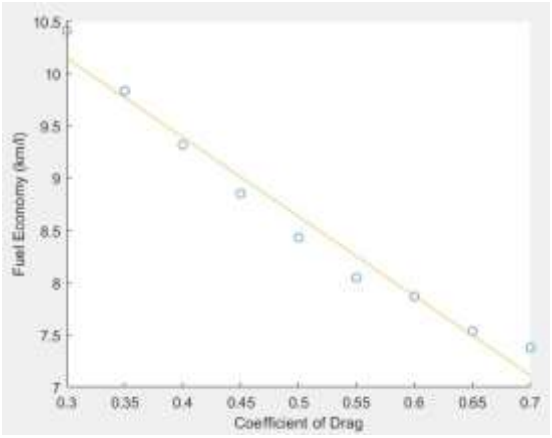


Fig. 4. FE versus C_D. relationship w.r.t JP10 driving cycle, curve-fitted linearly.

The line in yellow in Fig. 4. is the curve fitted graph for the fuel economy points under the JP10 driving cycle. The graph equation can be determined as mentioned in (27).

$$y = -7.588x + 12.43 \tag{27}$$

When considering the regression analysis, the R-squared value can be estimated as 0.971 which is a modest value whereas the norm of residuals lies as low as 0.5082. When delving into the graph of US06 driving cycle, it can clearly be seen that the fuel economy depicts an inverse linear relationship w.r.t coefficient of drag (C_D). Moreover, when analyzing the discrepancies between the plotted two graphs, it can be noted that the main reason of it is the difference in characteristics of the two driving cycles, especially the speed limitations. In US06 driving cycle, the maximum speed can be cited around 36 m/s whereas the top speed in JP-10 mode is around 20 m/s. As previously shown in (5), the tractive power required to overcome the drag resistance is proportional to the cube of speed and it portrays when the speed is increasing, the power requirement to overcome the drag resistance is mimicking a cubic relationship. This can be recommended as the main reason for the discrepancies between two graphs as shown in Fig.3 and Fig. 4.

IV. RELATIONSHIP BETWEEN FUEL ECONOMY AND THE VEHICULAR MASS

Mass of the vehicle or commonly referred to as the vehicular weight is another significant factor affecting the fuel economy of a vehicle. In this section, the relationship between the fuel economy and the vehicular mass will be evaluated in depth. Partially differentiating Equation (22) by M to obtain the relationship between M and $\dot{m}_f$, (28) can be derived:

$$\frac{\partial}{\partial M}\dot{m}_f = \frac{\partial}{\partial M}\left\{\frac{v}{\eta . q_c}\left\{[(C_r . g) + (\delta . a)].M + \left(\frac{C_D . A_F . \rho . v^2}{2}\right)\right\}\right\} \tag{28}$$

$$\frac{\partial}{\partial M}\dot{m}_f = \frac{v}{\eta . q_c}[(C_r . g) + (\delta . a)] \tag{29}$$

In (29), $\frac{\partial \dot{m}_f}{\partial M}$ represents the rate of change of $\dot{m}_f$ w.r.t M. In other terms, this depicts the gradient of the graph between $\dot{m}_f$ and M. Here when considering the R.H.S of Equation (29), C_r is kept constant. During the simulation test procedure, v and a are considered as constant average values since it's tested for a particular driving cycle. The gravitational acceleration, g, the mass correction factor, δ, engine efficiency, η and the specific energy of the fuel, $q_{c,}$ are estimated to be constant. Hence, the gradient of the graph between $\dot{m}_f$ and M is a constant and thus portrays a linear relationship.

$$\frac{\partial \dot{m}_f}{\partial M} = K\ (\in R) \tag{30}$$

Hence, it can be determined that, $\dot{m}_f$ and M have a linear proportional relationship between them. As stated in (17), the output should be expressed in terms of the operating fuel economy (FE) . As per (17), FE does have an inverse proportional relationship with $\dot{m}_f$. Consequently, FE portrays an inverse proportional relationship with M. The simulations are being carried out for US06 and JP10 driving cycles respectively. During the simulations the Drag Coefficient, the Coefficient of Rolling Resistance are kept constant.

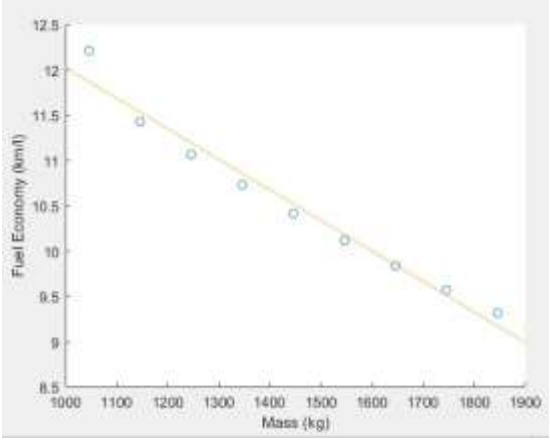


Fig 5. FE versus Vehicular Mass relationship under US06.

Fig. 5. denotes the curve fitted graph for the results of US06 driving cycle. The equation for the graph of US06 driving cycle can be determined as mentioned in Equation (31). The R-squared value for the regression analysis is lying around 0.9711 whereas the norm of residuals is around 0.4497.

$$y = -3.37x10^{-3}x + 15.39 \tag{31}$$

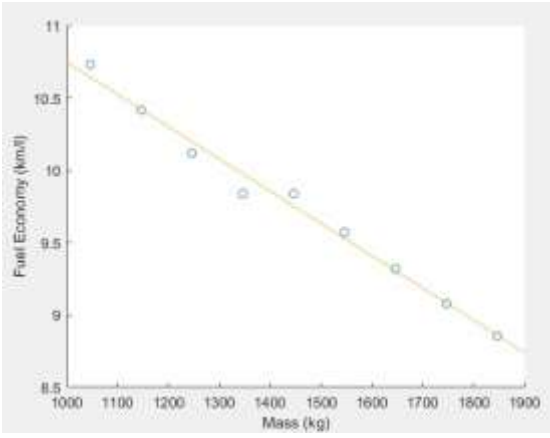


Fig 6. FE versus Vehicular Mass relationship under JP10 driving cycles.

The equation for the graph of JP10 driving cycle can be determined as mentioned in Equation (32) whereas its R-squared value can be estimated as 0.986 and the norm of residuals as 0.2057.

$$y = -2.23x10^{-3}x + 12.98 \quad (32)$$

From both the aforementioned equations, it can lucidly be seen that, fuel economy has a linear and an inverse relationship with the mass of the vehicle.

V. CONCLUSION

The study performs a single parametric modelling of the relationships between coefficient of rolling resistance and fuel economy, the coefficient of aero-dynamic drag and fuel economy and vehicular inertia and fuel economy and validates the developed models using FASTSim™ software simulations. The simulation has been carried out using two driving cycles, i.e., US06 driving cycle and JP10 driving cycle and subsequently the derived relationships have been validated. It can be concluded that the fuel economy portrays an inverse linear relationships w.r.t the variation of coefficient of rolling resistance, coefficient of aerodynamic drag and linear inertia. The said linear relationships can be seen since a single parametric study is conducted. It can be recommended that the coefficient of aerodynamic drag and vehicular mass should be optimized in vehicular design in order to obtain an enhanced fuel economy. Furthermore, the roughness of the terrain will also be impactful in terms of rolling resistance. Moreover, as a future study, attention should be provided towards the gradient of the road profile which too may provide an impact towards the operating fuel economy which is aspect which has not been appraised during the study since the driving cycle tests are conducted mainly on flat surfaces.

REFERENCES

[1] Sri Lanka Parliament, The *Motor Traffic Act 2009* (Section 122)

[2] Centre for Intelligent Transport Systems, "Feasibility Study and Market Research on Light Duty Electric Vehicles," University of Moratuwa, Sri Lanka, 2020.

[3] Iinternational Energy Agency, "Global EV Outlook 2019 - Scaling-up the transition to electric mobility," 2019.

[4] Engineering ToolBox, (2004). *Drag Coefficient*. [online] Available at:https://www.engineeringtoolbox.com/drag-coefficient-d_627.html [Accessed 10th November 2019].

[5] The Physics HyperText Book, (2019). *Aerodynamic Drag*. [online] Available at: https://physics.info/drag. [Accessed 11th November 2019].

[6] Kelly, K.B., and Holcombe, H.J. (1978). Aerodynamics for Body Engineers. Automotive Aerodynamics, Progress in Technology Series, 16 Society of Automotive Engineers.

[7] Wong, J. Y. (2001). *Theory of ground vehicles* (3rd ed., pp. 255–260). John Willey & Sons, Inc.

[8] Gough, V.E. (1971). Structure of the Tire. Mechanics of Pneumatic Tires. Monograph 122. Washington, DC: the National Bureau of Standards.

[9] Paterlini, G. (2015). Rolling Resistance Validation. [online] Available at: mndot.gov/research/TS/2015/201539.pdf. [Accessed on 13th November 2019].

[10] Bellman, M., Agarwal, R., Naber, J., and Chusak, L. (2010) Reducing energy consumption of ground vehicles by active flow control. In ASME 2010 4th International Conference on Energy Sustainability, pp 785- 793, American Society of Mechanical Engineers.

[11] Mayer, W., and Wickern, G. (2011) The new audi A6/A7 family-aerodynamic development of different body types on one platform, SAE International Journal of Passenger Cars-Mechanical Systems, Vol. 4(1), pp197-206.

[12] Ferguson, C., & Kirkpatrick, A. (1395). Internal Combustion Engines (3rd ed.). John Willey & Sons, Inc.

Field Deployable Additive Manufactured Housing for an Electro-optical System: A Case Study

Mark Holloway
Defence and Security Cluster, Optronic Sensor Systems
Council for Scientific and Industrial Research (CSIR)
Pretoria, South Africa
MHolloway@csir.co.za

Fernando Camisani-Calzolari
Defence and Security Cluster, Optronic Sensor Systems
Council for Scientific and Industrial Research (CSIR)
Pretoria, South Africa
FCamisaniCalzolari@csir.co.za

Hendrik Theron
Defence and Security Cluster, Optronic Sensor Systems
Council for Scientific and Industrial Research (CSIR)
Pretoria, South Africa
HTheron@csir.co.za

Abstract—The feasibility of the Additive Manufacture (AM) process for housings used on camera systems in outdoor applications is investigated. Aspects of the AM housing parts explored are stiffness-to-weight ratio and thermal transfer. The AM process of the housings allows the inclusion of functional features which are not easily achievable using traditional composite material manufacturing methods. A combination of strategically applied (to manage the final component weight) epoxy for sealing and composite fibre reinforcing was used to achieve the desired strength and durability. Comparative testing methods are used to evaluate the unprocessed AM and reinforced parts. These results are used to gauge the suitability of the post-processing materials and methods applied to the AM parts. In addition, the assembly process of the reinforced AM parts to the machined mechanical interface components was evaluated to determine whether this method of housing manufacture offered a viable alternative to traditional housing manufacture. A performance analysis of manufacturing methods is also presented. It was found that the AM process compared well to other manufacturing processes. Short-term field trials of the housings are planned to further evaluate the durability of the reinforced AM housing.

Keywords—additive manufacture, housing, composite reinforcing, strength

I. INTRODUCTION

Traditional design approaches for medium and long-range surveillance camera system housings, which are deployed in harsh environments, are based on machined or cast plate/profile aluminium assemblies (Fig. 11a) [1]. These assemblies typically require many fasteners to reduce warping between the plate interface junction surfaces where the environmental seals are located. Managing the effect of direct solar radiation, particularly in hot, cloudless environments, is achieved by using a separate sun-shield assembly fastened to the housing. This traditional approach is an effective and proven method of protecting the sensitive electro-optical hardware within the housing but is limiting in terms of form, cost, manufacture lead time, and weight.

An alternative to the aluminium assembly is a Glass Reinforced Plastic (GRP) [2] and/or Carbon Fibre Reinforced Polymer/Plastic (CFRP) [3] laminated composite housing (Fig. 11b) [4] [5]. Some benefits of using composites are more complex and tightly profiled shapes, significant weight saving, and integrated features such as sun-shields. The manufacture of composite housings is labour intensive and requires multiple sequential manufacturing steps; typically plug manufacture, mould manufacture, composite, and epoxy layup, trimming, and cementing of parts. Housing seal effectiveness can be highly dependent on component interface junction assembly quality. The interface junction of the machined or cast plate/profile aluminium assembly is inherently managed through CNC (Computer Numerical Control) machining of the parts and simple quality control metrology. Final composite housing assembly sealing, whether for single unit or small-scale production, relies on artisan skill. Manufacturing costs and lead times for the plug and moulds of complex housing profiles can be prohibitive, particularly with a single unit or small-scale production. The high stiffness and strength-to-weight ratio of composite materials make them particularly suitable for the manufacturing of housings.

Rapid advances in the field of Additive Manufacture (AM) and materials have allowed the application of the technology in previously unexploited fields [6]. This case study presents a hardware implementation of an AM composite reinforced housing used with a medium-range surveillance camera system in an outdoor application (Fig. 11c). Aspects of the AM housing parts explored are stiffness-to-weight ratio and thermal transfer. A complete AM composite reinforced housing has been manufactured for the TYTO medium-range surveillance camera System developed here at the CSIR (Fig. 11c).

II. DESIGN PHILOSOPHY

Fused Filament Fabrication (FFF) [7], more commonly referred to as “3D Printing” spans a user base from the hobbyist to the aerospace industry. Examples of housings, both cosmetic and functional, abound. The use of FFF in the industrial and scientific domains for permanent housings/structures has been sparse due to, amongst others, structural, material, and impermeability limitations of the technology [8].

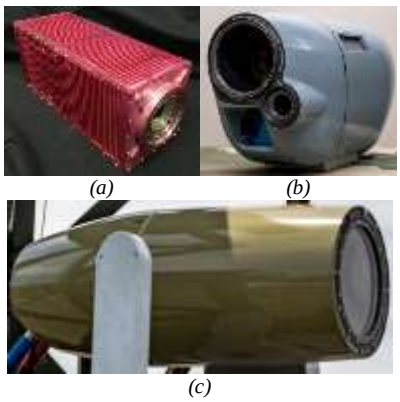


Fig. 11. (a) Traditional Plate Housing (DEROTATION). (b) Additive Manufactured Composite Reinforced Housing (TYTO). (c) Fibreglass/Carbon fibre Epoxy Laminated Composite Housing (RINO).

Selective Laser Sintering (SLS) [9] of powdered material with average grain sizes of 56 µm afford component structure dimensions not easily achievable with other 3D Printing or traditional laminated composite processes, while still retaining high strength and stiffness.

Utilisation of this manufacturing technique shifts its historical analysis and experimental nature into a real-world application.

The manufacture of niche market electro-optical housings, particularly for single unit or small-scale production, has historically been a slow, expensive, and labour-intensive process. Significant expenditure on Non-Recurring Engineering (NRE) cost is rarely recovered even when multiple components were manufactured.

The combination of existing AM manufacture, laminated composite, sealing, and interface processes are employed to produce a field deployable electro-optical system housing which addresses all the user requirements by reducing the operation count required to produce the component.

III. PROOF OF CONCEPT

An initial proof of concept, in the form of a side cover panel (core) was designed. The design comprises an inner skin which functions as the primary environmental barrier against contamination and moisture and together with machined interface flanges, an impermeable housing component. V-shaped ribs were added longitudinally to the outer face of the inner skin. These elements serve a dual purpose: imparting 1) longitudinal and transverse stiffness to the housing profile and 2) air channels for the forced cooling. Lastly, the outer skin is added to the profile, providing the substrate for the main structural layer of the housing which fulfils three functions: 1) a capping element for the air channels, 2) a thermal sun-shield against direct solar radiation, and 3) a mechanical surface resisting impact, abrasion, and the environment (Fig. *12*).

The housing side panel (core) was manufactured using the SLS Additive Manufacturing process on an EOS P 396 industrial 3D printer from EOS PA2200 (PA 12 Nylon Polyamide Polymer) material.

In the as-manufactured state, the component (core) is not suitable for extended exposure to a harsh environment and does not possess inherent structural strength. Post-processing of the component included dipping it in Axson EPOLAM 2022 Laminating Epoxy Resin and allowing the excess to drain off. Once cured, a varied number of twill weave e-glass (163 g/m²) layers were strategically positioned on the housing to achieve the desired stiffness and strength (Fig. 13). The purpose of this process is to understand the effects of sealing and composite processes on the cost, complexity, dimensions, weight and ultimately strength of the part (Fig. 14).

Results from this experimentation were used to select an optimised set of post-AM processes. The dominant challenge experienced was the identification of the optimal combination of 1) the sealing method, 2) the simplest placement of laminate composite, and 3) the retention of important dimensional features while achieving the best stiffness and strength-to-weight ratio at the lowest cost.

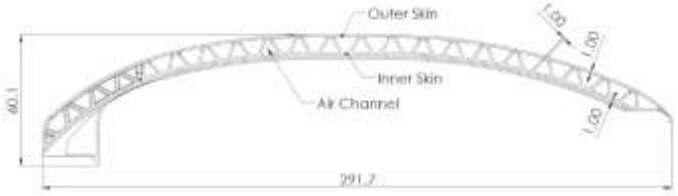


Fig. 12. TYTO Housing - Side Panel.

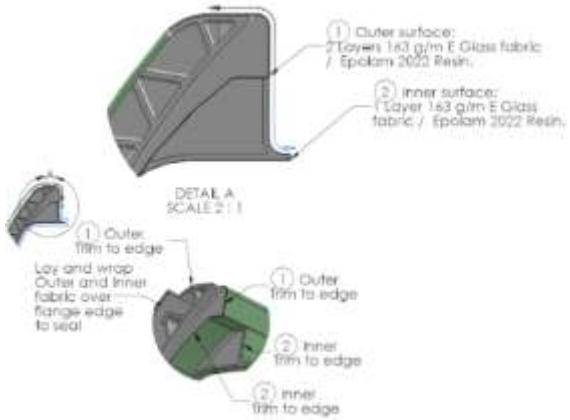


Fig. 13. Example of structural composite material application details.

A. *Sample Construction*

In this study, segment profiles for each of the samples were manufactured to be geometrically consistent with the housing application it represents. Four samples were prepared, namely, an aluminium stack (ALUSTD) (Fig. 15a and Fig. 16a) consisting of a machined pocket flat housing plate (EN AW 6082-T6) and sun-shield (EN AW 5754-H22) combined using fasteners. The GRP profile (Fig. 15b and Fig. 16b) was constructed from 15 layers of 163 g/m² glass fibre fabric, LR30 resin and LH30 hardener using a vacuum bag process. An as-manufactured SLS AM PA2200 Nylon Standard (AMSTD) segment (Fig. 15c and

DESCRIPTION	PROCESS	TARGET WEIGHT	MEASURED WEIGHT	NOTES
Component	SLS 3D Printed - PA 2200	170g		Clean component using compressed air (moderate pressure) to remove excess powder.
Coating	Brush on EPOLAM 2022 on external surfaces	175g		Ensure complete coverage of all internal surfaces. Hang with internal ribs in a vertical orientation to allow excess epoxy to flow off the component.
Lay-up - Outer surface and edge flanges	2 x layers 163 g/m² E Glass fabric and EPOLAM 2022	-	-	MAXIMUM layer build-up on offset surface 0.3mm. Fabric to wrap over top and bottom edges as per drawing detail. Flange outer and inner edges to encapsulated with fabric and epoxy to seal edge.
Lay-up - Inner surface and vent port	1 x Layer 163 g/m² E Glass fabric and EPOLAM 2022	-	-	Fabric to wrap up to top and bottom edges as per drawing detail. Vent port inner faces to be covered with fabric as per drawing and where accessible.
Post Curing	As per EPOLAM 2022 material data sheet up to a MAXIMUM of 50 °C	-	-	Hang component with internal ribs in a vertical orientation to MINIMISE core deformation / creep during 80 °C post cure.
Trim	Trim flange edges to specification.	240 g		
Primer	TBD	-		Spray inner surface with minimum primer layer thickness for environmental protection. Finish outer surface with featherlite spray filler, as necessary. Spray outer face with minimum primer layer thickness required as a key coat for the top coat.

Fig. 14. Data recording table.

Fig. 16c) and an SLS AM PA2200 Nylon + GRP composite reinforced (AMGRP) segment (Fig. 15d and Fig. 16d), with two layers on the outer skin and a single layer on the inner skin, of 163 g/m² twill weave e-glass and Axson EPOLAM 2022 Laminating Epoxy Resin. The samples were manufactured with two geometries, 250 x 50 mm for the three-point bending test, and the dimensional stability test and 120 x 120mm for the thermal insulation test.

IV. EXPERIMENTAL WORK

A limited number of mechanical tests were performed on selected sections of the housing manufacturing techniques to compare the effectiveness and viability of the AM composite reinforced housing against established traditional housing manufacturing methods. An exhaustive evaluation and characterization of the AM composite reinforced manufacturing is not part of this case study.

A. Stiffness-to-Weight-Ratio

Man-portable electro-optical systems, by definition, need to be lightweight and robust. As the housing typically functions as the handling surface for transfer to and deploying / recovery of the system on-site, it is required to be suitably stiff and lightweight.

To characterise the different manufacturing techniques, a non-destructive three-point bending test was conducted on each of the samples to determine the stiffness of the structure. The purpose of determining the stiffness, for this study, is to calculate a comparative stiffness-to-weight ratio rather than the flexural modulus of sample-specific manufacturing method. This metric is used to rank the housing manufacturing technique for inclusion in the key performance analysis.

1) *Experimental Setup*

The test setup (Fig. 15e) consisted of an extruded square beam, measuring 50 x 50mm, supported between two blocks. This beam forms the base for the lower support pins at a span of 224mm. Cylindrical support pins were used with the aluminium stack sample while rectangular support pins were employed for the rest of the samples. The applied load, of 35.9N, is applied to the sample by weights suspended from a y-shaped yoke, crossbar, and cylindrical loading pin. A Heidenhain MT30 digital linear probe and ND 221B Measured

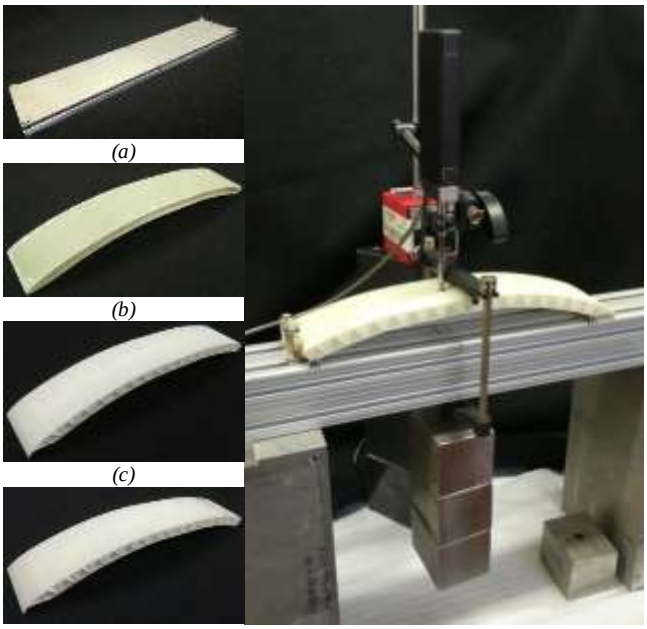


Fig. 15. Three point bending samples and experimental setup.

Value Display are used to measure the deflection. The maximum deflection was recorded for each of the ten loading cycles. A bending stiffness can be calculated using

$$K = \frac{p}{w}$$

where K is the bending stiffness [$N.mm^{-1}$], p is the applied force [N] and w is the deflection [mm].

The weight [W] in grams [g] of the samples was measured using a Mettler K4T precision balance. The stiffness-to-weight ratio, expressed in [$N.mm^{-1}.g^{-1}$], is calculated from the bending stiffness K and weight W.

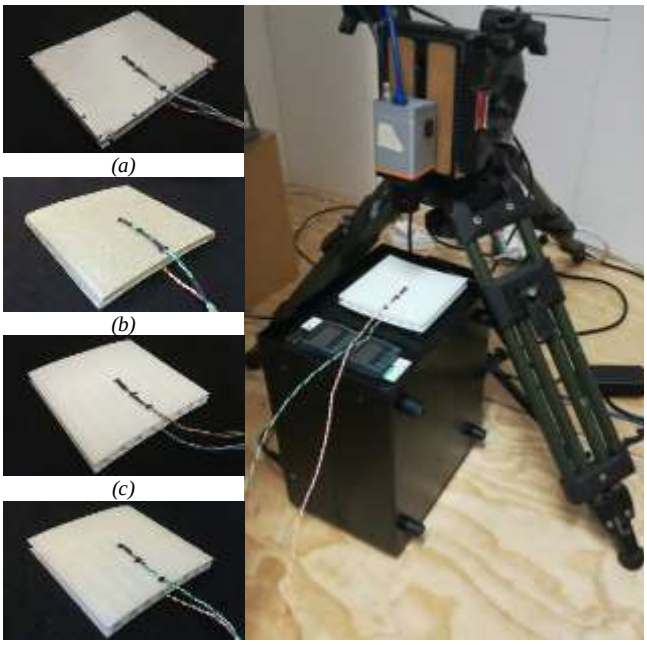
B. Thermal Transfer

Field deployable electro-optical surveillance systems are typically deployed at sites which are completely exposed to the elements and the housings are subject to direct solar radiation. Some systems are equipped with active cooling of the sensitive internal hardware, but the majority rely on natural convection. Therefore, minimising the convection transfer of this thermal load is a key performance requirement of the housing.

1) *Experimental Setup*

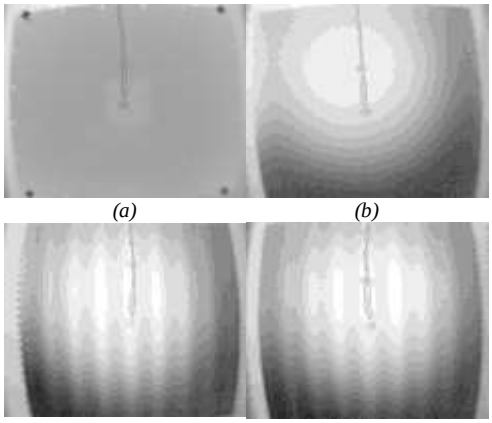
The experimental setup (Fig. 16e) consists of a Thermoteknix ThermaRef cavity black body orientated vertically and placed inside an insulated chamber equipped with a circulation fan and set at T = 333 K (60 °C) producing a radiant exitance of $M_e(T) = \sigma_e T^4 \Rightarrow M_e(333) = 697$ W.m^{-2}, where $\sigma_e = 5.6697 \times 10^{-8}$ W.m^{-2}.K^{-4} is the Stefan–Boltzmann constant [10].

A Xenix Gobi-640 un-cooled thermal camera was placed above the black body cavity and is used to observe the inner skin of the housing during the experiment. The thermal image was used to visualise the thermal transfer and observe any significant



(a) *(b)* *(c)* *(d)* *(e)*

Fig. 16. Thermal transfer samples and experimental setup.



(a) *(b)* *(c)* *(d)*

Fig. 17. Thermal transfer visualisation images.

gradients or localised regions during the heating cycle (Fig. 17a ALUSTD, Fig. 17b GRP, Fig. 17c AMSTD, Fig. 17d AMGRP). Analysing the gradients and their effect is not part of this study. Two thermocouples, one placed adjacent to the thermal camera and the second suspended from the enclosure ceiling, monitor the ambient temperature.

Each of the samples is instrumented with a 100 kΩ bead thermistor on the outer and inner skins/surface to measure the temperature gradient across the sample. Once the black body had achieved thermal equilibrium, the samples were placed on four domed nuts located around the black body aperture. The outer and inner skin thermistors as well as the thermal camera thermocouple were connected to an HP 3490A Data Acquisition / Switch unit. Data logging of the signals, at 30-second intervals, was performed on a laptop using Agilent BenchLink Data Logger 3 software.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Stiffness to Weight Ratio

TABLE 11 shows the deflections due to the applied load for the manufactured samples.

The comparative stiffness-to-weight ratio results are shown in Fig. 18. A higher ratio is desirable and the AMGRP sample produced the highest stiffness to weight ratio of all the samples with the intended design geometry for each of the electro-optical housing types. Significant hysteresis and creep were noted during the testing of the AMSTD sample. These properties do not form part of this study and can be addressed in future research.

B. Thermal Transfer

1) *Thermal Dynamic Modelling*

The thermal process is modelled from the source to the chamber temperature using a cascade approach. See Fig. 19.

The source q_s at 60 °C produces a radiant exitance of 697 $W.m^{-2}$ to cause a temperature change of the outer shell thermistor ($T_o(s)$) through the transfer function $g_{so}(s)$. Next, the outer shell temperature produces an inner shell temperature time series $T_i(s)$ through $g_{oi}(s)$. Finally, $T_i(s)$ produces the chamber temperature $T_c(s)$ through $g_{ic}(s)$. All temperatures are in °C and the source was kept constant for all experiments.

The form of the three transfer functions was selected based on the response of the respective curves of the thermal data and each was found to fit (using system identification) the structure in Fig. 20 well.

The structure comprises three first-order transfer functions in parallel followed by a low pass filter.

TABLE 11. DEFLECTION OF MANUFACTURED SAMPLES DUE TO THE APPLIED LOAD

Sample	Test Point									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Deflection in mm									
ALUSTD	0.469	0.447	0.447	0.441	0.444	0.443	0.446	0.446	0.442	0.434
GRP	2.474	2.471	2.447	2.422	2.433	2.440	2.441	2.430	2.433	2.424
AMSTD.	2.482	2.466	2.462	2.515	2.492	2.469	2.510	2.514	2.523	2.541
AMGRP	**0.540**	**0.525**	**0.522**	**0.523**	**0.521**	**0.523**	**0.524**	**0.541**	**0.534**	**0.522**

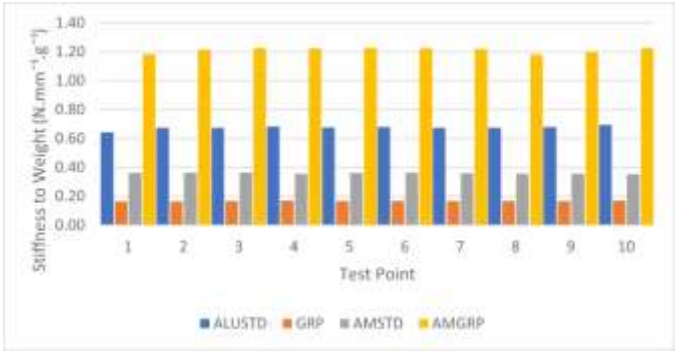


Fig. 18.Stiffness-to-weight ratio of manufactured samples.

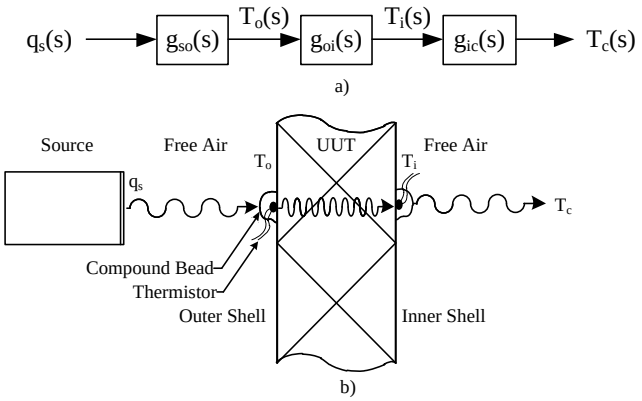


Fig. 19. Thermal modelling of the experimental process: a) transfer function description and b) description of the experimental setup. UUT – Unit Under Test.

The time constants [s] of the parallel transfer functions are typically such that $\tau_{p_1} \ll \tau_{p_2} \ll \tau_{p_3}$, and are assumed to be a combination of conduction, convection, and radiation interactions at the surfaces and within the materials.

For this work, the heat transfer through the outer shell to the inner shell ($g_{oi}(s)$) is assumed to be the dominant feature to thermally evaluate the performance of each manufacturing method, and the heat transfer from the source to the outer shell and the heat transfer from the inner shell to the chamber has been omitted here. These models have been fitted to the data gathered from the procedure in § IV.B.1). Fig. 21 shows input-output data, from one experiment, with the outer shell thermistor ($T_o(t)$) as the input and the inner shell thermistor ($T_i(t)$) as the output for the four manufacturing methods.

Fig. 22 shows experimental outputs and model outputs for the four manufacturing methods, where the outer shell temperature ($T_o(t)$ as in Fig. 21) was used as the input to the model, and all the responses were offset to start at 0 °C for comparison. The figure shows that the models compare well with the experimental outputs.

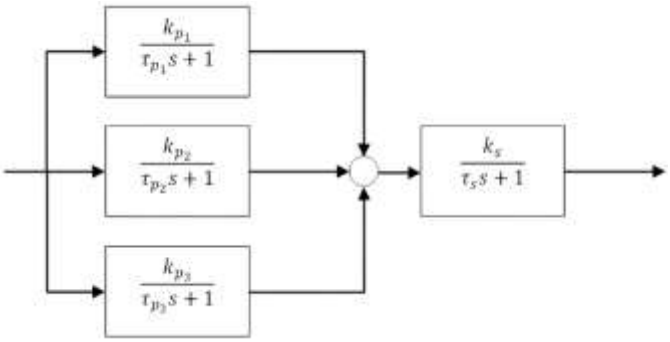


Fig. 20. Expanded structure for each of the models $g_{so}(s)$, $g_{oi}(s)$, and $g_{ic}(s)$.

Table 12 summarizes the fitted parameters for the models (in Fig. 20) for $g_{oi}(s)$ of each of the manufacturing methods. Note that the "fast" poles are zero, indicating a pure gain for the first parallel transfer function, and that the orders of magnitude of the time constant for the medium and slow poles are similar, respectively. It also shows the initial temperature at which the time series start for each respective manufacturing method.

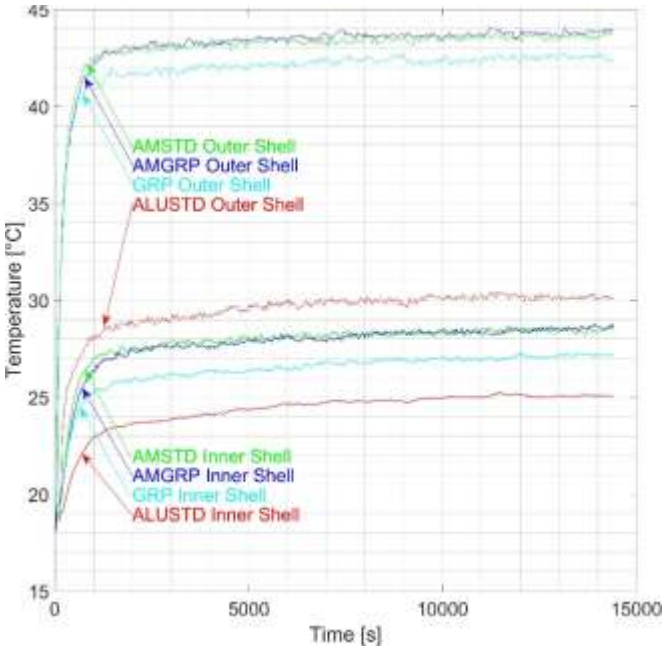


Fig. 21: Experimental measurement data showing time series of the outer shell temperatures and the inner shell temperatures for the different manufacturing methods (red - ALUSTD, green - AMSTD, blue - AMGRP, and cyan – GRP).

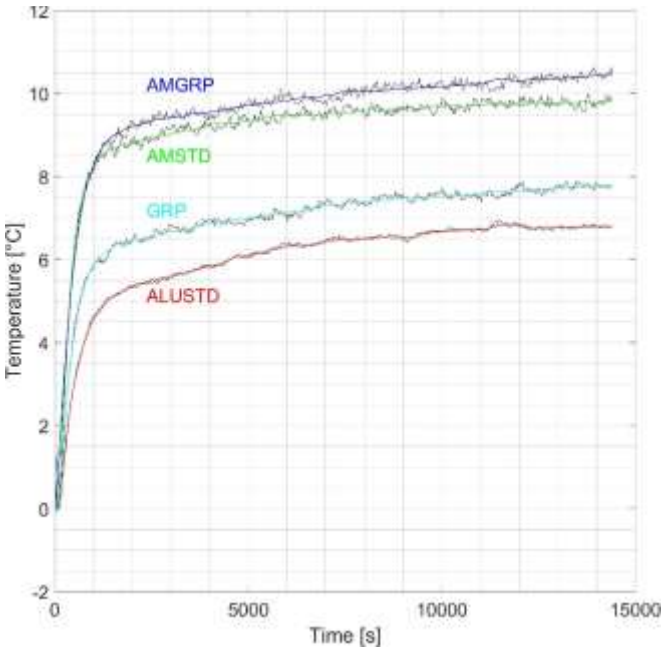


Fig. 22: Time series showing the measured outputs ($T_i(t)$, black lines) and model outputs (red - ALUSTD, green - AMSTD, blue - AMGRP, and cyan – GRP) for $g_{oi}(s)$ of the four manufacturing methods.

The parallel transfer functions of the model make it difficult to quantify a comparison between the manufacturing methods. Rather, the gain and bandwidth of the overall model in Fig. 20 should be considered.

Table 13 is an evaluation of the steady-state and dynamic thermal properties of the different materials, based on the fitted

models. The reciprocals of some quantities have been given - because of the slow nature of the process – for readability. The bandwidth is an indication of the transient nature of the process. The smaller this value, the slower the system will initially transfer energy from the outer to the inner shell, which is desired. The gain is an indication of the steady-state thermal behavior from the outer to the inner shell and a low as expected, while the others have significantly higher and similar ΔT. The table further shows that GRP exhibits both the best (slowest) transient response as well as the lowest steady- state performance. The AMGRP method has the second-best transient response and the AMSTD is second in steady-state performance. value is desired. The change in temperature from the outer to the inner shells at steady state (ΔT), the poles, and the zeros have been added for completeness. The ALUSTD has the lowest ΔT.

TABLE 12. IDENTIFIED PARAMETERS FOR $G_{OI}(S)$ FOR EACH OF THE MANUFACTURING METHODS. "-" INDICATES NO ADDITIONAL POLE.

Method	**Initial Temperature**	**Fast Pole**		**Medium Pole**		**Slow Pole**		**Low Pass Filter**	
Parameter	T(t=0) [°C]	k_{p_1} [°C/°C]	τ_{p_1} [s]	k_{p_2} [°C/°C]	τ_{p_2} [s]	k_{p_3} [°C/°C]	τ_{p_3} [s]	k_s [°C/°C]	τ_s [s]
ALUSTD	18.3	1.081	-	5.308	220	1.269	4649	0.0763	-
GRP	19.4	-	-	11.79	227	3.127	7209	0.0228	-
AMSTD	18.7	1.195	-	9.055	151	1.378	3798	0.0338	49
AMGRP	**18.2**	**0.757**	**-**	**3.875**	**255**	**0.761**	**9999**	**0.0789**	**-**

TABLE 13: COMPARISON OF THE DIFFERENT METHODS. "-" INDICATES NO ADDITIONAL POLE OR ZERO.

Method	**-3 dB Bandwidth [rad/s]**	**Gain [°C/°C]**	**ΔT [°C]**	**Poles [rad/s]**			**Zeros [rad/s]**	
ALUSTD	$\frac{1}{335}$	0.584	5.2	-	$\frac{-1}{220}$	$\frac{-1}{4649}$	$\frac{-1}{37}$	$\frac{-1}{3909}$
GRP	$\frac{1}{437}$	0.340	15.5	$\frac{-1}{227}$	-	$\frac{-1}{7209}$	-	$\frac{-1}{5745}$
AMSTD	$\frac{1}{214}$	0.393	15.1	$\frac{-1}{49}$	$\frac{-1}{151}$	$\frac{-1}{3768}$	$\frac{-1}{18}$	$\frac{-1}{3337}$
AMGRP	$\mathbf{\frac{1}{359}}$	**0.425**	**15.3**	-	$\mathbf{\frac{-1}{255}}$	$\mathbf{\frac{-1}{9998}}$	$\mathbf{\frac{-1}{41}}$	$\mathbf{\frac{-1}{8619}}$

VI. PERFORMANCE ANALYSIS

A set of criteria were established to compare the requirement performance of the manufacturing techniques (

TABLE 14) when applied to single or small-scale production of electro-optical system housings. Examples of existing electro-optical housings were used (Fig. 23).

These criteria are.

- Lighter and more compact surveillance system.
- Cost-effective system for a special application of a single unit or small-scale production run
- Flexible system housing to cater for a system upgrade or inclusion of newer sensor technology.
- Ability to include complex geometry and features.
- More reliable environmental protection for high-value sensitive electro-optical hardware.
- Shorter lead time from concept to housing hardware for a single unit or small-scale production run.
- Semi-skilled personnel can be utilised for composite laminate application.
- Simpler and more robust component interface junction assembly.
- Reduced NRE cost to manufacture housing hardware for a single unit or small-scale production run.
- Less reliance on labour for housing manufacture.

TABLE 14. PERFORMANCE ANALYSIS

	System		
	TYTO	***RINO***	***DEROTATION***
FIGURE OF MERIT	**8.0**	**73.6**	**151.6**
	All results are normalised to the TYTO baseline.		
Construction Density (1)	1.0	1.6	2.4
Internal Volume Utilisation (2)	1.0	1.6	1.5
Cost (3)	1.0	1.2(12)	0.6
Form and Fit Flexibility (4)	1.0	16.0	69.2
Complex Geometry (5)	1.0	40.0	69.2
Operational Reliability / Durability (6)	✘	✓	✘
Lead Time (7)	1.0	4.3	5.1
Manufacturing Skill (8)	1.0	1.1	1.2
Junction Interface (9)	0	6	1
Non-Recuring Engineering (NRE) (10)	0	kRand 104	0
Manufacture Automation (11)	1.0	3.0	1.5

(1) Ratio of Housing Mass to Volume.
(2) Ratio of Inner Volume Utilisation. Internal Component Volume to Inner Housing Volume.
(3) Actual Costs for Housing Component Manufacture.
(4) Function of Production Means (plugs, moulds, fixtures, jigs, and finishing) and geometry.
(5) Assessment of Part Complexity, Table 1 in [11]
(6) System Deployed in an Operational Environment.
(7) Actual Lead Time for Housing Component Manufacture.
(8) Manufacturing Skill Level Required for Manufacture Based on International Standard Classification of Occupations (ISCO-88) [12].
(9) Function of Number of Adjustments, Secondary Assembly Processes, Post-Assembly Processes Required.
(10) Cost of NRE Investment Required to Manufacture a Single Unit or Small-Scale Production Run.
(11) A Measure of the 'Levels of Automation' (LoA) Required to Manufacture the Housing [13].
(12) Metric Based on the Manufacture of 3 RINO Housings.

(a) *(b)* *(c)*

Fig. 23. (a) TYTO. (b) RINO. (c) DEROTATION.

VII. CONCLUSION

Building on the proof-of-concept manufacture technique, a complete field deployable Additive Manufactured composite reinforced housing for the TYTO electro-optical system was manufactured. From a requirement performance perspective, this manufacturing technique offers a viable alternative to traditional housings. The experimental results of selected properties have shown that the AMGRP technique provides the best stiffness-to-weight ratio. Results of the thermal transfer tests indicate further investigation into the construction and material selection of the AMGRP is necessary to compete effectively with the traditional ALUSTD housing.

REFERENCES

[1] "OBZERV Land Systems ARGC-750," [Online]. Available: http://www.obzerv.com/products/land-systems/argc-750/. [Accessed 04 Aug 2022].

[2] "Amiblu Glassfiber reinforced plastics (GRP)," [Online]. Available: https://www.amiblu.com/why-grp/. [Accessed 04 Aug 2022].

[3] "sgl carbon Carbon Fibers and Carbon Fiber-Reinforced Plastic (CFRP)," [Online]. Available: https://www.sglcarbon.com/en/carbon-fibers-and-cfrp/. [Accessed 02 Aug 2022].

[4] "Graflex CARBON FIBER ENCLOSURES," [Online]. Available: https://graflex.com/products/motorized-zoom-lenses/options/carbon-fiber-enclosures/. [Accessed 04 Aug 2022].

[5] "ACS Australia Aircraft Aerial Imaging Carbon Fibre Camera Housing," [Online]. Available: https://acs-aus.com/our-work/aircraft-aerial-imaging-carbon-fibre-camera-housing/. [Accessed 04 Aug 2022].

[6] "zortrax Getting Rid of Creative Restrictions with Zortrax 3D Printing – Gates Underwater," [Online]. Available: https://zortrax.com/blog/gates-underwater-products-and-zortrax-3d-printing/. [Accessed 04 Aug 2022].

[7] "All3DP Fused Filament Fabrication – Simply Explained," [Online]. Available: https://all3dp.com/2/fused-filament-fabrication-fff-3d-printing-simply-explained/. [Accessed 04 Aug 2022].

[8] S. Singh, G. Singh, C. Prakash and S. Ramakrishna, "Current status and future directions of fused filament fabrication," *Journal of Manufacturing Processes*, vol. 55, no. April, p. 288–306, 2020.

[9] "eos EOS SLS Techology for Plastics 3D Printing SLS 3D Printer for Additive Manufacturing," [Online]. Available: https://www.eos.info/en/industrial-3d-printing/additive-manufacturing-how-it-works/sls-3d-printing. [Accessed 04 Aug 2022].

[10] A. F. Mills, Heat Transfer, International Student ed., Burr Ridge, Illinois: Irwin, 1992.

[11] L. J. P. de Araújo, J. Atkin, Baumers and M. Baumers, "A part complexity measurement method supporting 3D Printing," in *The 32nd International Conference on Printing for Fabrication 2016 (NIP32)*, Manchester, UK, January 2016.

[12] "Intenational Labour Organization ISCO-08," [Online]. Available: https://isco-ilo.netlify.app/en/isco-08/. [Accessed 04 Aug 2022].

[13] "MAKINO The Roadmap to the Five Levels of Manufacturing Automation," [Online]. Available: https://www.makino.com/en-us/resources/content-library/white-papers/the-roadmap-to-the-five-levels-of-manufacturing. [Accessed 04 Aug 2022].

Recursive Least Squares Estimation of Battery Charge Capacity and State of Charge

Saeid Bashash
Department of Mechanical Engineering
San Jose State University
San Jose, CA, USA
saeid.bashash@sjsu.edu

***Abstract*—This paper presents a method for simultaneous state of charge and charge capacity estimation in electrochemical batteries using real-time current and voltage measurements. The method is based on the recursive least squares optimization process applied to an equivalent circuit battery model. The model is identified using pulse charge and discharge cycle data collected from a lithium iron phosphate polymer battery cell. During online estimation, all the parameters of the model are fixed to the identified values except for the initial SOC and the charge capacity, which are adapted in real-time using a set of compact parameter update laws. Test results indicate a highly accurate SOC and charge capacity estimation using the proposed method despite the simplicity of the model and the estimation algorithm. In addition, the estimator maintains a reasonable level of robustness with respect to uncertainty in the model parameters.**

Keywords—Battery state of charge estimation, battery state of health estimation, Recursive least squares optimization

I. INTRODUCTION

This paper presents a method for battery state of charge (SOC) and charge capacity estimation using real-time current and voltage measurements. Estimating battery SOC and charge capacity is a critical step toward the optimal operation of battery-powered systems. Many battery management algorithms use a fixed charge capacity value to estimate SOC, which can result in a drift in the accuracy of estimation over time due to battery capacity fade. This paper provides a method for simultaneous battery SOC and charge capacity estimation using recursive least squares optimization.

Electrochemical batteries play a central role in the advancement of consumer electronics and modern energy systems such as electric vehicles and microgrids. For an accurate estimation and prediction of battery's SOC and state of health (SOH), a dynamic model is usually employed and simulated in real-time. The most common battery modeling methods are the equivalent circuit representation [1, 2] and the first principles electrochemical modeling [3, 4]. Other methods such as machine learning techniques have also been used to model and predict battery health and performance [5].

SOC estimation methods can be classified into three methods: *conventional*, *model-based*, and *data-driven* [6]. For the conventional methods, coulomb counting [7] and open-circuit voltage (OCV)-based estimation [8] are the most commonly used methods. In these methods, simple calculations are run on the current or voltage measurement or both to determine the SOC of the battery. The performance of the coulomb counting process is limited by the inaccuracy of the initial SOC guess, capacity of the battery, and the current sensor bias. Therefore, it can only be used under a limited set of conditions, mainly for battery characterization purposes. The voltage-based estimation method can be used during rest periods, but its performance can be reduced under dynamic loading conditions.

Model-based estimation methods combine the benefits of both current and voltage measurements using a closed-loop state observer such as a Kalman filter [9, 10], a nonlinear adaptive state observer [11, 12] a sliding mode observer [13, 14], or a recursive least squares estimator [15, 16]. These methods involve coulomb counting with a voltage correction component which accounts for dynamic effects using a battery model. The most commonly used method for model-based estimation is Kalman filtering. Various Kalman filter formulations have been used for battery state estimation, which include the extended Kalman filter [9], sigma-point Kalman filtering [17], and the unscented Kalman filter [18], among others. While Kalman filter provides a powerful method for battery state estimation, its limitation lies in the assumption that statistics of the noise and uncertainties follow Gaussian distributions. Therefore, the optimality of the Kalman filter can be significantly affected by non-Gaussian errors such as parametric uncertainties or modeling errors.

Data-driven models such as neural networks [19, 20] and support vector machines [21, 22] have been used for battery SOC estimation as well. These methods can capture the complex nonlinearities of batteries using a highly parametrized mathematical function/network. However, they require significant amount of data to train such networks, and may lack robustness with respect to change of conditions, thereby requiring even more data to cover all the possible testing conditions and scenarios during training of the model.

One of the main weaknesses of the model-based battery SOC estimation algorithms is their dependency on an accurate charge capacity value. As a battery undergoes successive

charging and discharging cycles, its capacity to store energy degrades, thereby reducing the accuracy of SOC estimation. As a result, estimating charge capacity has been one of the key factors in maintaining the performance of the SOC estimation algorithm in the long run. Estimation methods based on Kalman filter [9], recursive least-squares estimation [16], and machine learning techniques [23] have been developed and used to estimate the battery's charge capacity along with SOC.

In this paper, we develop and experimentally validate a recursive least squares estimation algorithm for battery SOC and charge capacity estimation. This method, which is based on a previous work [16], presents a simplified and real-time-friendly version of the algorithm for embedded BMS applications with limited computational power. Experimental tests indicate the robustness of the algorithm with respect to perturbations in the model parameters.

II. EQUIVALENT CIRCUIT BATTERY MODEL

A battery cell or pack can be modeled as a voltage source in series with an internal resistance, R_0, and a resistor-capacitor pair as shown in Fig. 1.

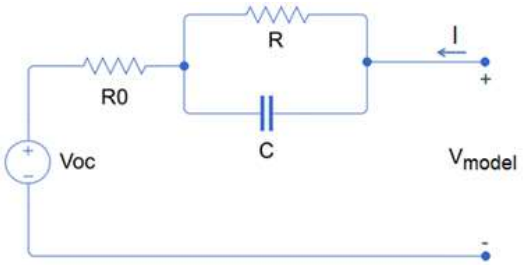


Fig. 1. Equivalent circuit battery model.

The equations of the circuit are given by [16]:

$$\begin{cases} \dot{s} = \dfrac{1}{Q} I, & s(0) = s_0 \\ \dot{V}_r = -\dfrac{1}{RC} V_r + \dfrac{1}{C} I, & V_r(0) = V_{r0} \end{cases} \tag{1}$$

where s is the state-of-charge (SOC), Q is the charge capacity, I is the applied current, and V_r is the voltage across the resistor-capacitor pair. This variable represents the relaxation (or the diffusion) voltage in the battery. In this paper, we only use one RC pair to represent the relaxation voltage for simplicity, but more RC pairs can be included to improve the model's accuracy without the loss of generality.

The battery terminal voltage is given by:

$$V_{model} = V_{oc}(s) + R_0 I + V_r \tag{2}$$

where V_{oc} is the SOC-dependent open-circuit voltage, which can be written as a polynomial function of SOC as follows:

$$V_{oc}(s) = a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n \tag{3}$$

where *a's* are known constant parameters identified for the battery cell through an experimental characterization process.

III. MODEL REVISION FOR SOC AND CHARGE CAPACITY ESTIMATION

The equations of the system represented by Eq. (1) in the previous section can be re-written as:

$$\begin{cases} \dot{x} = I, & x(0) = 0 \\ \dot{z} = -\dfrac{1}{T} z + \dfrac{1}{T} I, & z(0) = z_0 \end{cases} \tag{4}$$

where

$$x = Q(s - s_0), \qquad z = \frac{V_r}{R}, \qquad T = RC \tag{5}$$

The terminal voltage of the revised model is given by:

$$V_{model} = V_{oc}(s) + R_0 I + Rz \tag{6}$$

At a given SOC estimate, $\hat{s}$, the OCV function can be linearized as follows:

$$V_{oc}(s) \cong V_0 + C_0 s \tag{7}$$

where

$$C_0 = \left.\frac{\partial V_{oc}}{\partial s}\right|_{@s=\hat{s}} = a_1 + 2a_2\hat{s} + \cdots + na_n\hat{s}^{n-1} \tag{8}$$

$$V_0 = V_{oc}(\hat{s}) - C_0\hat{s} = a_0 - a_2\hat{s}^2 - \cdots - (1-n)a_n\hat{s}^n \tag{9}$$

We can re-write the linearized OCV equation as:

$$V_{oc}(s) \cong V_0 + C_0(s_0 + \alpha x), \text{ where } \alpha = \frac{1}{Q} \tag{10}$$

At a given time, the battery terminal voltage can then be written as:

$$V_{model}(t) \cong V_0(t) + s_0 C_0(t) + \alpha C_0(t) x(t) + R_0 I(t) + Rz(t) \tag{11}$$

The unknown parameters to be estimated through the proposed method in this paper include: $\underline{p} = \{s_0, \alpha\}$. In the next section, we derive the formulas for two cases:

- SOC estimation only: For this case, parameters $\alpha, R_0, R,$ and T must be made known to the estimator. By actively estimating s_0, the rest of the SOC trajectory can be estimated by a simple integration and scaling of the current according to the modified equations of the system.
- SOC and charge capacity estimation: In this case, parameters s_0 and α are estimated simultaneously, assuming R_0, R and T are known.

IV. RECURSIVE LEAST SQUARES ESTIMATION

The goal of the estimation algorithms in this effort is to minimize the integral of the difference between the measured

voltage and the model voltage, with a forgetting factor, as follows:

$$\underset{\underline{p}}{\text{minimize}}\ J(t) = \frac{1}{2}\int_0^t \lambda^{(t-\tau)}\big(V_{model}(\tau) - V(\tau)\big)^2 d\tau \quad (12)$$

where λ is a forgetting factor to be set very close to 1, e.g., λ = 0.99999. With this value, the weight of the data from one hour, one day, and one week prior to the current time is 96.5%, 42.1%, and 0.2%, respectively, compared to the weight of the most recent measurement. It is recommended this value to be chosen between 0.9999 and 0.999999 if the unit of time used is in seconds. If the forgetting factor is excluded, the estimator may not be able to stay up-to-date with battery's aging.

The cost function in Eq. (12) can be locally minimized using the gradient descent optimization, as follows:

$$\dot{\hat{p}}_i(t) = -k_i \frac{\partial J}{\partial p_i}\Big|_{@\underline{p}=\underline{\hat{p}}(t)}, \qquad i = 1,2 \quad (13)$$

where $\hat{p}_i$ is the estimate of the i^{th} unknown parameter within $\underline{p}$, the unknown parameter array, and k_i is the parameter adaption coefficient.

A. SOC Estimation Only

In this section, the formulas for the SOC estimation are derived assuming $\underline{p} = \{s_0\}$. All the other parameters are assumed to be known to the estimator. Using the linearized OCV equation, the cost function can be written as:

$$J(t) = \frac{1}{2}\int_0^t \lambda^{(t-\tau)}\Big(\big(V_0(\tau) + s_0 C_0(\tau) + \alpha C_0(\tau)x(\tau) + R_0 I(\tau) + Rz(\tau)\big) - V(\tau)\Big)^2 d\tau \quad (14)$$

The derivative of the cost function with respect to s_0 can be calculated as:

$$\frac{\partial J}{\partial s_o} = s_0\int_0^t \lambda^{(t-\tau)} C_0^2(\tau)d\tau + \int_0^t \lambda^{t-\tau} C_0(\tau)\begin{pmatrix} V_0(\tau) + \alpha C_0(\tau)x(\tau) + \\ R_0 I(\tau) + Rz(\tau) - V(\tau)\end{pmatrix} d\tau \quad (15)$$

The parameter update law is then derived as:

$$\dot{\hat{s}}_0(t) = -k\big(r_1(t) + \hat{s}_0(t)r_2(t)\big) \quad (16)$$

where r_1 and r_2 (known as regressors) are given by:

$$\begin{cases} r_1(t) = \int_0^t \lambda^{(t-\tau)} C_0(\tau)\begin{pmatrix} V_0(\tau) + \alpha C_0(\tau)x(\tau) + \\ R_0 I(\tau) + Rz(\tau) - V(\tau)\end{pmatrix} d\tau \\ r_2(t) = \int_0^t \lambda^{(t-\tau)} C_0^2(\tau)d\tau \end{cases} \quad (17)$$

As the initial SOC is computed through Eq. (16), the SOC estimate can be obtained from Eq. (5), as follows:

$$\hat{s}(t) = \hat{s}_0(t) + \alpha x(t) \quad (18)$$

For the implementation of the algorithm in an embedded system, the revised battery model (i.e., Eq. (4)), the parameter update law (i.e., Eq. (16)) and the regressors (i.e., Eq. (17)) must be discretized using a discretization method of choice with a sufficient accuracy.

B. SOC and Charge Capacity Estimation

In this section, the formulas for the simultaneous SOC and charge capacity estimation will be derived, for the unknown parameter array $\underline{p} = \{s_0, \alpha\}$. The rest of the model parameters, i.e., R_0, R, T and the OCV curve coefficients are assumed to be known to the estimator.

Applying the same procedure as the previous section to the unknown parameter array results in the following parameter update laws and the regressor equations:

$$\begin{cases} \dot{\hat{s}}_0(t) = -k_1\big(r_1(t) + \hat{s}_0(t)r_2(t) + \hat{\alpha}(t)r_3(t)\big) \\ \dot{\hat{\alpha}}(t) = -k_2\big(r_4(t) + \hat{s}_0(t)r_3(t) + \hat{\alpha}(t)r_5(t)\big) \end{cases} \quad (19)$$

where

$$\begin{cases} r_1(t) = \int_0^t \lambda^{(t-\tau)} C_0(\tau)\big(V_0(\tau) + R_0 I(\tau) + RZ(\tau) - V(\tau)\big)d\tau \\ r_2(t) = \int_0^t \lambda^{(t-\tau)} C_0^2(\tau)d\tau \\ r_3(t) = \int_0^t \lambda^{(t-\tau)} C_0^2(\tau)x(\tau)d\tau \\ r_4(t) = \int_0^t \lambda^{(t-\tau)} C_0(\tau)x(\tau)\big(V_0(\tau) + R_0 I(\tau) + Rz(\tau) - V(\tau)\big)d\tau \\ r_5(t) = \int_0^t \lambda^{(t-\tau)} C_0^2(\tau)x^2(\tau)d\tau \end{cases} \quad (20)$$

V. EXPERIMENTAL VALIDATION

In this section, we experimentally validate the proposed SOC and charge capacity estimation algorithm using a lithium ion phosphate (LFP) polymer battery cell.

A. Experimental Data and Model Identification

The battery cell used in this study is an 1100 mAh LFP polymer cell purchased from BatterySpace. The data has been collected under controlled temperature at 25 °C using an MTI battery analyzer with 1 mV resolution. The battery is cycled from a nearly full state of charge using a pulse discharge and charge cycle as seen in Fig. 2. The current pulses are applied at around C/4, C/2, and 1 C rates for 5 minutes followed by either a 10, 30, or 60 minute rest period in between. The voltage has been limited to the range of 3 to 3.5 V, covering around 90% of the battery's energy capacity.

The parameters of the battery model are identified using an offline least squares optimization process with a cost function similar to Eq. (12), except that R_0, R, T, and $a_0, \dots, a_n$ are all included in the optimization process as well. The details of the system identification procedure are omitted from this paper in

the interest of space, but the results for a 10th order OCV polynomial are depicted in Fig. 3 with the identified parameter values listed in Table I.

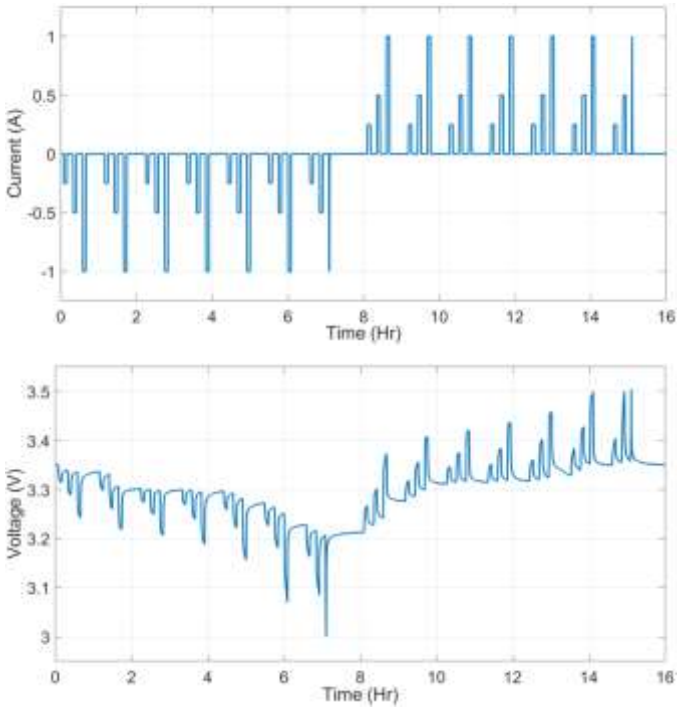


Fig. 2. Multi-rate pulse discharge and charge data from an LFP polymer battery.

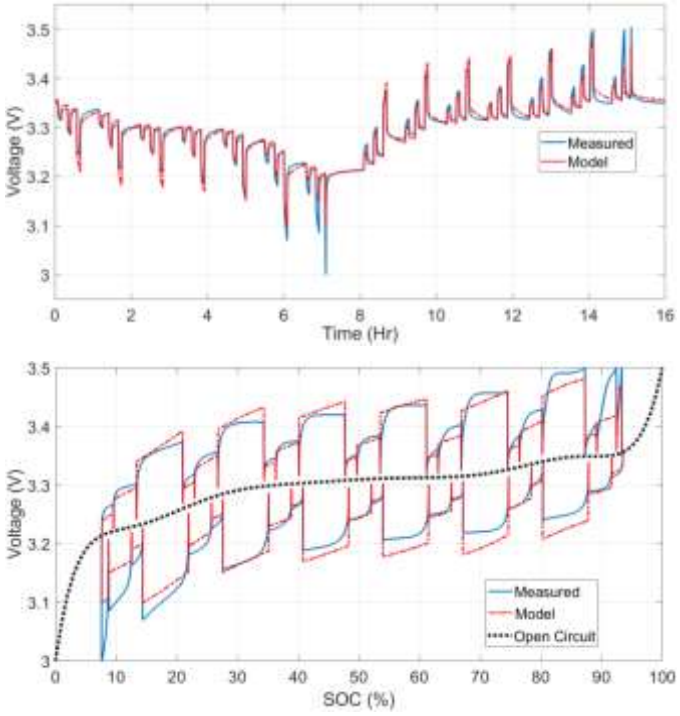


Fig. 3. Model vs. measured voltage as a function of time and SOC.

TABLE II. IDENTIFIED PARAMETERS OF THE BATTERY MODEL

Parameter	Value (Unit)	Parameter	Value (Unit)
Q	1.1 (Ah)	R	0.12 (Ω)
T	1330 (s)	R_0	0.1 (Ω)

As can be seen from Fig. 3, the model tracks the measured voltage curve with a fair accuracy. The present discrepancies are mainly due to the simplicity of the model compared to the complexities of the real electrochemical processes in a battery cell. Nonetheless, the model appears to be reasonably accurate for deployment in an online estimation algorithm.

Figure 3 also depicts the identified 10th OCV polynomial curve. To identify the parameters of the battery, including the initial conditions, the optimization has imposed a pair of constraints on the OCV curve at 3 V and 3.5 V, corresponding to 0% and 100% SOC, respectively. As can be seen from the figure, the OCV curve is fairly flat between 10-90% SOC, with only 0.1 V or 3% variation. This imposes a significant challenge on the performance of the SOC estimation algorithm, which relies heavily on the accuracy of the OCV curve. In the following sections, we review the performance of the proposed SOC and charge capacity estimation algorithm.

B. SOC Estimation

To validate the SOC estimation algorithm, we fix all the parameter values except for the initial SOC. To allow for the estimator to settle, data for two consecutive pulse discharge/charge cycles have been used. Figure 4 demonstrates the convergence of the estimated SOC trajectory from a number of different initial conditions to the SOC trajectory obtained from the coulomb counting process with an accurate initial condition. As can be seen from the figure, the SOC estimates converge to the coulomb counting-based SOC estimate within around 5 hours.

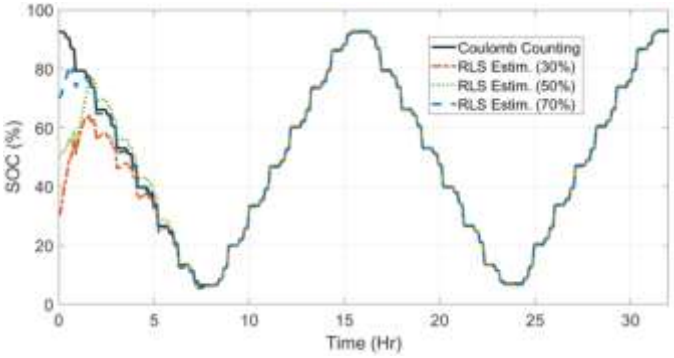


Fig. 4. Convergence of the SOC estimation from different initial conditions to the SOC trajectory obtained from the coulomb counting process.

A comparison of the proposed estimation with a simple state observer with the following equation is shown in Fig. 5:

$$\dot{\hat{s}} = \frac{1}{Q} I + L\big(V - (V_{oc}(\hat{s}) + R_0 I + V_r)\big), \quad \hat{s}(0) = \hat{s}_0 \qquad (21)$$

where L is the state observer gain. For simplicity, V_r is computed via an open-loop model, given its stable dynamics and near zero initial condition in this test.

Figure 5 shows that the recursive least squares estimator outperforms the state observer for three different values of the observer gain. When the gain is small, the observer takes a

longer time to converge, and when it is large, the model errors are amplified and reflected in the observer's response. The recursive estimator enables both fast and smooth convergence.

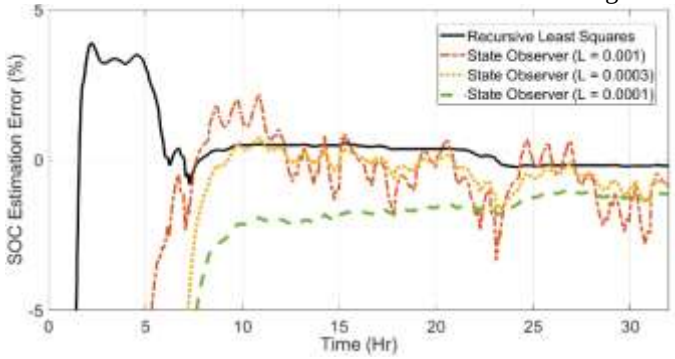


Fig. 5. Comparison of the recursive least square estimator with state observer.

C. SOC and Charge Capacity Estimation

To estimate the charge capacity and SOC simultaneously, we initialize SOC and Q estimates to different initial conditions (i.e. 0.3, 0.5, and 0.7 for SOC, and 0.9, 1.0, 1.1, 1.2, and 1.3 Ah for Q) and use the formulas in Sec. IV B to adapt SOC and Q estimates over time. Figure 6 shows the convergence of Q to the real value of 1.1 Ah regardless of the initial condition. Given that estimation of charge capacity is a slower process than that of SOC, it takes a longer time for the trajectories to settle, but they all are aligned and headed toward the same value after around 32 hours. The SOC trajectories converge to the coulomb counting trajectory in a similar fashion as before.

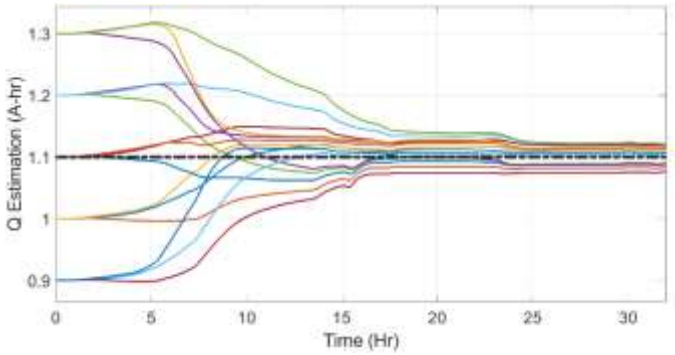


Fig. 6. Convergence of the charge capacity estimate from different initial conditions to the real value of 1.1. Ah.

D. Robustness with Respect to Model Uncertainty

To evaluate the accuracy of the proposed method in the presence of parametric uncertainties in the model, we can perturb the fixed parameters such as R_0, R and T from their identified values, and rerun the estimator. The estimation results for the cases of perturbing R_0 by ±25% and ±50% is shown in Fig. 7. The initial SOC and Q estimates are set to 50% and 1.2 Ah, respectively. Both SOC and charge capacity estimates converge toward the real values in a similar fashion as before, with a different transient behavior, depending on the amount of perturbation in R_0. At steady state, the SOC estimation is not affected by the perturbation of R_0. However, the charge capacity estimate seems to have found a slight offset, particularly for larger perturbations of R_0. Nonetheless, the offset is within 2% of the actual value, and the estimated values remain reasonably accurate.

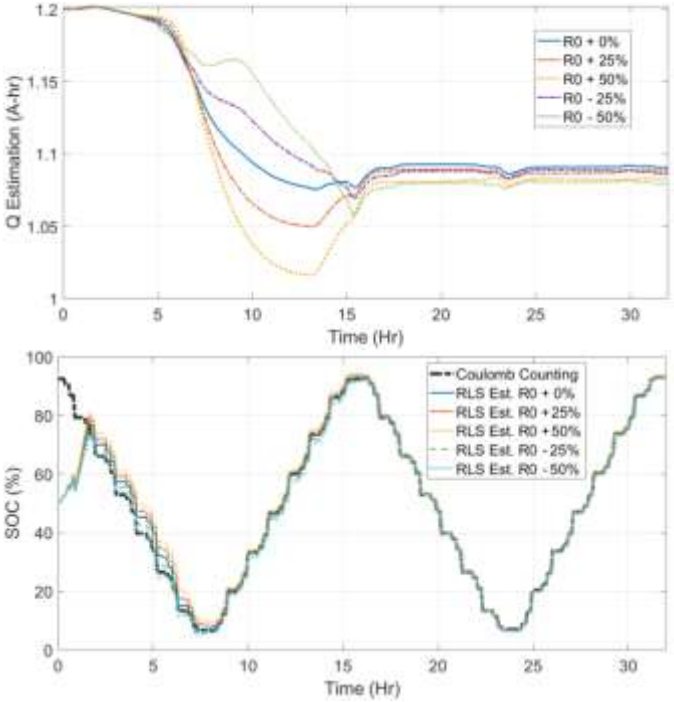


Fig. 7. Impacts of R_0 perturbations on the SOC and charge capacity estimation accuracies.

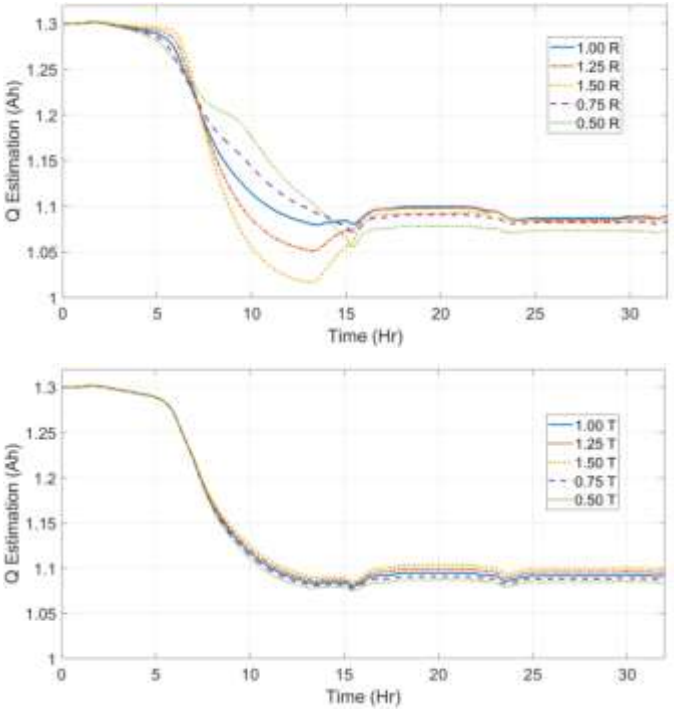


Fig. 8. Impacts of R and T perturbations on the charge capacity estimation accuracy.

Similar trends are observed for the perturbations of R and T, as shown in Fig. 8, which further endorses the robustness of the proposed method. In addition, Fig. 9 shows the charge capacity

estimation performance when all the three parameters (i.e., R_0, R, and T) are perturbed by at least 50% above or below the nominal values. All the eight possible combinations along with the initial estimation based on the nominal values (i.e., Dark thick line) are shown in the figure. The charge capacity estimation accuracy remains within ±3% for ±50% simultaneous perturbations in R_0, R, and T.

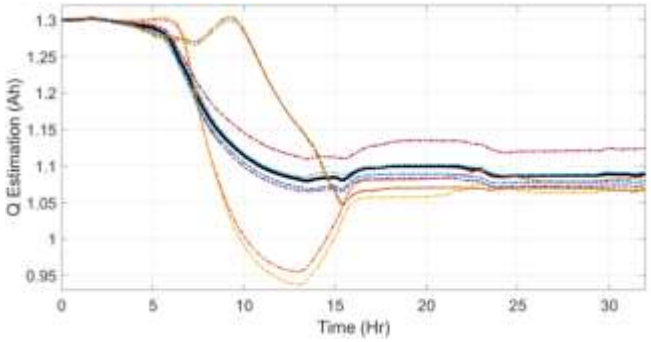


Fig. 9. Impacts simultaneous ±50% perturbation in R_0, R , and T on the charge capacity estimation accuracy.

More test temperatures and loading conditions and different battery chemistries must be investigated to fully evaluate the effectiveness of the proposed method in estimating charge capacity and state of charge of electrochemical batteries. In addition, the impacts of sensor bias and noise must be examined for field implementation.

VI. CONCLUSIONS

In this paper, a method for the simultaneous estimation of battery state of charge and charge capacity was developed and experimentally validated using an LFP polymer battery cell. The estimator is based on the recursive least squares estimation process, using an equivalent circuit battery model. The derived formulas are simple and real-time friendly. Experimental results indicate the method is effective and has reasonable robustness with respect to model inaccuracies. The estimator relies on an accurate OCV curve and sensor measurements. More investigation must be carried out to understand the impacts of sensor bias and noise, as well as OCV inaccuracies on the performance of the proposed estimation method.

REFERENCES

[1] X. Hu, S. Li, H. Peng, "A comparative study of equivalent circuit models for Li-ion batteries," *Journal of Power Sources*, vol. 198, pp. 359-367, 2012.

[2] M.-K. Tran, A. DaCosta, A. Mevawalla, S. Panchal, and M. Fowler, "Comparative study of equivalent circuit models performance in four common lithium-ion batteries: LFP, NMC, LMO, NCA," *Batteries*, vol. 7, no. 3, p. 51, 2021.

[3] J. C. Forman, S. Bashash, J. L. Stein, and H. K. Fathy, "Reduction of an electrochemistry-based Li-ion battery degradation model via constraint linearization and Padé approximation", *Journal of the Electrochemical Society*, vol. 158, pp. A93-A101, 2011.

[4] C. Li, N. Cui, C. Wang, C. Zhang, "Reduced-order electrochemical model for lithium-ion battery with domain decomposition and polynomial approximation methods", *Energy*, vol. 221, 119662, 2021.

[5] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, D.-I. Stroe, "A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery," *Applied Energy*, vol. 300, 117346, 2021.

[6] S. Park, J. Ahn, T. Kang, et al. "Review of state-of-the-art battery state estimation technologies for battery management systems of stationary energy storage systems," *Journal of Power Electronics*, vol. 20, pp. 1526–1540, 2020.

[7] K. S. Ng, C.-S. Moo, Y.-P. Chen, Y.-C. Hsieh, "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Applied Energy*, vol. 86, pp. 1506-1511, 2009.

[8] G. Fathoni, S. A. Widayat, P. A. Topan, A. Jalil, A. I. Cahyadi and O. Wahyunggoro, "Comparison of State-of-Charge (SOC) estimation performance based on three popular methods: Coulomb counting, open circuit voltage, and Kalman filter," *International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*, pp. 70-74, 2017.

[9] G. L. Pett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, pp. 277-92, 2004.

[10] D. D. Domenico, A. Stefanopoulou, and G. Fiengo, "Lithium-ion battery state of charge and critical surface charge estimation using an electrochemical model-based extended Kalman filter," *Journal of Dynamic Syst., Measurement, and Control*, vol. 132, pp. 061302, 2010.

[11] H. Chaoui, N.Golbon, I. Hmouz, R. Souissi, and S. Tahar, "Lyapunov-based adaptive state of charge and state of health estimation for lithium-ion batteries," *IEEE Transactions on Industrial Electronics*, vol. 62, no 3, pp. 1010-1018, 2015.

[12] B. M. Othman, Z. Salam and A. R. Husain, "Analysis of online Lyapunov-based adaptive state of charge observer for lithium-ion batteries under low excitation level," *IEEE Access*, vol. 8, pp. 178805-178815, 2020.

[13] O. Rezaei, H. A. Moghaddam, B. Papari, "A fast sliding-mode-based estimation of state-of-charge for Lithium-ion batteries for electric vehicle applications," *J. of Energy Storage*, vol. 45, 103484, 2022.

[14] M. Souaihia, B. Belmadani, F. Chabni, A. Gadoum "A comparison state of charge estimation between Kalman filter and sliding mode observer for lithium battery," Advances in Green Energies and Materials Technology. Springer Proceedings in Energy. Springer, Singapore, 2021.

[15] X. Sun, J. Ji, B. Ren, C. Xie, and D. Yan, "Adaptive forgetting factor recursive least square algorithm for online identification of equivalent circuit model parameters of a lithium-ion battery," *Energies*, vol. 12, no. 12, p. 2242, 2019.

[16] S. Bashash and H. K. Fathy, "Battery state of health and charge estimation using polynomial chaos theory," *Proceedings of ASME Dynamic Systems and Control Conference*, Palo Alto, CA, Oct 2013.

[17] G.L. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based hev battery packs," *Journal of Power Sources*, vol. 161, no. 2, pp. 1356-1368, 2006.

[18] Shuzhi Zhang, Xu Guo, Xiongwen Zhang, "An improved adaptive unscented kalman filtering for state of charge online estimation of lithium-ion battery," *J. of Energy Storage*, vol. 32, 101980, 2020.

[19] S. Tong, J. Lacap, J. W. Park, "Battery state of charge estimation using a load-classifying neural," *Journal of Energy Storage*, vol. 7, pp. 236-243, 2016.

[20] Z. Cui, L. Wang, Q. Li, K. Wang., "A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network," *International Journal of Energy Research*, vol. 46, pp. 5423-5440, 2022.

[21] J. C. Álvarez Antón, P. J. García Nieto, C. Blanco Viejo and J. A. Vilán Vilán, "Support vector machines used to estimate the battery state of charge," *IEEE Transactions on Power Electronics*, vol. 28, no. 12, pp. 5919-5926, 2013.

[22] J. Li, M. Ye, W. Meng, X. Xu and S. Jiao, "A novel state of charge approach of lithium Ion battery using least squares support vector machine," *IEEE Access*, vol. 8, pp. 195398-195410, 2020.

[23] C. Vidal, P. Malysz, P. Kollmeyer and A. Emadi, "Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, vol. 8, pp. 52796-52814, 2020.

Energy Management of Hybrid Microgrid System using Multi Agent System: A Distributed Control Approach

Balachennaiah P
Dept. of EEE
Annamacharya Institute of Technology and Sciences
Rajampet, A.P.,India
pbc.prudhvi@gmail.com

Chinna Babu J
Dept.of ECE
Annamacharya Institute of Technology and Sciences
Rajampet, A.P.,India
jchinnababu@gmail.com

***Abstract*—Low voltage networks, typically, established and managed by small consumers fall under the category of micro grids. These grids with components like distributed generators, variable loads, storage arrangements etc., can be operated in island mode or as a part of the grid. Renewable energy is becoming more preferable in view of its cleanliness and protecting the environment. The variety of devices interconnected leads to the requirement of a Multi Agent System (MAS) with a prime motive of guaranteeing a stable operation with the proposed system having two micro-grids of 1.5Kw (Wind), 1Kw (Solar), a 24V, 150Ah battery all connected to a local load. A simulation model for dynamic energy management (DEM) in the Java agent development (JADE) environment, is created which optimizes the course of action on hourly basis of the system. The simulation suggests that MAS can successfully manage dynamic loads, generated power in real-time micro-grids.**

Keywords— Micro-grids, Multi Agent systems (MAS), real time simulation, Energy Management system (EMS), Solar Power, wind Power, Java Agent development (JADE) frame work.

I. INTRODUCTION

The energy demand scenario around the world is a constant tussle between the ever rising demand and the rapidly depleting resources such as conventional fossil fuels. Weaning away from fossil fuels and building micro-grids powered by renewable energy sources are the latest developments to meet such requirements [1]. A micro-grid as a fraction of a power distribution network comprising modest generating capacities catering to local loads and having capability to function either independently or in tandem with the grid [2]. The focus of such an activity being demand supply matching, reduction in costs, energy procurement and storage facility enhancement. Effective and sophisticated management and controls are essential for energy management(EM) of micro-grids because of the nature of intermittent, low inertia characteristics of the micro-grid[3].

Planning and monitoring on the centralized controls for micro-grid power management was suggested as a predictive model control approach for the best functioning of the micro-grid [4]. To stabilize the micro-grid for various parameters such as under voltage and load variations, adaptive fuzzy logic and PI controllers were suggested [5], and in the islanded mode, the cooperative control of the direct control vector and the droop control approach are provided in [6], a rule-based EMS that was created to track and manage the true power flow within the smart microgid (SMG) system [7]. The energy management system (EMS) guarantees the energy stability of an AC/DC micro-grid which includes a battery and renewable energy sources (RES) [8].Few lacunae of the systems discussed above are - lack of run-time adaptive behaviour, communication overhead which could be overcome by effective communication and autonomous control mechanisms incorporated into the micro-grid monitoring process. An integrated solution technique addressing the distributed computing, communication, and data integration aspects, comprising numerous software agents known as a multi-agent system (MAS), which enables these agents to produce results cooperatively to solve problems that are beyond the scope of either one individually was proposed by Bogaraj and Kanakaraj [9].

In [10] authors proposed an ideal Micro-grid EMS featuring distributed storage and disaster recovery for lowering peak demand while lowering power costs. The advantage of MAS decentralised algorithm, lowers the cost of the power imbalance while taking customer preferences into account in the decision-making process thereby concluding that the proposed approach's decision-making time is quicker than the centralised approach [11].The issue of real time management(RTM) of smart grids was solved in [12] where in, a telecommunications system connects the energy management to the power system. An improved distributed energy management system for solar micro-grids was proposed in [13] using a smart grid architecture. Dynamic energy management is done through MAS implemented in JADE. In [14], authors explained how a multi-agent system (MAS) coordination approach was utilized to construct sophisticated demand side management of a solar micro-grid and distributed energy management. For the real-time operation of a hybrid micro-grid in both isolated and connected modes, an intelligent agent-based control is offered in [15].For hybrid micro-grids, this research proposed an agent-based management and control system. To build the multi agent system, JADE environment was employed. The system's agents

cooperate and interact with one another to achieve its energy management goal. The main objective of this paper is to develop and deploy an intelligent control for micro-grids in real time.

The section wise description of the structure of paper is: Sec. 2, describing a hybrid micro-grid. Sec. 3 a detailed explanation of multi agent system technology, Sec. 4, JADE description is provided, Sec. 5 provides a problem statement, Sec. 6 Implementation of EMS of a hybrid micro-grid in a distributed setting using JADE, Sec. 7 includes a case study and the simulation results. Section 8 provides a conclusion.

II. HYBRID MICROGRID SYSTEM

In hybrid micro-grid systems (HMGS), functioning in islanded and grid-connected modes, many dispersed resources are connected in parallel and have electronic controls. The most affordable solution for addressing the issue of power supply in distant locations that are located far from grids is HMGS based on renewable energy sources (RES) [16]. At EEE department of Annamacharya Institute of Technology and Sciences(AITS)' real-time hybrid micro-grids were installed comprising of 1.5 KW wind turbine generator systems and 1Kw solar panel systems. Battery banks are utilized as energy storage to power the 0.22-kv low voltage network. The loads in rooms 303, 304 and 305 are supplied by Microgrid 1. The loads in rooms 403 and 404 are supplied by Microgrid 2. An overview of the connection and management of DERs within the micro-grid is shown in Fig. 1.

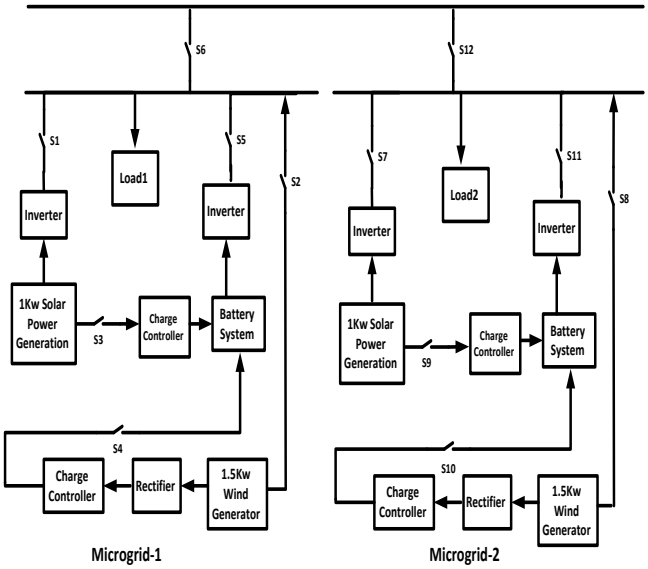


Fig.1. Block diagram of Hybrid microgrid system.

With sources of solar and wind being available, the user's demand can partially be met and the battery is charged which reduces the randomness in the supply such that the load always receives a constant supply of power. By using JADE, this paper suggested a control method for managing DERs. JADE's operating system independence makes it possible to implement coded algorithms on any personal computer as long as the Java virtual machine is present. This is the major justification for using JADE.

III. MULTI AGENT SYSTEM TECHNOLOGY

A multi-agent intelligent system accomplishes a common objective, through MAS by enabling a smart agent, to stands in for each component of a micro-grid, to ensure the three standard characteristics— proactive, reactive, and social abilities . These traits highlight the value of agent technology in creating complex systems by using agent communication language (ACL) [17]. They are able to act independently and behave in an object-oriented manner. Every intelligent representative would have the subsequent features:

(a) Autonomy/flexibility: To be able to work freely without outside assistance, but with a minimum of fundamental control over its decisions and actions.

(b) Reactivity: The capacity to recognise changes in the environment and act swiftly.

(c) Proactively: Ability to initiate to goal-oriented actions without relying on input reactions.

(d) Social aptitude: The capacity of an agent to communicate in a common language with humans and other agents.

(e) Data collection: Collecting enough data is indicative of environmental awareness, which helps decide to achieve its objectives.

(f) Protocols: A list of clearly defined protocols that recommend the appropriate way to communicate with humans or other agents as a part of the system.

(g) Helping agent: To offer assistance, share location, and work with other agents.

In power system engineering applications, multi-agent systems have been deployed since the past decade, in a growing number of varied applications. The principles, methods, technical issues, and possible benefits of applying multi-agent systems to power systems were well-explained in [18]. The authors spoke about the standards, resources, auxiliary technologies, and design approaches that may be used to implement MAS in power systems.

IV. JADE DESCRIPTION

JADE's simple distributed platform allows users to operate a micro-grid for managing and monitoring power balance as it is an open system with connector compatibility and extensibility [19].In this study, the FIPA (Framework of Intellectual and Physical Agent) standard-compliant JADE framework for intelligent agents is used. An interface platform is the setting in which software agents work [20]. A platform is made up of several containers, with an agent hosted in one of them. On enrolling with an AMS for a valid Agent ID, facilitates access the agent platform, which maintains a repository of Agent Identifiers (AIDs) and operator states. Agencies can use Directory Facilitator's (DF) platform's basic yellow page services to locate other agents in the system based on the services they wish to provide or receive. Fig. 2 illustrates the MAS model

This work is supported by Department of Science and Technology – Interdisciplinary Cyber Physical System, New Delhi, India.

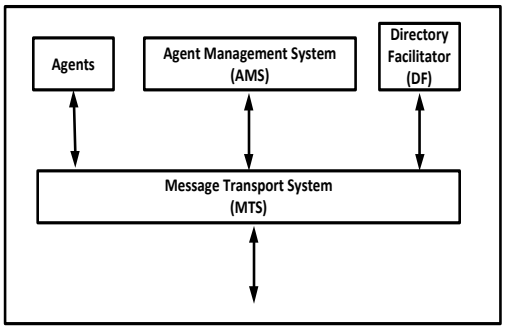


Fig.2. Multi agent system model

In JADE, both primary and secondary receptacles are present and the principal component is always started first. Agents for the Agent Management Platform (AMS) and DF are also set up instantly with the launch of main containers. The directory services, administration services are hosted by the main container, and are replicated on the other containers for redundancy, to build agent functionality, agent administration service, Directory Facilitation, and message delivery functions.

V. PROBLEM FORMULATION

The PV power and wind power are used to build the mathematical formalism of the EMS. The input signals are the load and the current battery state of charge(SOC), and the output signals are the grid on/off, load on/off, battery charging, and battery discharging.

The total energy produced by renewable sources is,

$$P_{pv} + P_{wp} = P_E \qquad (1)$$

where P_{pv} is the amount of electricity produced by solar panels, and P_{wp} is the amount of power produced by wind power.

And total load,

$$P_{L11} + P_{L22} = P_L \qquad (2)$$

where, P_{L11} is load in Room Nos: 303 and 304, P_{L22} is load in Room No: 305. After comparing P_E and P_L

If $P_E > P_L$,

then charge battery through renewable energy sources.

If $P_E < P_L$, battery Capacity is more than 0.9,

then discharge battery to supply any additional load that cannot be met by renewable energy.

If SOC<0.9.

then check P_{grid} to supply extra load

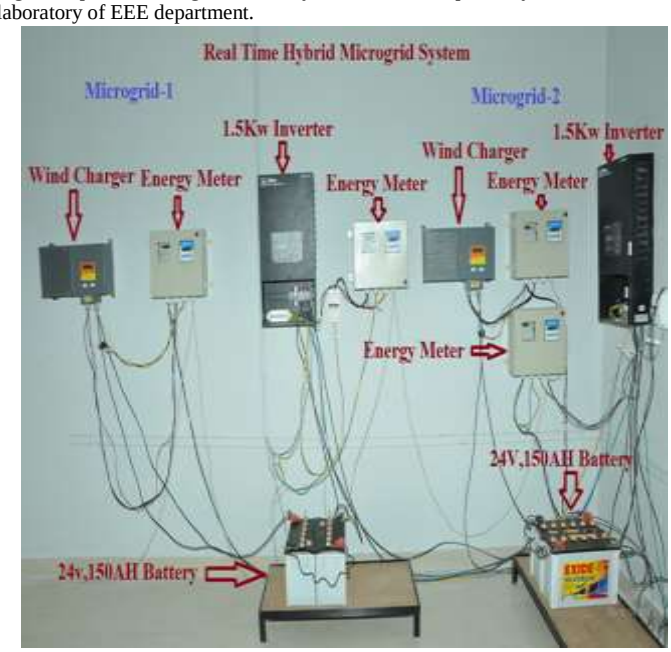
If $P_{grid} > 0$,

then give supply to extra load. In this situation, turn off a reasonable load if P_{grid} does not have extra power to give.

VI. IMPLEMENTATION OF MAS FOR ENERGY MANAGEMENT OF HYBRID MICROGRID SYSTEM

JADE is needed to implement the suggested agent-based system. JADE is an open-source framework for creating a variety of co-systems that is based on Java. Peer-to-peer technology is the foundation of the JADE architecture. Through wireless or wired networks, it is feasible to evenly divide intelligence, initiative, information, resources, and command among a variety of hosts and devices. Each agent has the ability to speak with and bargain with its peers in order to discover a solution to an issue that benefits both parties. The two micro-grid systems considered are a 1 kW solar PV system, a 1.5 KW wind turbine generating system, a 24 V, 150 AH battery bank system, and local load. 1kw rated Solar PV systems and 1.5 Kw rated wind turbine generator system are installed in the Roof top of EEE department, control systems, measuring instruments and sensors are installed in the Power system research laboratory of EEE department. Power is being provided by Microgrid-1 to connected loads in Room Nos. 303, 304, and 305 of the EEE department and by Microgrid-2 to linked loads in Room Nos. 403, 404 of the EEE department. Proposed micro-grid control system is shown in Fig. 3.

Fig.3. Proposed micro-grid control system installed in power system research laboratory of EEE department.



The sensors monitor in real-time solar and wind energy on an hourly basis taking into account all electrical loads. Figs. 4 and 5 depict the load, wind, and solar power graphs for microgrids 1 and 2. Based on the parameters of the solar power, wind power, load, and state of charge (SOC) of the battery, monitored hourly, the agent implements the most practicable actions for dynamic energy management of the hybrid micro-grids in a distributed environment, automatically. Fig. 6 shows the flow chart incorporating all potential uses of micro-grids.

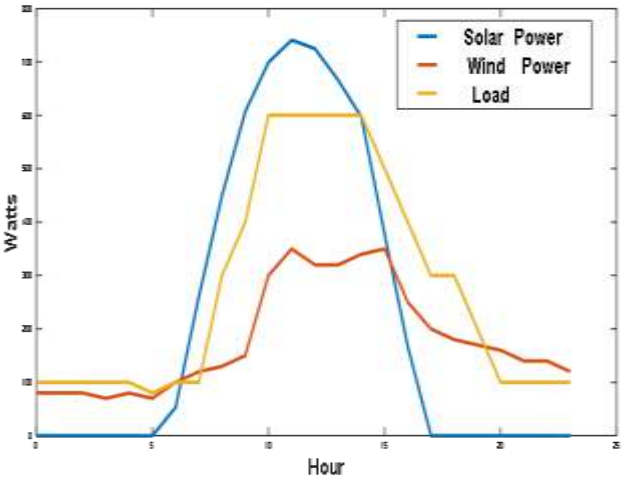


Fig.4. Solar power, wind power and load for Microgrid-1

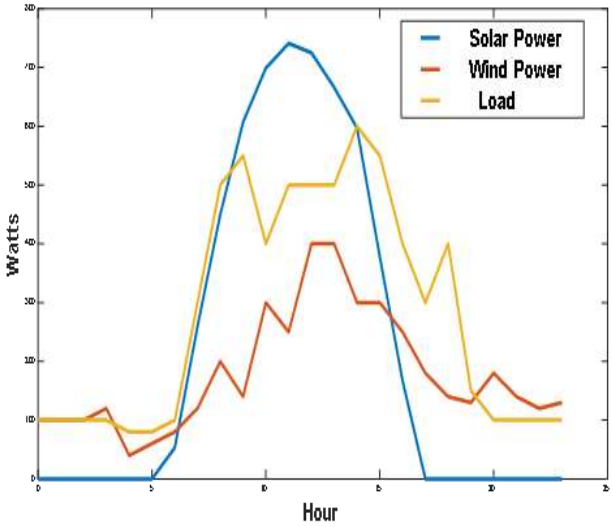


Fig.5. Ssolar power, wind power and load for Microgrid-2

The agents given below mimic a MAS in a JADE environment:

Load Agent (LA): Load Agent or Purchasing Agent decides quantity of power to be bought and communicates with the generating agent by searching the DF yellow pages providing power distribution solutions. In this work , the microgrid-1 load agent (LA1) and the microgrid-2 load agent (LA2) were considered.

Generation agent (GA): The LA requests the generation agent (GA) considering the microgrid-1 solar power agent (SA1), microgrid 2 solar agent (SA2), microgrid-1 wind power agent (WA1), and microgrid 2 wind power agent (WA2).

Battery Agent (BA): Battery Agent (BA) coordinates the condition of the battery's charge, communicates to and from with other agents about the availability and demand for power. In our scenario, we take into account the microgrid-1 Battery Agent (BA1) and the microgrid -2 Battery Agent (BA2).

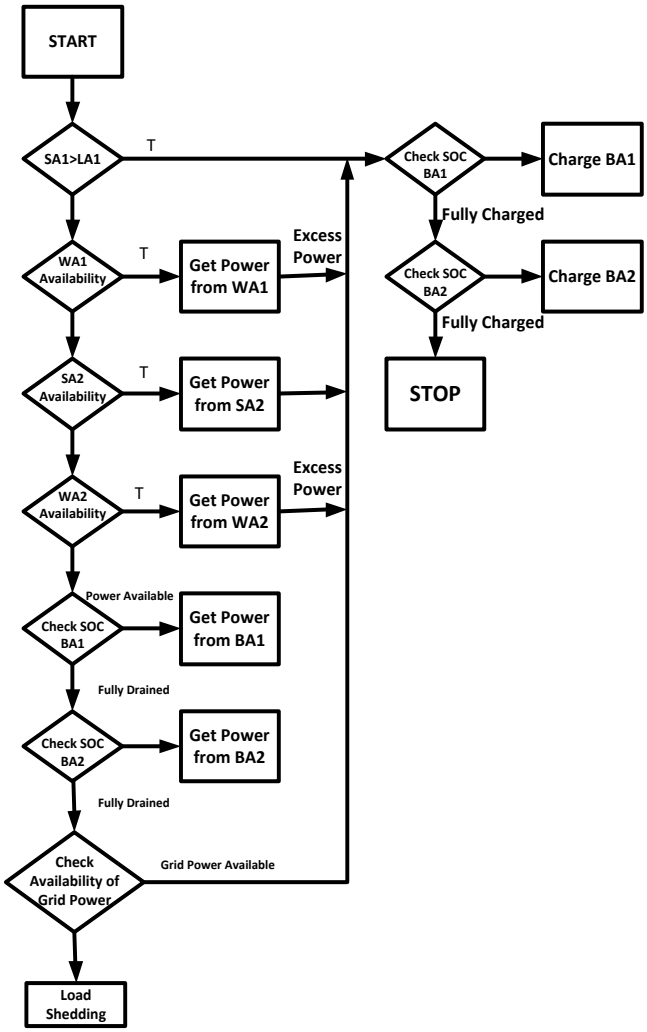


Fig.6. Flow chart of EMS of the micro-grids using MAS

Control agent (CA): Monitoring, managing, negotiating, and executing power exchange across the micro-grids are the tasks catered by the control agent (CA). Fig. 7 depicts the agent's relationship map. The procedure adopted is as follows:

Step 1: SA1 requests power from LA1 by a ACL message, SA1 delivers power to LA1. Power is provided to BA1 when power is in excess. BA2 receives extra power. Thus, the LA makes local decisions and coordinates with other agents to make global decisions.

Step 2: In the absence of enough power at SA1, LA1 contacts the WA1 and obtains the power that is available.

Step 3: LA1 contacts SA2 and receives more power if needed.

Step 4: In case of power is still needed, LA1 contacts WA2, which then provides LA1 with required power.

Step 5: In case of need of power further , the charge of BA1 is verified , BA1 powers LA1 if the battery is fully charged.

Step 6: If BA1 is not fully charged, verify BA2's charge status. On BA2 being fully charged provide LA1 with power .

Step 7: In case BA1 and BA2 are devoid of any charge, in storage, then the control agent interacts with the grid agent to charge the batteries.

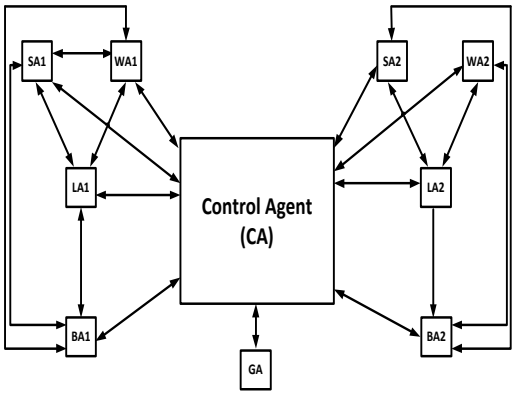


Fig.7. Agents relationship

Step 8: LA1 gets in touch with BA1 to fulfil the necessary load requirement. In the event of deficiency still , LA1 calls BA2, which then provides LA1 with the necessary power .

Step 9: The agent chooses the optimum energy management option for a distributed environment every hour depending on the load need and the availability of wind and solar power.

Step 10. Similar procedures are repeated for the load agent (LA2) in the microgrid-2, with ACL used for all communication which results in the multi agent system on the JADE platform to be used dynamically to manage the energy of the solar and wind based hybrid micro-grid for distributed optimization.

VII. SIMULATION RESULTS AND CASE STUDY

Hybrid micro-grids were operated in a single cycle for 24 hours, taking into account all potential outcomes. All of the processes are taken into account in accordance with the flowchart, and for these scenarios, all seven agents of the solar and wind-based hybrid micro-grids are programmed in Java using the JADE framework and run in the IntelliJIdea Environment. Numerous situations are taken into account, and the analysis of agent interaction and transaction data are shown in sniffer diagrams and console output. Snapshot of a communication exchange between a sniffer agent and an agent is shown in Fig. 8. The energy of the solar power, wind power, battery, and load for micro-grids 1 and 2 at 2 pm are taken into account in this case study. Below is a list of the operations in order:

(i) Micro-grid 1 load requires 600 W; however the total amount of power that could be generated is 897 W, including 300 W from wind and 597 W from solar. The extra power is therefore 297 W.

(ii) Micro-grid 2 load requires 600 W; however the total amount of power that could be generated is 937 W, including 340 W from wind and 597 W from solar. The extra power is therefore 337 W.

(iii) There is surplus power in both Micro-grids. Therefore, the extra power will be utilised to recharge the batteries.

(iv) Finally, the control agent checks the SoC of the batteries and charges the batteries with the lowest SoC first using the excess power from both Micro-grids 1 and 2.

(v) In case the batteries are fully charged as in this instance, the control agent does not continue to charge them. The console output for a particular scenario is given in table 1.

TABLE III. CONSOLE OUTPUT

S. No.	Description	Microgrid-1	Microgrid-2
1	Load	600 W	600 W
2	Power tapped from solar agent	597 W	597 W
3	Power tapped from the wind agent	300 W	340 W
4	Power remaining	297 W	337 W
5	Power needed	0 W	0W
6	Power tapped from microgrid- 2	-	0 W
7	Power sent to Microgrid-2	-	0 W
8	Power tapped from microgrid -1	0 W	-
9	Power sent to Microgrid-1	0W	-
8	Power tapped from battery	0 W	0 W
9	Power tapped from battery	98.0%	98.0%

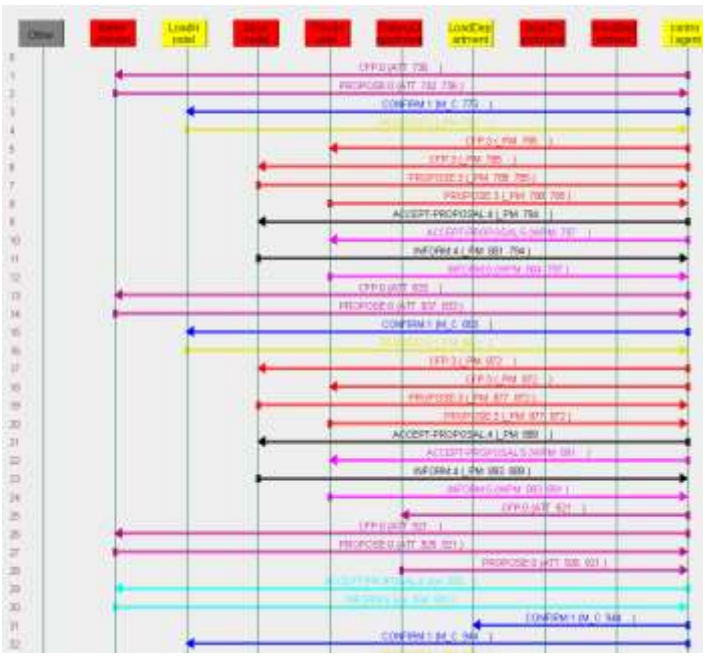


Fig.8. Sniffer agent and agent message exchange snapshot

VIII. CONCLUSION

This study provides a summary of the micro-grid designed and demonstrated at AITS. The advantages of the MAS based control system for the micro grid are well explained. In this simulation research, we successfully created agents that communicate and decide via message exchanges , based on the acquired data. Results prove that, the proposed control system based multi agent is efficient in handling energy during instantaneous micro-grid procedure. In order to undertake software verification before performing hardware verification on the AITS micro-grid, JADE will be integrated with MATLAB/Simulink/Labview in a future study, coupled with Internet of things to make it more efficient.

ACKNOWLEDGMENT

The DST-Interdisciplinary Cyber Physical System, New Delhi, India, provided financial assistance for the authors, which they warmly acknowledge (Grant No: DST/ICPS/CPS-Individual/2018/607).

REFERENCES

[1] ND Hatziargyrion, Micro-grids Architecture and control, Wiley-IEEE press, 2014.

[2] C. Chen, S. Duan, T. Cai, et al., Smart energy management syst. for optimal microgrid economic operation, IET Renewable Power Generation, vol. 5, no. 3, pp. 258–267, 2011.

[3] Microgrid Institute,http://www.microgridinstitute.org/, 2022 accessed: 08-09-2022.

[4] D.E. Olivares, A.Claudio, et al., A Centralized Energy Management Syst. for Isolated Microgrids, IEEE Trans. on smart grid, vol. 5, no. 4, july 2014.

[5] Subhashree Choudhury,Abhijeet Choudhury, et al., Optimal control of islanded microgrid with adapt. fuzzy logic & PI contr. using HBCC under various voltage & load variation, International Confer. on Circuit, Power and Computing Tech. (ICCPCT), pp.1-8, 18-19 March 2016.

[6] Malek Ramezani, Shuhui Li, Voltage and frequency control of islanded microgrid based on combined direct current vector control and droop control, IEEE PES General Meeting (PESGM), pp.1-5, 17-21 July 2016.

[7] Yuan Meng, A Rule-Based Energy Managemen. Syst. for Smart Micro-Grid, Australasian Universities Power Eng. Confer. (AUPEC), 26-29 November 2019.

[8] AyberkCalpbinici, Erdal Irmak,ErsanKabalcı, Ramazan Bayındır, Design of an Energy Management Syst. for AC/DC Microgrid, 3rd Global Power, Energy and Communication Confer. (GPECOM), 05-08, October 2021.

[9] T.Bogaraj , J.Kanakaraj, Intelligent energy management cont. for independent microgrid, Sadhana, 41(7):755–69, 2016.

[10] HK Nunna , S. Doolla, Energy manageme. in microgrids using demand response and distributed storage – a multiagent appro.. IEEE Trans. Power Deliv., 2013, 28(2):939–47.

[11] S.Ghorbani , R.Rahmani , Unland R. Multi-agent autonomous decision making in smart micro-grids energy management: a decentralized appro., German conference on multiagent syst. tech. Springer; 2017. p. 223–37.

[12] Roberto S. Netto , Guilherme R. Ramalho , Benedito D. Bonatto , Otavio A. S. Real-Time Framework for Energy Managem. Syst. of a Smart Microgrid using Multiagent Syst., energies, 2018.

[13] Leo Raju, R.S. Milton and Antony Amalraj Morais1, Multi Agent Syst. based Distributed Energy Manageme. of a Micro-Grid, I J C T A, 8(5), 2015, pp. 1947-1953.

[14] Leo raju, R.S.Milton , Senthilkumaran mahadevan, Implementation of energy manag. and demand side management of a solar micro-grid using a hybrid platform, Vol.25, 2017.

[15] H. N. Aung, A. M. Khambadkone, D. Srinivasan, T. Logenthiran, Agent-based intelligent control for real-time operation of a microgrid, Joint Int. Conference on Power Electronics, Drives and Energy Syst. & Power India, 20-23 December 2010.

[16] Sihem Amara; Sana Toumi; Chokri Ben Salah, Modeling and Simulation of Hybrid Renewable Microgrid System, 2020 17th Int. Multi-Conference on Syst., Signals & Devices (SSD), 20-23 July 2020.

[17] M.Pipattanasomporn, H. Feroze, and S. Rahman, Multi-agent syst. in a distributed smart grid: Design and implement., in IEEE/PES Power Syst. Confer. and Exposition, Seattle, WA, USA, pp. 1–8, 2009.

[18] W. Zhang, Y. Xu, W. Liu, C. Zang, and H. Yu, Distributed online optimal energy manag. for smart grids, IEEE Trans. on Industrial Informatics, vol. 11, no. 3, pp. 717–727, 2015.

[19] JADE tool kit available from: http://www.jade.tlab.com

[20] The Foundation for Intelligent Physical agent (FIPA):www.fipa.org.

A Spectrogram-based CNN Algorithm for Denoising ECG Signals

Sahar Keshavarzi
Department of Electrical & Computer Engineering
Isfahan University of Technology
Isfahan, Iran
s.keshavarzi@alumni.iut.ac.ir

Mohammad Soltanian
Department of computer science
Kharazmi University
Tehran, Iran
m.soltanian@khu.ac.ir

Mahmoud Keshavarzi
Department of Psychology
University of Cambridge
Cambridge, UK
mahmoud.keshavarzi.ir@ieee.org

***Abstract*—This paper proposed an algorithm based on the convolutional neural network (CNN) for enhancing noisy Electrocardiogram (ECG) signals. The proposed algorithm (called CNN III) took the spectrogram of ECG signal as the input of network and predicted the enhanced signal as the output. Three case studies were investigated to assess the performance of the proposed algorithm and to compare it with three other models including a recurrent neural network (RNN) and two CNN-based models (CNN I, CNN II). The size of data set utilized to train the models were different across the three case studies. We performed some simulations to assess the performance of investigated algorithms at different SNRs (from -5 dB to 15 dB). The performance was quantified based on two different measures including Pearson correlation and mean squared error. The simulation results suggested that the proposed algorithm (CNN III) outperformed the other three algorithms for enhancing ECG signal in the presence of noise. We particularly found that the mean correlation at low SNRs (-5 dB to 1 dB) were 0.915, 0.883, 0.756 and 0.859 for CNN III, CNN II, CNN I and RNN, respectively. Our results also showed that the performance diminished when reducing the SNR level. Furthermore, increasing the size of training data set led to improvement in performance for all the four investigated algorithms.**

Keyword—ECG signal, denoising, convolutional neural network, recurrent neural network, spectrogram

I. INTRODUCTION

Electrocardiogram (ECG) signal as a physiological signal representing the heart activities, is obtained through an electrically recording process in clinical environments and consists of five main waves known as P, Q, R, S and T. This signal is widely used as a useful tool for monitoring and diagnosing heart diseases.

The quality of ECG signals, however, is usually affected by noises resulted from different sources including power line interference, motion and muscle artefacts, noisy interference caused by body movement of the subject, poor contact of electrodes, baseline wander, and instrumentation noise. The distortion caused by these sources can negatively affect the morphological characteristics of ECG signals, leading to inaccurate diagnosis of heart diseases and clinical evaluations. Aiming to address this issue, thus, a wide range of methods have been proposed to improve ECG signals corrupted by noise. These ECG denoising methods include Empirical Mode Decomposition (EMD), techniques based on wavelet transform, hybrid wavelet techniques combined with filtering methods, adaptive filters, Savitzky Golay filter, Kalman filters, recursive filtering, Non-Local Means (NLM) filter, independent component analysis (ICA), and a group of machine learning-based algorithms such as support vector regressions (SVRs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs).

In [1], a real-time approach based on the EMD was proposed to detect the noise associated with the ECG signal. This method separates the clean ECG signal from the noisy signal in two steps. In stage 1, the first-order intrinsic mode function (F-IMF) corresponding to each of the two signals (clean and noisy signals) is obtained through the EMD decomposing process applied to signals. In stage 2, three statistical metrics including the Shannon entropy, mean, and variance are applied to the F-IMF obtained in stage 1, and the resulted measurements are considered to find the level of randomness and variability as the characteristic features which are used for detection of noise over a thresholding operation, leading to distinguish the noisy ECG from the clean ECG signal. The thresholds values were determined based on the data from 15 healthy participants with 24-h Holter recordings. Blanco-Velasco et al. [2] proposed an EMD-based method to enhance the ECG signal degraded by noise by removal of high frequency noises and baseline wander. This method performs the denoising procedure in four steps including: (1) Separating the QRS complex; (2) Applying a proper windowing to keep the QRS complex; (3) Employing statistical tests to find the number of IMFs contributing to the noise; (4) Filtering the noise through partial reconstruction. In [3], the denoising procedure is conducted in EMD and Discrete Wavelet Transform (DWT) domains. This approach, particularly, employed a windowing process in EMD domain to reduce the noise from the initial IMFs and to preserve the QRS complex information in the first three high frequency IMFs. The resulted signal was then transformed in DWT domain, where an adaptive soft thresholding procedure was applied to further reduction of noise and to reconstruct the original ECG signal with a higher time resolution.

In [4], a threshold-based method was presented to enhance ECG signal degraded by noise using the wavelet transform. This method used Daubechies wavelet (db4) through a multi resolution decomposing process where the signal was decomposed into five levels of the wavelet transform. An experimental threshold value was then calculated over a loop to achieve a result where the error between the detailed coefficients

of noisy ECG signal and those of clean ECG signal is minimized. The result obtained by this method was then compared with another denoising approach named Donoho's method, and a better result was reported for ECG signal denoising by the approach presented in [4]. Nilolaev et al. [5], proposed a two-stage denoising method combining the wavelet shrinkage with the Wiener filtering in translation-invariant wavelet domain. In the first stage, the noisy signal was decomposed in wavelet domain, the wavelet coefficients were shrinked using wavelet filters, and finally coefficients in shift-dependent wavelet domain were estimated. In the second stage, the results of the first stage were used to create an optimal Wiener filter to denoise the input noisy ECG signal.

Thakor and Zhu [6] used the adaptive filter for cancelling ECG noise and detecting arrhythmia. They employed a basic structure of adaptive filter to iteratively minimize the mean squared error (MSE) between the noisy ECG signal (primary input) and a reference input which was either a noise correlated with the noise associated with the input or a signal correlated only with the ECG signal in the the input.

Chakrabortya and Das [7] presented a method based on the Savitzky-Golay filter to enhance the ECG signal. They applied Savitzky-Golay filter to noisy ECG signal and significantly filtered out the signal's high frequency components along with the noise. The performance of the Savitzky-Golay filter was then compared with the band-pass filter from Pan-Tompkins algorithm of QRS detection, and a better performance was reported for the Savitzky-Golay filter.

In [8], an approach based on an extended Kalman smoothing (EKS) filter combined with a differential evolution (DE) technique was investigated to denoise ECG signals. This approach used the DE technique to automatically select ten optimized features of the ECG signal. These features were then utilized by the EKS for developing a state equation and for initialization of the EKS parameters. The performance of this approach was evaluated at different SNRs and compared with techniques such as adaptive filtering, extended Kalman filter, EKS, wavelet soft threshold-based technique, NLM, and conventional filtering at different SNR. They reported a better SNR improvement for the proposed method (EKS+DE). In [9], a method based on a recursive filtering was proposed for real time ECG signal denoising, with low computation cost in terms of memory and time consumption. Tracey and Miller [10] introduced an NLM-based technique to recover the original ECG signal from noisy observations. The results suggested that this technique performs ECG denoising while minimizing signal distortion and preserving the characteristic details of the ECG signal.

Kuzilek et al. [11] presented an ICA-based method which combines the JADE source separation and the binary decision tree for identifying noise and removing it from the signal. The results reported that the proposed method achieved a promising result in eliminating standard noises such as power line interference, baseline wander, Electromyography noise, and significantly better result in removing uncommon noises such as artefact caused by electrode cable movement.

In [12], two different algorithms including a SVR and an RNN were proposed to enhance noisy ECG signals. Both algorithms used the signal samples in the time domain as their input. The accuracy of the two algorithms was assessed and compared using three case studies which were different in size of training data set. The results showed the RNN outperformed the SVR. Arsene et al. [13] introduced two algorithms including a CNN and an RNN to denoise noisy ECG signals. Both algorithms took the time domain representation of signal as their input. A wavelet technique was also applied as the benchmark to be compared with the two proposed algorithms. The results demonstrated that the CNN achieved a better performance compared to both RNN and the wavelet technique. Antczak [14] proposed an RNN-based model for ECG signal denoising. This model consisted of a deep RNN and a denoising autoencoder, and it achieved SNR = 7.71 dB for noisy ECG signals presented at SNR = -8.82 dB.

In the current study, a spectrogram-based CNN algorithm was proposed to denoise ECG signals. The network took the spectrogram representation (2D data) of frames (each frame contained 300 time-samples of the ECG signal, and each period of signal contained 83 time-samples) of noisy ECG signal as the input and estimated the enhanced ECG signal as its output. Three models including the RNN proposed by Keshavarzi [12], CNN I (a CNN model with the input of 100 time-samples), CNN II (a CNN model with the input of 200 time-samples) were also considered as the benchmarks in our study.

This paper is structured as follows. Our proposed method is described in section II. Section III describes the simulation results and the case studies used to evaluate the performance of the proposed method. Finally, the conclusions are provided in section IV.

II. PROPOSED METHOD

Artificial neural networks (ANNs) are multi-layer structures which have the ability of learning patterns from the data. Each layer contains a number of neurons with parameters (such as weights) that are optimized during a training procedure [15]. ANNs are utilized to map an input data with multiple dimensions into multi-dimensional output data with an arbitrary degree of abstraction [15]. They have been widely applied for solving problems in different areas such as vision, hearing, medical diagnosis, weather forecasting, etc. In general, ANNs can be classified into three main architectures including feed-forward DNNs, RNNs and CNNs [16]. Of these architectures, RNNs have proved promising results in processing sequential data such as speech, language, and time series data. On the other hand, CNNs have been used as very successful class of models for processing two-dimensional (or three-dimensional) data in visual recognition tasks like image classification, image segmentation, etc.

In this paper, we proposed a denoising algorithm based on CNN to enhance ECG signals. The CNN took the spectrogram of noisy signal and estimated the enhanced signal in time domain. It included a 2D convolution layer (Kernel size: 4×4, Stride: 1×1), a max-pooling layer (Pooling size: 2×2), a flatten

layer and two fully connected layers. Number of units used in the first fully connected layer was 300, and in the second fully connected layer, number of units was equal to the number of time samples of each ECG data sample. The optimizer algorithm was Adam optimizer [17] (learning rate = 0.001, β1 = 0.9, β2 = 0.999, ε = 10^{-7}), and the cost function was MSE. The batch size and number of epochs were also 13 and 17, respectively. We used two python libraries including TensorFlow [18] and Keras [19] to build, train and test the proposed CNN model. Fig. 1 illustrates the block diagram of our proposed model.

III. SIMULATION RESULTS

We investigated three case studies to assess the efficiency of our proposed model (CNN III, with the input based on frames of 300 time-samples and an overlapping of 200 time samples between successive frames) and to compare it with three models including CNN II (a CNN model with the input based on frames of 200 time-samples and an overlapping of 100 time samples between successive frames), CNN I (the CNN model with the input based on frames of 100 time-samples and no overlapping between successive frames), and RNN ([12]). Both clean ECG data and noise (white Gaussian noise) was simulated using MATLAB. A MATLAB function named *awgn*() was used to generate the noise. The noisy ECG data was generated by adding noise to the clean ECG data at SNRs of -5 dB to 15 dB (with a step size of 2 dB).

For the three CNN models (CNN I, CNN II, CNN III), the spectrogram of the windowed signal was used as the input of network in all the three case studies. However, the RNN took the time-domain waveform of the windowed input signal as its input.

The investigated case studies varied in terms of the size of data set (number of data samples) used to train the network. The Number of data samples for the investigated case studies were 4400, 6600, and 8800. The number of data samples used as the testing data set in the three case studies was 2200. It

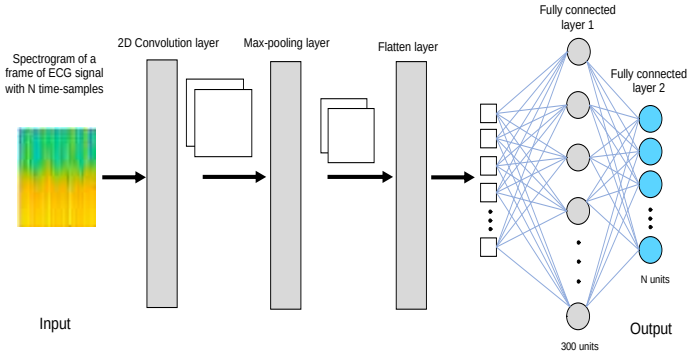


Fig 1. Block diagram of the proposed model (CNN III).

should be noted that each SNR value (-5 dB to 15 dB) contributed to the data equally. The MSE and Pearson correlation were used as the objective metrics to assess the performance of algorithms. The MSE between clean (x) and estimated ($\hat{x}$) signal is given as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[x(i) - \hat{x}(i)]^2$$

where n refers to the number of time-samples in each ECG data sample. The correlation, r, between clean and estimated signal is also calculated as:

$$r = \frac{\sum_{i=1}^{n}(x(i) - \bar{x})(\hat{x}(i) - \bar{\hat{x}})}{[\sum_{i=1}^{n}(x(i) - \bar{x})^2 \sum_{i=1}^{n}(\hat{x}(i) - \bar{\hat{x}})^2]^{0.5}}$$

where $\bar{x}$ is mean of clean signal and $\bar{\hat{x}}$ refers to mean of estimated signal.

A. *Case Study 1*

In case study 1, the models were separately trained using 4400 data samples (400 samples per SNR) and tested on 2200 data samples (200 samples per SNR). For the proposed model (CNN III), each data sample consisted of the spectrogram of a frame of ECG signal with 300 time-samples. For the CNN II, each data sample was the spectrogram of a frame of signal with 200 time-samples. For CNN I, each data sample consisted of the spectrogram of a frame with 100 time-samples. For the RNN, each data sample was a frame with 100 time-samples.

We quantified the performance of the four models using correlation and MSE. Fig. 2 shows the performance based on MSE and correlation versus SNR (dB) for the first case study. As shown in Fig. 2A, the correlation value for all algorithms rises when the SNR increases. A similar trend can be seen for the MSE measure (see Fig. 2B). Comparing the performance of algorithms based on the correlation metric, it is evident that CNN III outperformed the other three models, in particular for the low SNRs.

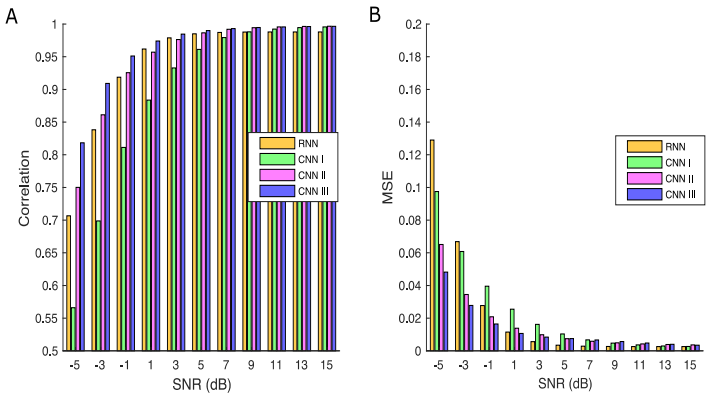


Fig 2. The performance of RNN, CNN I, CNN II, CNN III in terms of Pearson correlation and MSE in case study 1.

B. *Case Study 2*

In case study 2, we trained the models using 6600 data samples (600 samples per SNR) and tested on 2200 data samples

(200 samples per SNR). The testing data set was the same as testing data set used in case study 1. For CNN III, each data sample consisted of the spectrogram of a frame of ECG signal with 300 time-samples. For the CNN II, each data sample was the spectrogram of a frame of signal with 200 time-samples. For CNN I, each data sample consisted of the spectrogram of a frame with 100 time-samples. For the RNN, each data sample was a frame with 100 time-samples.

Fig. 3 shows the correlation score and MSE against SNR for the second study and for all the four algorithms. According to Fig 3, the efficiency of the algorithms, based on both correlation and MSE, increases when the SNR increases. Comparing the algorithms based on correlation value, the CNN III generally achieved a better performance than the other three models particularly for the low SNRs.

C. *Case Study 3*

In case study 3, we trained our proposed algorithm (CNN III) using 8800 data samples (800 samples per SNR) and tested it on 2200 data samples (200 samples per SNR). The testing data set was the same as the testing data set utilized in case studies 1 and 2. For CNN III, each data sample was the spectrogram of a frame of ECG signal with 300 time-samples. For the CNN II, each data sample was the spectrogram of a frame of signal with 200 time-samples. For CNN I, each data sample was the spectrogram of a frame with 100 time-samples. For the RNN, each data sample was a frame with 100 time-samples.

Fig. 4 illustrated the correlation score and MSE as functions of SNR for the third study and for all the four algorithms. As demonstrated in Fig 4, both correlation and MSE metrics suggested improvement in performance for all algorithms when increasing the SNR. This improvement is particularly remarkable for the low SNR values. In addition, it is evident that the proposed algorithm (CNN III) achieved a better performance than the other three models (RNN, CNN I, CNN II).

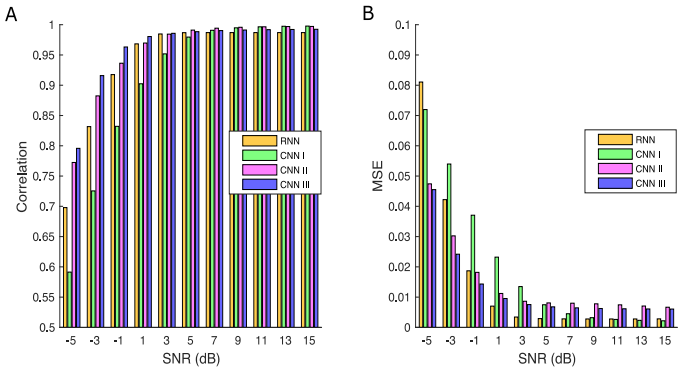


Fig 3. The performance of RNN, CNN I, CNN II, CNN III in terms of Pearson correlation and MSE in case study 2.

D. *Numerical results of case studies*

Comparing the results of the case studies investigated here, we found that the performance of the proposed algorithm (CNN III) improved when we increased the size of training data. This enhancement was greater for CNN III compared to the three other models, and more remarkable for the low SNRs (-5 dB to 1 dB) and for the correlation measure.

Table 1 shows the mean correlation values at the low SNRs across SNRs of -5 dB to 1 dB for all algorithms and for the three case studies. The mean MSE values across the SNRs of -5 dB to 1 dB, for all algorithms and for the three case studies, are mentioned in Table 2.

The proposed algorithm (CNN III) processed the spectrogram of ECG segments with 300 time-samples, and this segment length was greater than those which were processed by the other three models. Each data sample of CNN III, therefore, included more information about the ECG signal compared to the other investigated models. This might be a potential factor allowing CNN III to achieve a better performance.

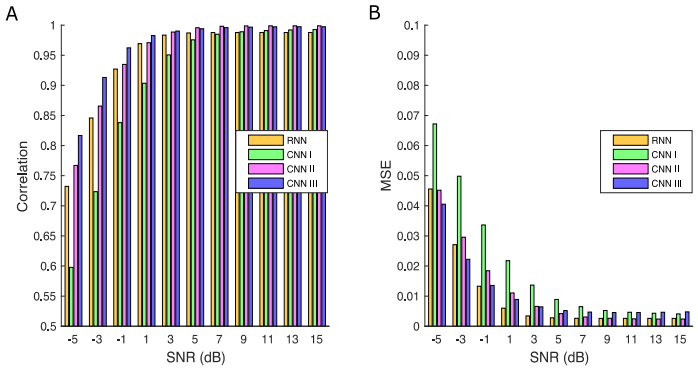


Fig 4. The performance of RNN, CNN I, CNN II, CNN III in terms of Pearson correlation and MSE in case study 3.

Table I. Mean correlation value across low SNRs for the three case studies.

	First study	Second study	Third study
CNN I	0.739	0.763	0.766
CNN II	0.873	0.890	0.885
CNN III	0.913	0.914	0.919
RNN	0.856	0.854	0.869

Table II. Mean MSE value across low SNRs for the three case studies.

	First study	Second study	Third study
CNN I	0.0558	0.0465	0.0431
CNN II	0.0336	0.0267	0.0260
CNN III	0.0258	0.0234	0.0213
RNN	0.0588	0.0372	0.0229

IV. CONCLUSIONS

In this paper, we proposed an algorithm based on CNN to enhance ECG signals in the presence of noise. The CNN took the spectrogram of noisy ECG signals and estimated the enhanced signal in time domain. Three case studies, which varied in size of the training data, were simulated to assess the efficiency of the proposed algorithm (CNN III) and to compare it with three other algorithms. Two objective measures including Pearson correlation and MSE were employed to quantify the performance across different SNRs (-5 dB to 15 dB). Either measures suggested that CNN III had superiority over the three

other models for enhancing ECG signals corrupted by noise. Additionally, our results demonstrated that the performance of CNN III improved when we increased the size of training data, and this enhancement was more remarkable for low SNRs. We also found that the performance of CNN III was degraded when the SNR reduced.

REFERENCES

[1] J. Lee, D. D. McManus, S. Merchant, and K. H. Chon, "Automatic motion and noise artifact detection in Holter ECG data using empirical mode decomposition and statistical approaches," IEEE Transactions on Biomedical Engineering, 59(6), pp. 1499–1506, 2011.

[2] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," Computers in biology and medicine, 38(1), 1-13, 2008.

[3] M. A. Kabir, and C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," Biomedical Signal Processing and Control, 7(5), pp. 481-489, 2012.

[4] M. Alfaouri, K. Daqrouq, "ECG signal denoising by wavelet transform thresholding," American Journal of applied sciences, 5(3), pp. 276–281, 2008.

[5] N. Nikolaev, Z. Nikolov, A. Gotchev, and K. Egiazarian, "Wavelet domain Wiener filtering for ECG denoising using improved signal estimate." 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100). Vol. 6. IEEE, 2000.

[6] N. V. Thakor and Y. S. Zhu, "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection," IEEE transactions on biomedical engineering, 38(8), pp. 785–794, 1991.

[7] M. Chakraborty and S. Das, "Determination of signal to noise ratio of electrocardiograms filtered by band pass and Savitzky-Golay filters," Procedia Technology, 4, pp. 830–833, 2012.

[8] D. Panigrahy and P. K. Sahu, "Extended Kalman smoother with differential evolution technique for denoising of ECG signal," Australasian physical & engineering sciences in medicine, 39(3), pp. 783-795, 2016.

[9] S. Cuomo, G. De Pietro, R. Farina, A. Galletti, and G. Sannino, "A revised scheme for real time ecg signal denoising based on recursive filtering," Biomedical Signal Processing and Control, 27, 134-144, 2016.

[10] B. H. Tracey, and E. L. Miller, "Nonlocal means denoising of ECG signals," IEEE transactions on biomedical engineering, 59(9), pp. 2383-2386, 2012.

[11] K. Jakub, V. Kremen, F. Soucek, and L. Lhotska, "Independent component analysis and decision trees for ECG holter recording de-noising," PLoS One, 9(6), e98450, 2014.

[12] S. Keshavarzi, "Comparison of Two Algorithms for ECG Signal Denoising: A Recurrent Neural Network and A Support Vector Regression," International Journal of Simulation--Systems, Science & Technology, 23(1), 2022.

[13] C. T. Arsene, R. Hankins, and H. Yin, "Deep learning models for denoising ECG signals," European Signal Processing Conference, pp. 1–5, 2019.

[14] K. Antczak, "Deep recurrent neural networks for ECG signal denoising," arXiv preprint arXiv:1807.11551, 2018.

[15] M. Keshavarzi, T. Goehring, R. E. Turner, B. C. J. Moore, "Comparison of effects on subjective intelligibility and quality of speech in babble for two algorithms: A deep recurrent neural network and spectral subtraction," The Journal of the Acoustical Society of America, 145(3), pp. 1493-1503, 2019.

[16] M. Keshavarzi, T. Goehring, J. Zakis, R. E. Turner, and Brian C. J. Moore, "Use of a deep recurrent neural network to reduce wind noise: effects on judged speech intelligibility and sound quality," Trends in hearing, 22, 2331216518770964, 2018.

[17] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," Preprint arXiv:1312.6199, 2014.

[18] M. Abadi, A. Agarwal, P. Barham, E. Brevado, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," arXiv 1603.004467, 2016.

[19] F. Chollet. https://Github.Com/Fchollet/Keras. Last accessed March 2021.

Mapping Researcher Activity based on Publication Data by means of Transformers

Zineddine Bettouche
Deggendorf Institute of Technology
Dieter-Goerlitz-Platz 1
94469 Deggendorf
zineddine.bettouche@th-deg.de

Andreas Fischer
Deggendorf Institute of Technology
Dieter-Goerlitz-Platz 1
94469 Deggendorf
andreas.fischer@th-deg.de

***Abstract*—Modern performance on several natural language processing (NLP) tasks has been enhanced thanks to the Transformer-based pre-trained language model BERT. We employ this concept to investigate a local publication database. Research papers are encoded and clustered to form a landscape view of the scientific topics, in which research is active. Authors working on related topics can be identified by calculating the similarity between their papers. Based on this, we define a similarity metric between authors. Additionally, we introduce the concept of self-similarity to indicate the topical variety of authors.**

***Index Terms*—document similarity, document encoding, BERT, Natural Language Processing, clustering, K-Means, Keyword-extraction.**

I. INTRODUCTION

One of the central themes in natural language processing (NLP) is the task of text representation, which is a kind of rule for converting natural language input information into machine-readable data. Today, the most advanced text models use Transformers to teach how to represent text. Transformers are a type of neural network that are increasingly finding their use in various branches of machine learning, most often in sequence transduction problems, that is, such problems when both the input and output information is a sequence.

BERT (Bidirectional Encoder Representations from Transformers) [1] has stirred up the machine learning world since its publication by highlighting innovative results in a wide range of NLP tasks. The main technical advancement of BERT is the combination of left-to-right and right-to-left training with Transformer's bidirectional training (attention model) for language modeling. The study's findings demonstrate that bidirectionally trained language models can comprehend context and flow of language more deeply than single-direction language models. The data could be clustered more effectively by using the embeddings-format produced from running textual data via BERT as opposed to more conventional clustering techniques like topic extraction and text similarity.

Therefore, we attempt in this work to cluster a dataset of research papers (institute's internal database) into topics using Transformer models. The obtained clusters would provide an overview of the published-research landscape. Having met an acceptable clustering efficiency, authors working in related topics could be linked together, based on the similarity between their papers.

As for this paper's structure, section II introduces the technologies used in it as a general background. Section III discusses the previous papers that dealt with document similarity and clustering, BERT, and keyword extraction. Section IV provides an analysis on the data used in this work. Section V discusses the methodology and the overall implementation of the processes implemented for this paper. Section VI presents the experiments done in this work, their implementation, and the rationale behind them. Finally, section VII concludes the work and sets up the future developments, which could be built on our findings, or, in the least, complement it in certain areas.

II. BACKGROUND

In this section, the main techniques used in this paper are discussed.

A. Transformer Models

A deep learning technique for natural language processing (NLP) called Bidirectional Encoder Representations from Transformers (BERT) aids artificial intelligence (AI) programs in comprehending the context of ambiguous words in text.

By evaluating text in the left-to-right and right-to-left directions, BERT-using applications can forecast the accurate meaning of a synonym. Deep learning neural networks may use unsupervised learning methods to develop new NLP models thanks to BERT's bidirectionality, a masking strategy, and learning how to anticipate the meaning of an ambiguous term. This natural language understanding method (NLU) is so effective that Google advises customers to use it to train an innovative question and answer system in a short amount of time if there is sufficient training data available.

B. UMAP: Dimensionality Reduction

Like t-SNE, the dimensionality-reduction method known as Uniform Manifold Approximation and Projection (UMAP) [2] can be utilized for visualization as well as generic non-linear dimensionality reduction. The following suppositions about the data form the basis of the algorithm:

- The Riemannian metric is constant locally, and the data are uniformly distributed on the Riemannian manifold.
- The manifold is connected locally.

These presumptions allow one to construct a fuzzy topological model of the manifold. Finding the embedding involves looking for a low-dimensional projection of the data that has the most similar fuzzy topological structure to the original data.

C. K-means Clustering

One of the most straightforward and well-liked unsupervised machine learning methods is K-means clustering [3]. Unsupervised algorithms typically draw conclusions from

datasets using only input vectors without considering predetermined or labelled results.

The K-means algorithm finds k centroids, keeps the centroids as small as feasible, and then assigns each data point to the closest cluster. Finding the centroid is what 'means' in the K-means algorithm indicates: averaging the data. The K-means technique in data mining uses a first set of centroids that are randomly chosen as the starting points for each cluster to process the learning data.

The program then performs iterative (repetitive) calculations to optimize the positions of the centroids. It ends creating and optimizing clusters when either the centroids have stabilized, or the defined number of iterations has been achieved.

III. RELATED WORK

Concerning related work, Beltagy et al. have published SciBERT [4], a pretrained language model for scientific text based on BERT, in a paper titled 'SciBERT: A Pretrained Language Model for Scientific Text'. They evaluated SciBERT on a suite of tasks and datasets from scientific domains. SciBERT significantly outperformed BERT-Base and achieves new SOTA results on several of these tasks, even compared to some reported BIOBERT results on biomedical tasks.

In a paper [5] titled 'Aspect-based Document Similarity for Research Papers' published on a related subject, Ostendorff et al. apply pairwise multi-label multi-class document classification to scientific papers to determine an aspect-based document similarity score. Based on the paper's title and abstract, the investigated models are trained to forecast citations and the recognized label. Over two scientific corpora, they assess the Transformer models BERT, CovidBERT, SciBERT, ELECTRA, RoBERTa, and XLNet along with an LSTM baseline. Overall, SciBERT outperformed all other models in the tests. SciBERT predicted the aspect-based document similarity with F1-scores of up to 0.83 despite the difficult assignment. Transformers are highly adapted to accurately compute the aspect-based document similarity for research papers, according to their empirical investigation.

Chandrasekaran and Mago conducted a survey [6] titled 'Evolution of Semantic Similarity - A Survey' in which they stated that, measuring semantic similarity between two text snippets has been one of the most challenging tasks in the field of Natural Language Processing. They concluded that, most recent hybrid methods have shown promising results over other independent models. While the focus of recent research is shifted towards building more semantically aware word embeddings, and the transformer models have shown promising results, the need for determining a balance between computational efficiency and performance is still a work in progress.

The problem of semantic textual similarity in medical data was addressed by Kades et al. in a paper [7] titled 'Adapting Bidirectional Encoder Representations from Transformers (BERT) to Assess Clinical Semantic Textual Similarity: Algorithm Development and Validation Study.' The authors developed three methods to address this issue. They suggested enhancing BERT with new features and comparing several regression models based on the BERT result and other features. The use of M-Heads and an effort to automatically extrapolate medical knowledge from the training data was another concept. They noticed the underlying dataset significantly impacted the effectiveness of the techniques.

'Measurement of Semantic Textual Similarity in Clinical Texts: Comparison of Transformer-Based Models' is an article published in the journal Clinical Text by Yang et al. [8], in which they showed how to measure clinical STS using transformer-based models, and created a system that can employ several transformer algorithms. In comparison to other transformer models, their experiment findings demonstrate that the RoBERTa model performed the best. The study also showed how well transformer-based models performed when used to evaluate the semantic similarity of clinical content.

In another paper [9] titled 'Extracting Keywords from Publication Abstracts for an Automated Researcher Recommendation System' Kretschmann et al. presented a keyword assignment system based on an older version of the DIT publication database. It handles low volume data and missing keywords by extending the data volume using information from online publication databases, extending the total volume to 6500 items. A prototype keyword assignment system was built, that uses random oversampling in preprocessing and LightGBM as classifier with binary relevance as transformation method. As an enhancement of this system, Transformer models could be used to uncover relations between papers and authors on another level.

Therefore, our work investigates the use of Transformer models, SciBERT, in the last-mentioned recommendation system, as an attempt to enhance the quality of recommendations, through understanding the relations between abstracts deeply, using BERT models and K-means clustering.

IV. EXPLORATORY DATA ANALYSIS

The data used in this work are encoded in a JSON file with a size of 11.735 KB. Contained are 7548 references of several types: the file includes references to newspaper articles, blog entries, talks, along with scientific articles such as conference papers or journal articles. Only the latter ones come with abstracts, comprising 1500 documents, chosen as the data basis for this investigation.

Each of the selected entries has at least a title, a list of authors, a date, and an abstract. Figure 1 shows an example of an entry. Each author is marked by a name. It is possible to distinguish between internal authors (i.e., employees) and external authors: Internal authors are identified by their e-mail address. External authors, on the other hand, cannot reliably be distinguished. Two entries with the same name as an external author might stem from the same person or from two persons with the same name.

The character count of every abstract present in the database was calculated, and the overall distribution is presented in Figure 2. Note that the distribution of data represents a long-tail distribution, with most documents falling in the interval of 500 to 1500 characters per abstract-text. The entry at the far right is an outlier, in which the entire paper content is recorded in the abstract field.

```
{
    "id":"019844ce-e696-0c48-a2ac-1821047639e0",
    "abstractText":"A new facility designed to perf...",
    "title":"Calibration facility for airborne imaging spectrometers",
    "date":"30.06.2009",
    "referenceAuthors":[
        {"person":{"firstname":"P.","lastname":"G[redacted]"},
            "notes":null,"rank":0},
        {"person":{"firstname":"J.","lastname":"F[redacted]"},
            "notes":null,"rank":1},
        {"person":{"firstname":"P[redacted]","lastname":"S[redacted]"},
            "notes":"p[redacted]@th-deg.de","rank":2},
        {"person":{"firstname":"H.","lastname":"S[redacted]"},
            "notes":null,"rank":3}
    ]
}
```


Fig. 1. Paper-Object Example

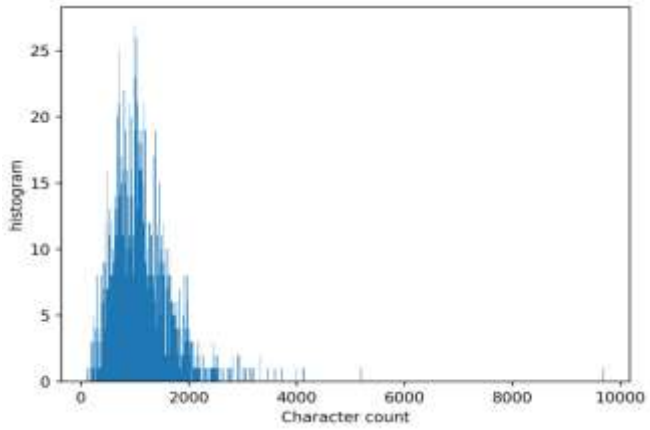


Fig. 2. Histogram plot of the character count in each abstract

V. METHODOLOGY

In this section the general approach is described, followed by a discussion of the metric used to calculate author distances.

A. General Approach of Processing

The goal of the implementation is to cluster the encodings of the abstracts and visualize it, to assess the performance of BERT when dealing with long textual data. The starting point of the implementation is getting the data as input. The data, as previously mentioned, is in the form of paper objects. Every object contains an ID and an abstract text. Therefore, the abstract-texts are encoded with a BERT encoder, which gives in return the vector format (768-dimension) for each text. The obtained vectors' dimensions must be reduced into 2D for visualization to take place. Hence, K-means can be applied on the 768D vectors, and then the dimensionality-reduction is done for graphing. The implementation overview is displayed in Figure 3.

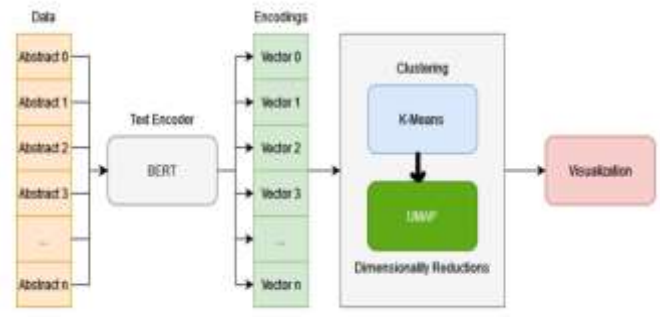


Fig. 3. Implementation Overview

To obtain similar papers, Euclidean distance is calculated between the encodings of each abstract-text. This attempt can display the efficiency of clustering textual data based on its BERT encoding, rather than traditional words sequence assessment.

B. Calculating distances between authors

The embedding of documents produced by transformer architectures such as the BERT family has, for the use case outlined in this paper, the advantage that a distance between authors can be directly derived from document distances. We define the distance between two authors as the mean pair-wise distance of their respective papers. I.e., let P1 be the set of papers by author 1 and P2 likewise be the set of papers by author 2. Then the distance between authors 1 and 2 is defined as:

$$\frac{\sum_{p_1 \in P_1} \sum_{p_2 \in P_2} dist(p_1, p_2)}{|P_1| \cdot |P_2|} \quad (1)$$

There are a few border cases to be discussed here. First, there will be a distance for every pair of authors, provided both have published a paper. If an author is recorded in the database without any publication, no useful distance can be computed to any other author. This is to be expected.

Second, papers may appear both in P1 and P2—for co-authors. In this case, the corresponding distance for that paper will, of course, be zero. This lowers the overall distance, and is an expected effect.

Third, an author can be compared to itself. The overall result would not be zero (as could be expected), but instead report the average distance among his papers. We argue that this is useful, as it provides a measure of the self-similarity, i.e., the topical variety of the papers from a given author.

VI. EXPERIMENTS

In this section, the experiments done throughout this work and both the rationale behind them, and the results obtained from them, are discussed.

A. SciBERT: From Abstracts To 768D-Vectors

As we have already mentioned in the background and related-work sections, SciBERT scores best when dealing with scientific papers. We attempt now to obtain vector representations of each of the abstracts in the data, which are visualized after being reduced to 2D points by UMAP. Figure 4 shows the resulting scatter graph.

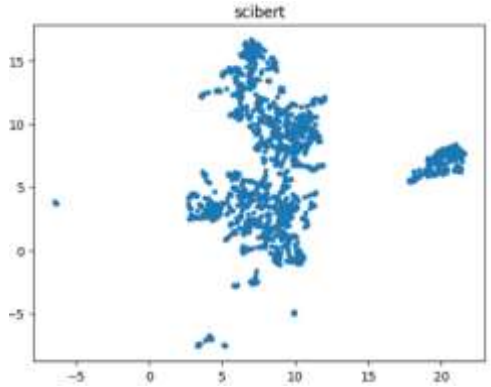


Fig. 4. 2D representations of 768D-encodings resulting from SciBERT encoding the abstracts

An initial remark could state that, the points experience a few condensations. This indicates the presence of clusters in the data. Therefore, we proceed to cluster the data using K-means.

B. Initial Raw Clustering: K-means Application

Before we indulge in describing how the different techniques in this work (such as K-means and UMAP) are used together, it is worth to mention that the process of clustering is implemented in our code, in a way that automates the discovery of clusters-number in the data, instead of assuming it. The program loops n times (n is in the range [10, 30]) and selects the number of clusters attached to the best silhouette score. The silhouette value [10] is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation).

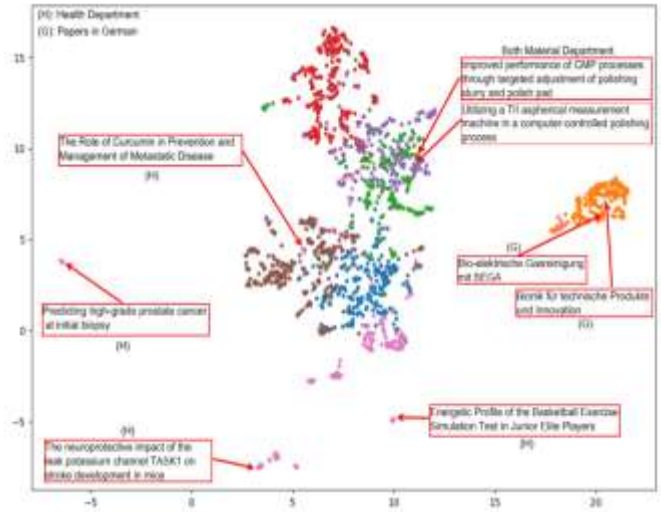


Fig. 5. Detailed points (768D-vector clustering)

After encoding the abstract-texts and obtaining the 768Dvectors, they can be clustered first, then mapped into 2Dvectors, which are then visualized. A hypothesis can state that clustering directly the 768D-vectors would give meaningful clusters, on the condition that the similarity between two abstract-texts is inversely proportional to the Euclidean distance between each of their vectors. It is critical to understand that the compactness of clusters does not imply the efficiency the encodings of the abstract-texts. Figure 5 shows the obtained clusters and the titles of some interesting cases, that we would like to comment on as follows:

- In the case of health-related papers (pink), although some points were farther from the centroid of their cluster than most of that cluster's points, K-means was accurate enough to assign them properly. This indicates that, the inaccurate representation of the distance on the graph (i.e., when it does not indicate the actual similarity between two paper), is due to the nature of UMAP, when it attempts to map 768D-vectors into a 2D points. This is regarded as a positive result concerning the relevancy of the 768D-vectors to one another when their original abstract-texts are similar.
- It is remarked from the distinct cluster on the righthand side (orange) that, the encoder used (SciBERT) has separated papers written in German from papers written in English. This can be further investigated by extracting the German papers, and running the encoding process on them separately.
- Concerning the other clusters, we can say that, both the distances between the points and their assigned clusters were reasonably accurate; as they were grouped in the form of: optics in red, networks in blue, and media in brown (including image/video processing).
- Finally, we have reason to believe that the automated silhouette method, which determines clusters-number for K-means, has made an unnecessary split, resulting in two clusters (purple/green) that belong to the same department (Material Department). Going further, clusters-number can be manually defined at this point, which will shed lighter onto their global cluster, when rerunning K-means.

Therefore, in this paper, we would proceed by casting the German papers aside, calculating cluster-metrics, and finally extracting keywords for each obtained cluster.

C. Clusters: Metrics and Keywords

As a first step in extracting the keywords of each cluster, and potentially, the topic of each, the German data was separated from the rest of the data (mostly English). The latter was adjusted further, based on what we perceived as one cluster divided unnecessarily in half. Figure 6 shows the adjusted clusters and a few keywords of each cluster.

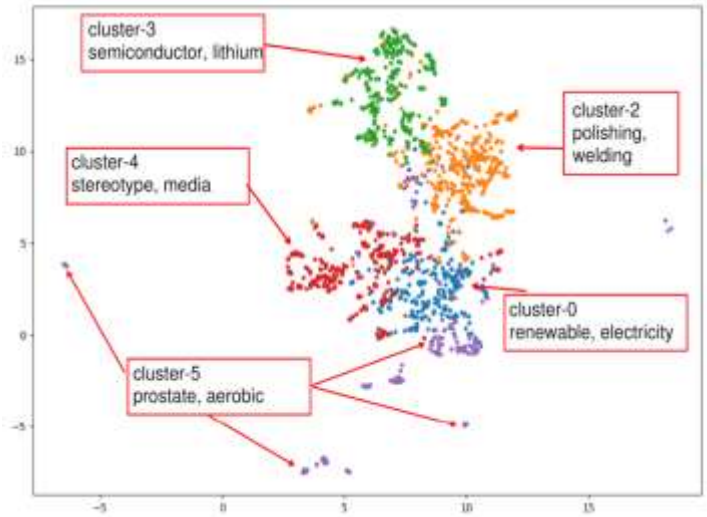


Fig. 6. Adjusted clusters

To finally assess the clusters with common metrics, some of the clustering metrics were calculated and are as follows:

- Silhouette: 0.103
- Calinski-Harabasz: 129.237
- Davies-Bouldin: 3.085

Keywords could be extracted using the KeyBERT library, we then attempt to assign a topic to such keywords accordingly. In addition, a further evaluation of the clusters can be set in terms of the radius and the standard-deviation of each cluster. Table I shows the result of our calculations and keyword-extraction.

Table 1
CLUSTERS-METRICS TABLE

English Papers					
label	768D-radius	768D-std	points-count	keywords	topic
0	4.903	0.708	251	renewable, photovoltaic, electricity	Power
1	3.846	1.080	178	german papers (ignored)	german papers (ignored)
2	5.048	0.830	369	polishing, welding, piezoelectric	Production/Industry
3	4.773	0.647	275	semiconductor, nanowire, lithium	Material
4	4.801	0.694	279	stereoscopic, 3dtv, multimedia	Media
5	5.140	0.832	148	prostate, aerobic, schizophrenia	Health

D. Distances between Authors

As explained before, how close two authors are, is related to how similar are their papers. Therefore, to evaluate whether the distances, obtained in this work between authors, are reasonable, we have investigated the difference between three cases: the average distance each other has to himself (self-distance), the average distance between co-authors (coauthor-distance), and the total average distance between all authors. Figure 7 shows the author-distance distributions in the three mentioned cases. As expected, self-distance and coauthors-distance are close, with coauthors-distance slightly higher. However, there is a significant difference to the average distance between all authors. Ergo, the defined distance metric for authors captures semantic relationship and, as such, is a useful tool to indicate the degree of overlapping research for two given authors.

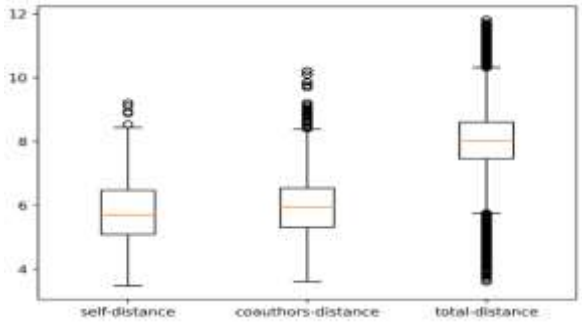


Fig. 7. Author-distance distributions in the three mentioned cases

For further analysis, only active authors with ten publications or more are considered to avoid bias resulting from a few publications. Figure 8 shows the distribution of paper count per author. As seen, most authors have 1 to 15 published papers, whereas the highest paper count is 106. Logarithmic scale was used on the y-axis, due to the stark difference between authors with number of papers lower and higher than 10.

To further visualize the self-distance of authors, we excluded the authors having fewer than 10 published papers (100 out of 4426 authors), because having very few papers could distort the distance calculations. Figure 8 shows the distribution of paper count per author.

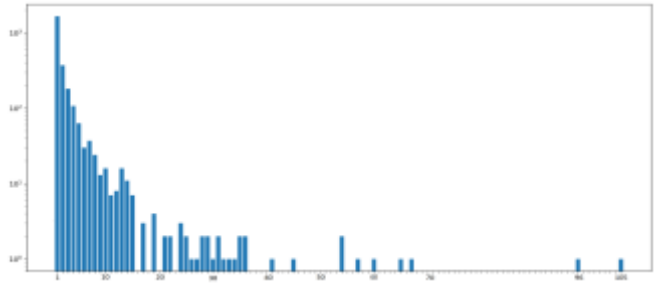


Fig. 8. Distribution of paper count per author

German-papers were again excluded to avoid confusion between topical clusters and language clusters. The two authors with the highest vs. the lowest self-similarity are selected and visualized in Figure 9. The author represented by the red blobs (highest self-distance with 30 papers) is active in several topical areas, such as production, industry, and materials engineering. However, the author represented by the yellow crosses (lowest self-distance with 10 papers) is focused on the narrower field of Materials.

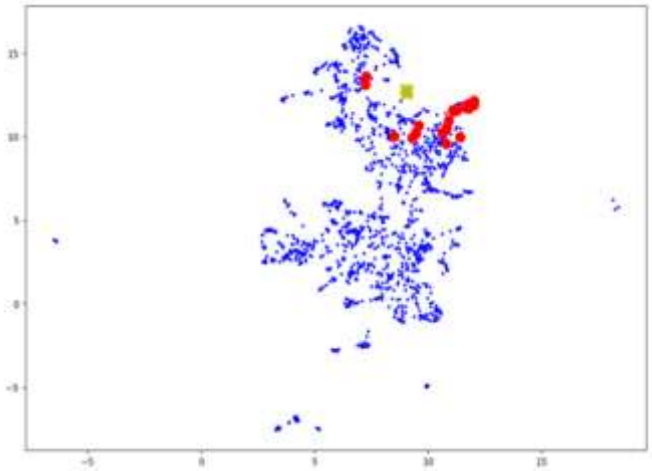


Fig. 9. The papers of the two authors having the highest and lowest self-distances (red and yellow, respectively)

VII. CONCLUSION

In this paper, SciBERT encodings were used to obtain vector representations of paper abstracts. Semantic similarity between papers is encoded in the pairwise distance of vectors. With K-Means, topical research clusters were identified in a publication database. Visualization using UMAP highlights the topical areas. By extracting keywords from the clustered papers, the research areas could be identified. Based on distance between papers, a distance metric between authors was introduced. The data indicates that this distance metric is a useful tool to indicate topical relationships between authors.

For this paper, a cluster of purely German articles was ignored to avoid confusion between topical and linguistical similarity. Future work includes developing an approach that can reliably handle multilingual data. Further investigation is also needed in how to apply the author similarity metric in a recommender system. Finally, more research is required on the keyword extraction mechanism used for cluster labeling: The current approach is based on manual investigation of extracted keywords. An ontology-based system might be able to automate this process and, such, scale to larger publication databases.

ACKNOWLEDGEMENT

This paper has received funding from the state of Bavaria in the context of project SEMIARID, funding no. DIK-2104-0067//DIK0299/01.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for

language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[2] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2018, cite arxiv:1802.03426Comment: Reference implementation available at http://github.com/lmcinnes/umap. [Online]. Available: http://arxiv.org/abs/1802.03426

[3] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," JSTOR: Applied Statistics, vol. 28, no. 1, pp. 100–108, 1979.

[4] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3615–3620. [Online]. Available: https://aclanthology.org/D19-1371

[5] M. Ostendorff, T. Ruas, T. Blume, B. Gipp, and G. Rehm, "Aspect-based document similarity for research papers," in Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 6194–6206. [Online]. Available: https://aclanthology.org/2020.coling-main.545

[6] D. Chandrasekaran and V. Mago, "Evolution of semantic similarity—a survey," ACM Comput. Surv., vol. 54, no. 2, feb 2021. [Online]. Available: https://doi.org/10.1145/3440755

[7] K. Kades, J. Sellner, G. Koehler, P. M. Full, T. Y. E. Lai, J. Kleesiek, and K. H. Maier-Hein, "Adapting bidirectional encoder representations from transformers (Bert) to assess clinical semantic textual similarity: Algorithm development and validation study," JMIR Med Inform, vol. 9, no. 2, p. e22795, Feb 2021. [Online]. Available: https://medinform.jmir.org/2021/2/e22795

[8] X. Yang, X. He, H. Zhang, Y. Ma, J. Bian, and Y. Wu, "Measurement of semantic textual similarity in clinical texts: Comparison of transformer-based models," JMIR Med Inform, vol. 8, no. 11, p. e19735, Nov 2020. [Online]. Available: http://medinform.jmir.org/2020/11/e19735/

[9] M. Kretschmann, A. Fischer, and B. Elser, "Extracting keywords from publication abstracts for an automated researcher recommendation system," Digitale Welt, vol. 4, pp. 20–25, 01 2020.

[10] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53–65, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

Localization of the Solid-Solid Interfaces in A Three-Layer Material

Guillermo F. Umbricht
Dpto. De Matematica, FCE
Universidad Austral-CONICET
Rosario, Argentina
guilleungs@yahoo.com.ar

Diana Rubio
CEDEMA-Escuela de Ciencia y Tecnología
ITECA-UNSAM-CONICET
Buenos Aires, Argentina
drubio@unsam.edu.ar

Domingo A. Tarzia
Dpto. De Matematica, FCE
Universidad Austral-CONICET
Rosario, Argentina
dtarzia@austral.edu.ar

***Abstract*— In this work, an analytical technique is proposed for the simultaneous determination of the interfaces positions between two materials in a three-layer body. The estimate is made from two noisy temperature measurements, one in the middle of the body and the other on its right edge. A bound for the estimation error is found, which depends on the noise in the temperature measurements. Moreover, the local dependency of the estimated parameters on data is study by means of an elasticity analysis. Numerical examples with different characteristics are used to show the performance of the proposed method.**

Keywords*— *inverse problem, parameter estimation, multilayer material, heat transfer

I. INTRODUCTION

Heat transfer problems in multilayer or solid-solid interface materials have been extensively studied in recent years due to its multiple and different applications found in science and engineering [1], [2], [3], [4], [5], [6]. These problems have direct applications in several industries, such as metallurgical [7], technology and electronics [8], automotive [9], aerospace [10] and aviation [11], among the most important.

The location of the solid-solid interface in multilayer materials, by using data from heat transfer processes, has different applications in the field of engineering. It can be applied in problems that arise in chemical engineering for the determination of impurities [12], [13] and for the separation of metals by means of polymers [14], [15]. Also in pharmaceutical engineering, for the identification of impurities in medicines [16] and in the cosmetic industry [17], to name a few. This article proposes the simultaneous determination of the two contact points in a three-layer body.

II. DIRECT PROBLEM

A three-layer material having consecutive sections of homogeneous and isotropic materials, namely *A, B and C*, with constant thermal diffusivity coefficients denoted by α_A^2, α_B^2, α_C^2 (m^2/s), is considered. A process of heat transfer through the material is analyzed assuming that it is one-dimensional and it can be modeled by the transport of thermal energy in a bar of length L (m) and diameter d (m) totally isolated on its lateral surface, where $L \gg d$.

The length of the left section (occupied by the material *A)* is denoted by l_1 (m), the middle part (occupied by the material *B)* has a length $l_2 - l_1\,(m)$ and the length of the right section (occupied by the material *C*) is $L - l_2$ (m).

The sections are assumed to be perfectly assembled (no cracks or roughness are present), so that there is no thermal resistance at the interfaces. Therefore, temperature and thermal flow are continuous on the entire bar. Moreover, it is also assumed that at the left edge of the bar, the temperature is maintained constant at F (ºC) while the right edge of the bar remains free, in contact with the fluid, giving rise to the phenomenon of convection.

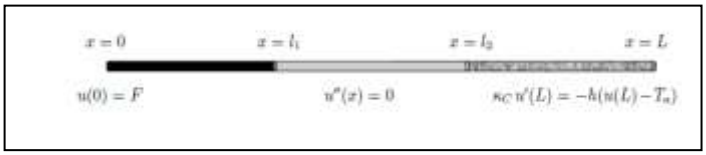


Fig.1. Diagram of the three-layer material

The problem described above can be modeled using the following system of elliptic equations with the boundary and interface conditions:

$$\begin{cases} u''(x) = 0, & 0 < x < l_1, \\ u''(x) = 0, & l_1 < x < l_2, \\ u''(x) = 0, & l_2 < x < L, \\ u(x) = F, & x = 0, \\ u(x^-) = u(x^+), & x = l_1, \\ u(x^-) = u(x^+), & x = l_2, \\ \kappa_A\, u'(x^-) = \kappa_B\, u'(x^+), & x = l_1, \\ \kappa_B\, u'(x^-) = \kappa_C\, u'(x^+), & x = l_2, \\ \kappa_C\, u'(x^-) = -h(u(x) - T_a) & x = L, \end{cases} \tag{1}$$

where u denotes the temperature, κ_i the thermal conductivity of the i-th material, h the convective heat transfer coefficient, T_a the temperature medium $(T_a < F)$, and

$$u(x^-) = \lim_{s \to x^-} u(s), \qquad u(x^+) = \lim_{s \to x^+} u(s),$$

$$u'(x^-) = \lim_{s \to x^-} u'(s), \qquad u'(x^+) = \lim_{s \to x^+} u'(s). \tag{2}$$

An analytical expression is sought for the temperature function u. Since its second derivative vanishes in the intervals $(0, l_1)$, (l_1, l_2) and (l_2, L), it turns out that u is piecewise linear, i.e., linear on each interval. Namely,

$$u(x) = \begin{cases} b_1 + a_1 x, & 0 < x < l_1, \\ b_2 + a_2 x, & l_1 < x < l_2, \\ b_3 + a_3 x, & l_2 < x < L, \end{cases} \quad (3)$$

where a_1, a_2, a_3, b_1, b_2 and b_3 are constant coefficients that depend on the parameters of the problem and are determined from the boundary and interface conditions of the System (1). Hence, the following linear matrix equation is obtained,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & l_1 & -l_1 & 0 \\ 0 & 1 & -1 & 0 & l_2 & -l_2 \\ 0 & 0 & 0 & \kappa_A & -\kappa_B & 0 \\ 0 & 0 & 0 & 0 & \kappa_B & -\kappa_C \\ 0 & 0 & h & 0 & 0 & \kappa_C + hL \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} F \\ 0 \\ 0 \\ 0 \\ 0 \\ hT_a \end{pmatrix}. \quad (4)$$

Solving the system (4) for a_1, a_2, a_3,b_1, b_2 and b_3 it follows that:

$$u(x) = \begin{cases} F + \zeta x, & 0 \le x \le l_1, \\ F + \zeta \left[l_1 \left(1 - \frac{\kappa_A}{\kappa_B} \right) + \frac{\kappa_A}{\kappa_B} x \right], & l_1 \le x \le l_2, \\ F + \zeta \left[l_1 \left(1 - \frac{\kappa_A}{\kappa_B} \right) l_2 \left(\frac{\kappa_A}{\kappa_B} - \frac{\kappa_A}{\kappa_C} \right) + \frac{\kappa_A}{\kappa_C} x \right], & l_2 \le x \le L, \end{cases} \quad (5)$$

Where,

$$\zeta = -\frac{F - T_a}{L\,\zeta_0},$$

With,

$$\zeta_0 = \frac{\kappa_A}{hL} + \frac{l_1}{L}\left(1 - \frac{\kappa_A}{\kappa_B}\right) + \frac{l_2}{L}\left(\frac{\kappa_A}{\kappa_B} - \frac{\kappa_A}{\kappa_C}\right) + \frac{\kappa_A}{\kappa_C}. \quad (6)$$

Let us X denote a particular homogeneous material. Consider three-layer materials of the form X-Copper-Nickel and Lead-Silver-X of length L = 3 m and solid-solid interfaces located at l_1 = 0.8 m; l_2 = 2.1 m. The solution to the heat transfer process described by (1) with F = 100 ºC and T_a = 25ºC for different materials X are plotted in Fig. 2. The heat transfer coefficient (h) is determined as in [18], assuming that the convective fluid is air at an atmosphere of pressure. The termal conductivity values for the materials considered here were taken from [19].

In Fig. 2 it can be seen the changes in the slope at the interface points of the piecewise linear temperature functions for the different materials. Moreover, it can be observed that, in the second and third sections of the material, the temperature profiles are "almost" parallel segments which are due to the fact that the second and third materials are the same (Copper-Nickel) in all cases. It is also observed that, for Lead-Copper-Nickel the stationary temperature at the right edge of the bar is $u(L)$ = 80.5 ºC and for the Silver-Copper-Nickel bar we have that $u(L)$=90.5 ºC, approximately. This difference in temperature is consistent to the thermal property of the materials, since conductive materials favor heat conduction, and hence, the temperature reaches higher values. Lastly, since all materials considered for this example are of the form X-Copper-Nickel, the material X will be decisive in terms of the temperature reached by the three-layer material on its right edge.

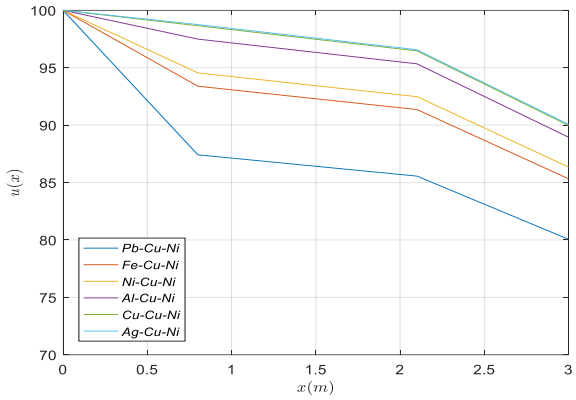


Fig. 2. Temperature Profiles for three-layer materials X-Cu-Ni.

Now, we consider three-layer materials of the form Lead-Silver-X of length L = 3 m and solid-solid interfaces located at l_1 = 0.3 m; l_2 = 0.4 m. The solution to the heat transfer process described by (1) with F = 100 ºC and T_a = 25ºC for different materials X are plotted in Fig. 3. It is observed that, in the first and second sections, the stationary temperature profiles are similar due to the fact the materials at these sections are the same. Now, the material X of the third section determines the temperature value reached at the right edge, being higher for more conductive materials X. For instance, for X=Lead, that is, a three-layer material Lead-Silver-Lead, the temperature at the right edge of the bar is $u(L)$ = 72.8 ºC while for X=Silver, i.e. Lead-Silver-Silver, $u(L)$ = 90.7 ºC.

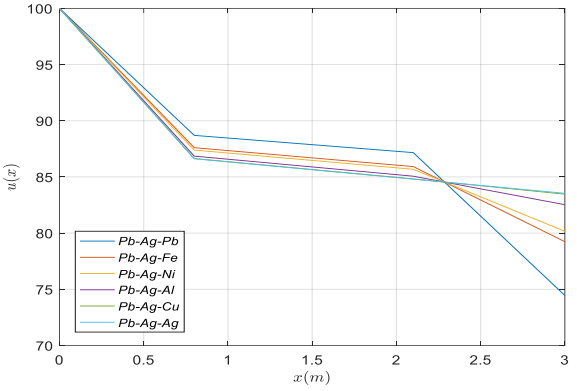


Fig.3. Temperature profiles for three-layer materials Pb-Ag-X.

III. INVERSE PROBLEM

In this section, it is addressed the problem of the localization of the two solid-solid interfaces, (*A*-*B*) and (*B*-*C*) [20]. This inverse problem can be stated as the simultaneous estimation of l_1 (m) and l_2 (m). In this work, two noisy temperature data, one in the middle of the bar ($x = L/2$) and another at the right edge of it ($x = L$), are used to solve the identification problem.

An analytical expression is obtained for the approximation of the location of each interface point. A bound for the estimation errors is derived which depends on the error in the temperature data. Moreover, the local dependency of the estimated parameters on the data is study by means of an

elasticity analysis. The use of this technique is shown through a numerical example.

A. Parameter estimation

Assuming that the interface points satisfy the inequality

$$l_1 < L/2 < l_2 \tag{7}$$

the estimation of l_1, l_2 is made based on two noisy temperature data, T_1^ϵ and T_2^ϵ, at the middle point ($x = L/2$) and at the right edge ($x = L$), respectively, imposing the following conditions:

$$\begin{cases} |T_1 - T_1^\epsilon| = \epsilon\,(F - T_a), \\ |T_2 - T_2^\epsilon| = \epsilon\,(F - T_a), \end{cases} \tag{8}$$

where $u(L/2) = T_1$ (ºC) and $u(L) = T_2$ (ºC) are the exact values and ϵ denotes the noise level, which represents the error introduced by the measuring instruments, among other possible errors in data.

Using the expressions in (4) to calculate $u(L/2)$ and $u(L)$ it follows that:

$$\begin{cases} T_1 = F + \zeta\left[l_1\left(1 - \dfrac{\kappa_A}{\kappa_B}\right) + \dfrac{\kappa_A}{\kappa_B}\dfrac{L}{2}\right], \\ T_2 = F + \zeta\left[l_1\left(1 - \dfrac{\kappa_A}{\kappa_B}\right) + l_2\left(\dfrac{\kappa_A}{\kappa_B} - \dfrac{\kappa_A}{\kappa_C}\right) + L\dfrac{\kappa_A}{\kappa_C}\right]. \end{cases} \tag{9}$$

From the expressions given in Eq. (7), operating algebraically it is possible to obtain the following expressions for the parameters

$$\begin{cases} l_1 = \dfrac{\kappa_A\,\kappa_B}{\kappa_B - \kappa_A}\left(\dfrac{1}{h}\dfrac{F - T_1}{T_2 - T_a} - \dfrac{L}{2\,\kappa_B}\right), \\ l_2 = \dfrac{\kappa_B\,\kappa_C}{\kappa_C - \kappa_B}\left[\dfrac{1}{h}\dfrac{T_1 - T_2}{T_2 - T_a} + \left(\dfrac{1}{2\,\kappa_B} - \dfrac{1}{\kappa_C}\right)\right]. \end{cases} \tag{10}$$

The necessary and sufficient conditions for the existence of the solution to the stated inverse problem can be derived as follows. Using the expression given in (10) in the inequality (5) four cases must be considered

Case 1: $(\boldsymbol{\kappa_A < \kappa_B})$**,** for this case we get:

$$F - \tfrac{L\,h}{2\,\kappa_A}(T_2 - T_a) < T_1 < F - \tfrac{L\,h}{2\,\kappa_B}(T_2 - T_a). \tag{11}$$

Case 2: $(\boldsymbol{\kappa_A > \kappa_B})$**,** for this case we get:

$$F - \tfrac{L\,h}{2\,\kappa_B}(T_2 - T_a) < T_1 < F - \tfrac{L\,h}{2\,\kappa_A}(T_2 - T_a). \tag{12}$$

Case 3: $(\boldsymbol{\kappa_B < \kappa_C})$**,** for this case we get:

$$T_2 + \tfrac{L\,h}{2\,\kappa_C}(T_2 - T_a) < T_1 < T_2 + \tfrac{L\,h}{2\,\kappa_B}(T_2 - T_a). \tag{13}$$

Case 4: $(\boldsymbol{\kappa_B > \kappa_C})$**,** for this case we get:

$$T_2 + \tfrac{L\,h}{2\,\kappa_B}(T_2 - T_a) < T_1 < T_2 + \tfrac{L\,h}{2\,\kappa_C}(T_2 - T_a). \tag{14}$$

Combining the Eq's. (9) - (12) we obtain:

$$T_m(T_2) < T_1 < T_M(T_2), \tag{15}$$

where

$$\begin{cases} T_m = \max\{F - \frac{L\,h}{2\,\kappa_m^1}(T_2 - T_a), T_2 + \frac{L\,h}{2\,\kappa_M^2}(T_2 - T_a)\}, \\ T_M = \min\left\{F - \frac{L\,h}{2\,\kappa_M^1}(T_2 - T_a), T_2 + \frac{L\,h}{2\,\kappa_m^2}(T_2 - T_a)\right\}, \end{cases} \tag{16}$$

with

$$\begin{cases} \kappa_m^1 = \min\{\kappa_A, \kappa_B\}, \\ \kappa_M^1 = \max\{\kappa_A, \kappa_B\}, \\ \kappa_m^2 = \min\{\kappa_B, \kappa_C\}, \\ \kappa_M^2 = \max\{\kappa_B, \kappa_C\}. \end{cases} \tag{17}$$

Note that the Eqs. (15-17) indicate the relationship that T_1 and T_2 must meet for the estimation problem to have a solution.

Remark: For the estimation of l_1 and l_2 we have assumed that $l_1 < L/2 < l_2$. Analogous results are obtained for different situations. In a real specific application, although the exact locations of the interface points are unknown, it is assumed that some information about the interface positions is available in order to decide where the temperature data can be taken for the estimation.

B. Error analysis

In this section, an analytical expression is obtained for the bounds of the error made when approximating the locations of the interface points l_1 and l_2, using two measurements noisy temperature signals T_1^ϵ and T_2^ϵ. Regardless of whether the measured temperatures T_1^ϵ and T_2^ϵ meet the necessary and sufficient conditions given by Eqs. (13)-(15), there will be an error in the estimation of the contact points that will depend directly on the error *(ϵ)* in temperature measurements.

The exact temperature values T_1, T_2 and their respective measurements $T_1^\epsilon, T_2^\epsilon$ are considered. Then we have:

$$\left|l_1 - \widehat{l_1}\right| = \frac{\kappa_A\,\kappa_B}{h\,|\kappa_B - \kappa_A|}\left|\frac{F - T_1}{T_2 - T_a} - \frac{F - T_1^\epsilon}{T_2^\epsilon - T_a}\right|. \tag{18}$$

$$\left|l_2 - \widehat{l_2}\right| = \frac{\kappa_B\,\kappa_C}{h\,|\kappa_C - \kappa_B|}\left|\frac{T_1 - T_2}{T_2 - T_a} - \frac{T_1^\epsilon - T_2^\epsilon}{T_2^\epsilon - T_a}\right|. \tag{19}$$

Operating algebraically on the Eqs. (16)-(17) and Eq. (6), the following expressions are obtained:

$$\left|l_1 - \widehat{l_1}\right| = \frac{\kappa_A\,\kappa_B}{h\,|\kappa_B - \kappa_A|}\,\frac{F - T_a}{T_2 - T_a}\left(1 + \frac{F - T_1^\epsilon}{T_2^\epsilon - T_a}\right)\epsilon, \tag{20a}$$

$$\left|l_2 - \widehat{l_2}\right| = \frac{\kappa_B\,\kappa_C}{h\,|\kappa_C - \kappa_B|}\,\frac{F - T_a}{T_2 - T_a}\left(1 + \frac{T_1^{\epsilon} - T_a}{T_2^{\epsilon} - T_a}\right)\epsilon, \qquad (20b)$$

or equivalently

$$\left|l_1 - \widehat{l_1}\right| = \frac{1}{M_1}\,\frac{\kappa_A\,\kappa_B}{h\,|\kappa_B - \kappa_A|}\left(1 + \frac{F - T_1^{\epsilon}}{T_2^{\epsilon} - T_a}\right)\epsilon, \qquad (21a)$$

$$\left|l_2 - \widehat{l_2}\right| = \frac{1}{M_1}\,\frac{\kappa_B\,\kappa_C}{h\,|\kappa_C - \kappa_B|}\left(1 + \frac{T_1^{\epsilon} - T_a}{T_2^{\epsilon} - T_a}\right)\epsilon, \qquad (21b)$$

where $M_1\ \varepsilon\ (0,1)$ is a constant that satisfies:

$$M_1 < \frac{T_2 - T_a}{F - T_a}. \qquad (22)$$

Expressions (20-22) provide a bound for the error made in the estimation of each parameter, which depends on the error in the data. Note that even though the errors in data are small, the estimation error can be significant large when the materials involved have similar thermal conductivities.

C. Local dependency of the estimated parameters on the data

Expressions (8) indicates that the estimated values depend on the parameters of the problem and data, as expected There are some useful tools to study the influence of data on the estimated parameters. Among them, sensitivity([21], [22]) and elasticity ([20], [22], [23]) analyses are frequently used in the literature. The latter is widely used in economics and provides the percentage error in the estimate for an error of 1% in the measured data.

Denoting by *P* the parameter to be estimated and *d* the data, the elasticity function is defined by:

$$E(d) = \frac{d}{P}\,\frac{\partial P}{\partial d}. \qquad (23)$$

In this case, given that two parameters are estimated from two different data, we are interested in studying four elasticity functions. The elasticity of parameter l_i on the noisy temperature data T_j is defined as,

$$E_{l_i}^{T_j}(T_1, T_2) = \frac{T_j}{l_i(T_1,T_2)}\,\frac{\partial l_i(T_1,T_2)}{\partial T_j}, \qquad (24)$$

where $i = 1, 2$ (two parameters are estimated), $j = 1, 2$ (two noisy data are used of temperature). The functions defined in (24) can be calculated from (8) to obtain:

$$E_{l_1}^{T_1}(T_1, T_2) = \frac{T_1}{T_1 - F + \frac{L\,h}{2\,\kappa_B}(T_2 - T_a)}, \qquad (25)$$

$$E_{l_1}^{T_2}(T_1, T_2) = \frac{T_2}{T_a - T_2 + \frac{L\,h}{2\,\kappa_B}\frac{(T_2 - T_a)^2}{F - T_1}}, \qquad (26)$$

$$E_{l_2}^{T_1}(T_1, T_2) = \frac{T_1}{T_1 - T_2 + h\,L(T_2 - T_a)\left(\frac{1}{2\,\kappa_B} - \frac{1}{\kappa_C}\right)}, \qquad (27)$$

$$E_{l_2}^{T_2}(T_1, T_2) = \frac{T_2}{\left(\frac{T_2 - T_a}{t_a - T_1}\right)\left[T_1 - T_2 + h\,L(T_2 - T_a)\left(\frac{1}{2\,\kappa_B} - \frac{1}{\kappa_C}\right)\right]}. \qquad (28)$$

IV. NUMERICAL EXAMPLE

In this section a numerical example is included in order to illustrate the performance of the estimation technique proposed in this work. The forward problem is solved assuming that l_1 and l_2 are given. Then, analytical values for T_1 and T_2 are obtained by using the expression in (7). Random noise, normally distributed with zero mean and deviation σ, is added to simulate experimental measurements T_1^{ϵ} and T_2^{ϵ} ; from which the estimated values $\widehat{l_1}$ and $\widehat{l_2}$ are obtained by using the simulated values for the temperatures in (8).

For the examples presented here it is considered *L* = 10 m; *F*=100°C; T_a = 25°C. The convective heat transfer coefficients *(h)* are determined as explained in [18]. Under these considerations, the inverse estimation problem consists in estimate l_1 and l_2 for a Lead-Nickel-Silver bar, where data is simulated for $l_1 = 4\,m$ and $l_2 = 9\,m$.

TABLE I. RELATIVE ESTIMATION ERRORS OF l_1 AND l_2

T_1(°C)	T_2(°C)	$\frac{\lvert l_1 - \widehat{l_1}\rvert}{l_1}$	$\frac{\lvert l_2 - \widehat{l_2}\rvert}{l_2}$
64.9	52.0	0.065	0.012
65.0	52.1	0.053	0.009
65.1	52.2	0.041	0.007
65.2	52.3	0.029	0.005
65.3	52.4	0.017	0.003
65.4	52.5	0.006	0.000
65.5	52.6	0.005	0.001
65.6	52.7	0.017	0.003
65.7	52.8	0.028	0.005
65.8	52.9	0.040	0.007
65.9	53.0	0.051	0.009

Table 1 shows the relative estimation errors for data close to the true analytical temperature values, which in this case are T_1 = 65.45 ºC and T_2 = 52.55 ºC.

It can be seen that the good recoveries are obtained, depending on the temperature data values used. As point out before, the estimation worsens when the temperature data are far

from the true. In this range of temperatures, a maximum error of 6% is obtained for the estimation of l_1 and 1% for that of l_2.

The elasticities of l_1 and l_2 with respect to the temperature data T_1^ϵ and T_2^ϵ are shown in Figs. (4)-(7), for the three-layer material Lead-Nickel-Silver described before.

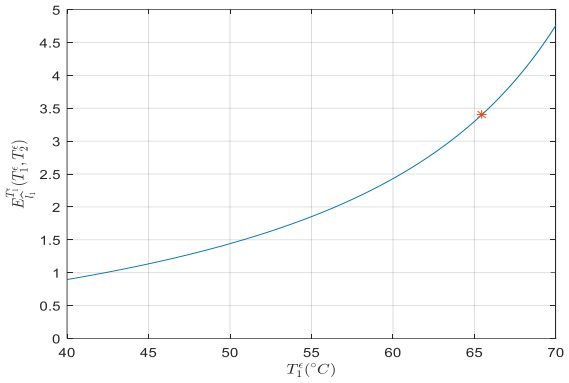


Fig.4. Elasticity of l_1 with respect to T_1

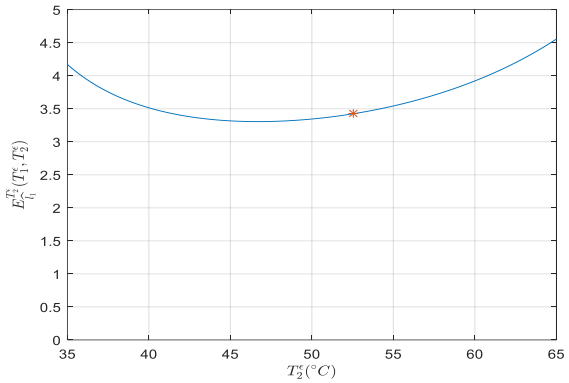


Fig. 5. Elasticity of l_1 with respect to T_2

Figs. (4-7) indicate that a measurement error of 1 % in the temperature T_1 translates into an error of 3.3 % in the estimate of (l_1) and an error of 3.0 % in the estimation of (l_2). Similarly, a 1 % measurement error in the temperature T_2 translates into a 3.4% error in the estimate of (l_1) and a 3.6 % error in the estimate of (l_2). In this case, it can be observed that l_1 and (l_2) have the same order of sensitivity with respect to errors in the data.

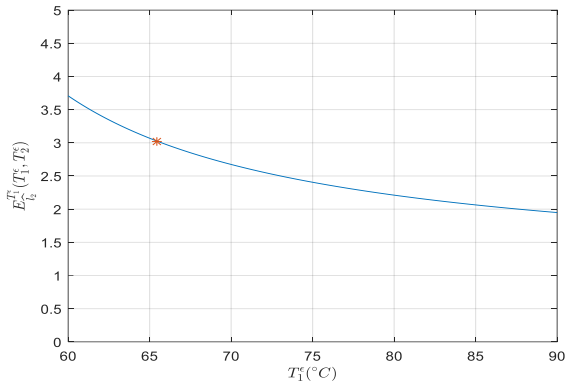


Fig. 6. Elasticity of l_2 with respect to T_1

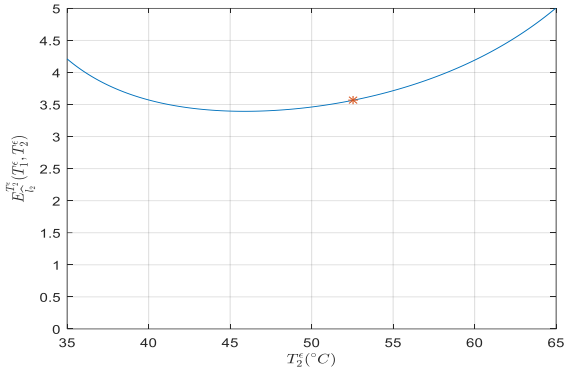


Fig. 7. Elasticity of l_2 with respect to T_2

V. CONCLUSION

In this work, it is dealt with the simultaneous estimate of the location of the solid-solid interfaces a three-layer materials considering a stationary heat transfer problem. A technique is proposed for the estimation based on two noisy temperature over-conditions, one at the middle and the other on the far right of the body. Necessary and sufficient conditions are derived for the existence and uniqueness of the solution to this inverse problem. Moreover, an analytical bound is obtained for the error in the estimations. Using the elasticity function, the local influence of the data on the estimated parameters is studied. The numerical example suggests that the approach introduced here behaves well although the determination is very sensitive to measurement errors.

ACKNOWLEDGMENT

The first and second authors acknowledge support from SOARD/AFOSR through grant FA9550-18-1-0523. The third author acknowledges support from European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant Agreement No. 823731 CONMECH and by the Project PIP No. 0275 from CONICET-UA, Rosario, Argentina.

REFERENCES

[1] A. M. Clausing, and B. T. Chao, "Thermal contact resistance in a vacuum environment," Journal of Heat Transfer 87(2) (1965), pp. 243-250.

http://dx.doi.org/10.1115/1.3689082

[2] K. Kim, S. Mun, M. Jang, J. Sok, and K. Park, "Thermoelectric properties of Ni/Ge-multilayer-laminated silicon," Applied Physics A 127(1) (2021), pp. 1-7.
https://doi.org/10.1007/s00339-020-04200-2

[3] O. Krotov, P. Gromyko, M. Gravit, S. Belyaeva, and S. Sultanov," Thermal conductivity of geopolymer concrete with different types of aggregate," IOP Conference Series: Materials Science and Engineering 1030(1) (2021), pp. 12-18.
https://doi.org/10.1088/1757-899X/1030/1/012018

[4] D. Rubio, D. A. Tarzia, and G.F. Umbricht, "Heat Transfer Process with Solid-solid Interface: Analytical and Numerical Solutions," WSEAS Transactions on Mathematics 20 (2021), pp. 404–414.
http://dx.doi.org/10.37394/23206.2021.20.42

[5] L. Zhou, M. Parhizi, A. Jain, "Analytical solution for temperature distribution in a multilayer body with spatially varying convective heat transfer boundary conditions on both ends," Journal of Heat Transfer 143(3) (2021), 034501.
https://doi.org/10.1115/1.4048968

[6] L. Zhou, M. Parhizi, A. Jain, "Temperature distribution in a multi-layer cylinder with circumferentially-varying convective heat transfer boundary conditions," International Journal of Thermal Sciences 160(3) (2021), 106673.
https://doi.org/10.1016/j.ijthermalsci.2020.106673

[7] H. Y. Ma, P. X. Zhu, S. G. Zhou, J. Xu, W. F. Huang, H. M. Yang, and M. J. Chen, "Preliminary research on Pb-Sn-Al laminated composite electrode materials applied to zinc electrodeposition," Advanced Materials Research 150–151 (2010), pp. 303-308.
http://dx.doi.org/10.4028/www.scientific.net/AMR.150-151.303

[8] D. G. Cahill, W. K. Ford, K. E. Goodson, G. D. Mahan, A. Majumdar, H. J. Maris, R. Merlin, and S. R. Phiipot, "Nanoscale thermal transport," Journal of Applied Physics 93(2) (2003), pp. 793–818.
http://dx.doi.org/10.1063/1.1524305

[9] A. M. Clausing, and B. T. Chao, "Thermal contact resistance in a vacuum environment," Journal of Heat Transfer 87(2) (1965), pp. 243–250.
http://dx.doi.org/10.1115/1.3689082

[10] V. Barturkin, "Micro-satellites thermal control-concepts and components," Acta Astronautica 56(1–2) (2005), pp. 161–170.
http://dx.doi.org/10.1016/j.actaastro.2004.09.003

[11] R. K. Shervedani, A. Z. Isfahani, R. Khodavisy, and A. Hatefi-Mehrjardi, " Electrochemical investigation of the anodic corrosion of Pb-Ca-Sn-Li grid alloy in H2SO4 solution," Journal of Power Sources 164(2) (2007), pp. 890–895.
https://doi.org/10.1016/j.jpowsour.2006.10.105

[12] A. Aziz, S. Jan, F. Waqar, B. Mohammad, M. Hakim, and W. Yawar, "Selective ion exchange separation of uranium from concomitant impurities in uranium materials and subsequent determination of the impurities by ICPOES," Journal of Radioanalytical and Nuclear Chemistry 284(1) (2010), pp. 117-121.
http://dx.doi.org/10.1007/s10967-009-0444-5

[13] C. L. Luke, and M. E. Campbell, "Determination of impurities in germanium and silicon," Analytical Chemistry 25(11) (1953), pp. 1588–1593.
https://doi.org/10.1021/ac60083a004

[14] T. P. Rao, D. Sobhi, S. Daniel, and J. M. Gladis, "Tailored materials for preconcentrationor separation of metals by ion-imprinted polymers for solid-phase extraction (IIP-SPE)," TrAC Trends in Analytical Chemistry 23(1) (2004), pp. 28–35.
https://doi.org/10.1016/S0165-9936(04)00106-2

[15] S. Zhai, P. Zhang, Y. Xian, J. Zeng, and B. Shi, "Effective thermal conductivity of polymer composites: theoretical models and simulation models," International Journal of Heat and Mass Transfer 117 (2018), pp. 358–374.
https://doi.org/10.1016/j.ijheatmasstransfer.2017.09.067

[16] S. Görög, "Identification and determination of impurities in drugs," Elsevier, Amsterdam (2000).
https://doi.org/10.1016/s1464-3456(00)x8001-5

[17] V. Andrisano, V. Cavrini, P. Summer, and S. Passuti, "Determination of Impurities in oxidation hair dyes as raw materials by liquid chromatography (HPLC)," International Journal of Cosmetic Science 17(2) (1995), pp. 53–60.
https://doi.org/10.1111/j.1467-2494.1995.tb00109.x

[18] G. F. Umbricht, D. Rubio, R. Echarri, and C. El Hasi, "A technique to estimate the transient coefficient of heat transfer by convection," Latin American Applied Research 50(3) (2020), pp. 229-234.
https://doi.org/10.52292/j.laar.2020.179

[19] Y.A. Cengel, "Heat and mass transfer: a practical approach," McGraw-Hill, New York (2007).
https://www.academia.edu/38856854/Heat_and_Mass_Transfer_A_
Practical_Approach_3rd_Edition_by_Cengel20190418_8592_13b2vml

[20] G. F. Umbricht, D. Rubio, and D.A. Tarzia, " Estimation technique for a contact point between two materials in a stationary heat transfer problem. Mathematical Modelling of Engineering Problems 7(4) (2020), pp. 607–613.
http://dx.doi.org/10.18280/mmep.070413

[21] H.T. Banks, S. Dediu, and S. L. Ernstberger, "Sensitivity functions and their uses in inverse problems," Journal Inverse and Ill-posed Problems, 15 (2007), pp. 683-708.
https://dx.doi.org/10.1515/jiip.2007.038

[22] K. Sydsaeter,and P. J. Hammond, "Mathematics for Economic Analysis," Prentice Hall, New Jersey (1995)
ISBN 013583600X, 9780135836002

[23] G. F. Umbricht, D. Rubio, D. A. Tarzia, "Estimation of a thermal conductivity in a stationary heat transfer problem with a solid-solid interface," International Journal of Heat and Technology, 39(2) (2021), pp. 337-344.
https://doi.org/10.18280/ijht.390202

In vitro Assessment of Mechanical Heart Valve Performance in Concomitant Presence of Discrete Subaortic Stenosis Using Particle Image Velocimetry System

Othman Smadi
Biomedical Engineering Dept.
Faculty of Engineering
The Hashemite University
Zarqa, Jordan
Othman.smadi@hu.edu.jo

Baha Al-Deen El-khader
Mechanical Engineering Dept.
Faculty of Engineering
Pennsylvania State University
University Park, PA, USA
bahaaelkhader98@gmail.com

***Abstract*— Due to presence of congenital or acquired sever subaortic stenosis, concomitant left ventricular septal myectomy during aortic valve replacement is possible. However, in some patients, subaortic stenosis may recur and request regular clinical evaluation (recurrence rate might reach up to 37%) . in this study, using particle image velocimetry and cutting-edge cardiac simulator, hemodynamic impact of symmetric discrete subaortic stenosis on flow downstream of bileaflet mechanical heart valve was investigated. For this purpose, thirteen 3D printed subaortic stenosis models were fabricated and tested under physiological flow conditions. The study revealed the significant impact of SAS on the flow patterns in terms of velocity, vorticity dynamics, Reynolds shear stress magnitude, and Doppler echocardiographic parameters.**

Keywords*— *Subaortic Stenosis, Subvalvular Stenosis, Prosthetic Heart Valve, Particle image velocimetry, Vorticity, Reynolds Shear Stress, Platelet Activation, Effective Orifice Area, Transvalvular Pressure Gradient. Introduction

I. INTRODUCTION

Subaortic stenosis (SAS) is an obstructive lesion in the left ventricle outflow tract (LVOT), just below the aortic valve, which causes a drop in the blood flow across the left ventricular outflow tract. Although it's usually a congenital disorder seen at birth, it could still arise due to acquired heart valve disease [1,2]. Subvalvular stenosis is the second most prevalent type of aortic stenosis accounting for 8–30% of total pediatric aortic stenosis and in 60% of cases it's associated with congenital defects. However, it's rarely diagnosed before Adolescence [3,4]. Different medical imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and echocardiography, are employed in the diagnosis process. Because of its availability, non-invasiveness, radiation-free nature, and cost-effectiveness, echocardiography is the most commonly used preliminary assessment method [21,22]. In some cases, utilizing both MRI and echocardiography are recommended for more accurate assessment [23]. However, their ability to adequately characterize leaflet dynamics and small-scale flow characteristics affected by their low spatial resolution. In addition, 2-D images from trans-thoracic echocardiography (TTE) might be less accurate when assessing the severity and geometry of stenosis, especially in infants; pregnant women; and cases of SAS with concomitant aortic stenosis [5].

In-Vitro studies are a common approach used to assess the performance of prosthetic heart valves and understand the etiology of various aortic stenosis diseases. In such studies, Blood flow behavior is investigated in terms of turbulence, shear stress, and coherent structure. Additionally, effective orifice area (EOA) and Transvalvular pressure gradient (PG), two clinically relevant echo Doppler measures, are also investigated [6]. Particle Image Velocimetry technique is commonly used with in vitro models for more in depth understanding of cardiovascular flow. Various parameters including velocity jets, coherent structures, vortex formation, leaflet kinematics, and shear stress magnitude are investigated revealing important thrombogenic and hemodynamic behaviors around and downstream of prosthetic valves in both physiological and pathological conditions [7].

In the current study, the subvalvular stenosis will be represented as a symmetrical discrete SAS, and flow patterns downstream of bileaflet mechanical heart valve will be studied under physiological flow conditions and by introducing different levels of subvalvular stenosis severity.

I. METHODS AND MODELS

A. Experimental setup

In this study, the ViVitro Labs Inc. cardiac simulator was used to create physiological flow conditions The cardiac output was set as normal flow condition at rest (4.9 L/Min)., with a heart rate of 70 beats per minute and systolic/diastolic pressure of 120/80 mmHg. Also, On-X aortic heart valve with standard sewing ring was used in the current study (Fig. 1 A&B).

Three different models of discrete SAS were 3D printed representing three levels of area reduction; 25%, 50%, and 75%. The material used to mold the SAS models was Thermoplastic polyurethane (better known as TPU) (Fig. 1(c)).

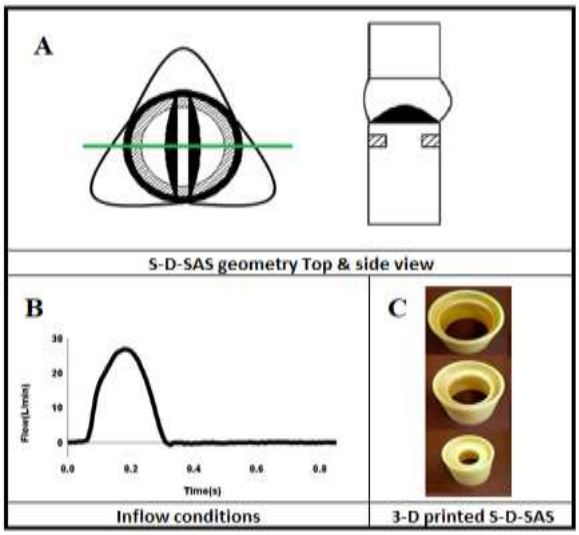


Fig. 1. PIV experimental setup. Location and shape of laser sheet and SAS (A), flowrate waveform (B), 3D printed SAS (C).

B. Particle image velocimetry

In this study, a planar 2D PIV analysis was carried out in a flow seeded with fluorescent PMMA-Rhodamine B particles with diameters ranging from 20 to 50 μm. The flow region was illuminated using a laser sheet generated by a double-pulsed Nd:YAG laser with a maximum output energy of 1200mJ (LaVISON GmbH, Göttingen, Germany). Images were captured using a PIV camera (Imager LX 2M GigE) with a 44 fps at full resolution of 1608 X 1208. Images acquired for the PIV analysis were in double frame mode, in which each frame contains one pair of images with a controllable time period between them. 250 ensembles with a spatial resolution of 0.039 mm/pixel. Davis 10.2.0 software was used to post process the acquired images and calculate the PIV measurements.

C. Measured hemodynamic parameters

Vorticity magnitude:

$$\omega_z = -\left(\frac{du}{dy} - \frac{dv}{dx}\right) \quad (1)$$

Reynolds shear stress (RSS):

$$RSS = \rho\sqrt{\left(\frac{\acute{u}\acute{u}-\acute{v}\acute{v}}{2}\right)^2 + (\acute{u}\acute{v})^2} \quad (2)$$

Peak Velocity (Vpeak):

It is the maximum measured velocity downstream of the valve.

Peak Pressure Gradient (PG):

$$PG = 4V_{peak}^2 \quad (3)$$

Peak Effective Orifice Area (EOApeak):

$$EOApeak = \frac{measured\ peak\ volume\ flowrate}{V_{peak}} \quad (4)$$

II. RESULTS AND DISCUSSION

Peak systole holds the peak values of different hemodynamic parameters (e.g. velocity, vorticity, Reynolds shear stress…) and onset of turbulence happens mostly at the peak systole. Therefore, only data at instant of peak systole will be presented in the current study.

Velocity magnitude at peak systole for four different levels of SAS severity is shown in Fig. 2. The normal case had three velocity jets (one central and two lateral jets). Also, introducing 25% severity of SAS did not, significantly, alter the flow patterns and velocity magnitude. Additionally, cases with 50% severity of SAS had similar flow patterns with a considerable increase in velocity jet's magnitude. However, at 75% severity of SAS, flow patterns and velocity magnitude had changed with the disappearance of the central velocity jet and profound presence of the two lateral velocity jets. The peak velocity magnitudes were proportional to the severity of stenosis and ranged from 1.67 to 4.33 m/s. In case of 25% severity of SAS, there was no significant increase in peak velocity magnitude compared to the normal case (1.69 vs. 1.67 m/s). Additionally, in case of 50% severity of SAS, there was a noticeable increase in velocity magnitude compared to normal case (1.67 vs. 2.04 m/s). However, in case of 75% severity of SAS, there was a dramatic increase in velocity magnitude compared to normal case (1.67 vs. 4.33).

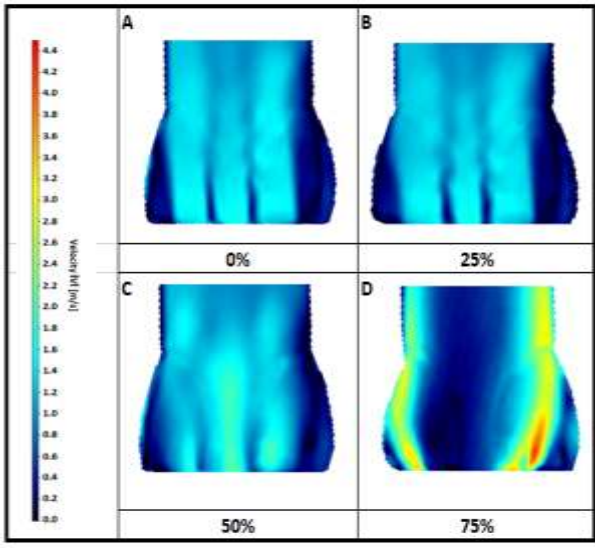


Fig. 2. Velocity magnitude at peak systole and for four different severities of SAS.

Vorticity magnitude at peak systole for four different levels of SAS severity is shown in Fig. 3. In all cases, vorticity magnitudes were proportional to the severity of SAS, and the highest magnitude of vorticity was found in 75% severity of SAS (peak vorticity magnitude = 3600 s^{-1}). At the case of 25% severity of SAS, vortex patterns showed an overall agreement compared to the normal case (maximum values were 1200 and 1600s^{-1}). Moreover, vortex patterns for the case of 50% severity of SAS was slightly different than normal case with localized high vorticity magnitude around the leading edges of the two leaflets. moreover, a reasonable increase in vortex magnitude compared to the normal case was observed (maximum value was 1800 vs. 1200 s^{-1}). However, a significant increase in vorticity magnitude as well as major changes in vortex patterns were noticed at 75% severity of SAS compared normal case. As a result, central shear layers were shifted towards the wall with significantly higher vorticity magnitudes (maximum value was 3600 s^{-1} vs. 1200 s^{-1}).

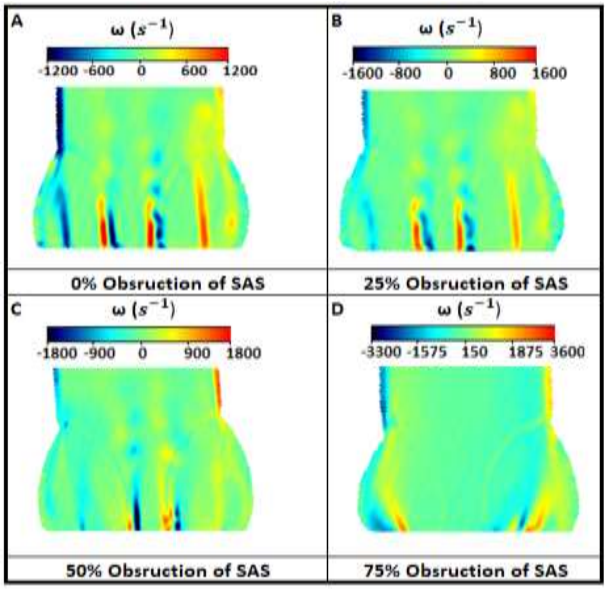


Fig. 3. Vorticity magnitude at peak systole and for four different severities of SAS.

Fig. 4 shows RSS contours at peak systole for four different levels of SAS severity. RSS magnitudes were proportional to the severity of SAS and a significant increase in RSS magnitude occurred at 75% severity of SAS. For the cases of 25% and 50% severities of SAS, no significant increase in RSS magnitude was noticed. However, in case of 75% severity of SAS, the high RSS magnitude regions around the leaflets became more lateral with a significant increase in magnitude as well (maximum value reached up to 400 Pa).

Fig. 5 displays the data for Vpeak, PG, and valve EOA as a function of the severity of SAS for the Doppler-Echocardiographic parameters obtained using the PIV system. According to the standards of the American Society of Echocardiography (ASE), the figure's dashed lines show the cutoff values for Doppler-Echocardiographic Parameters [8].

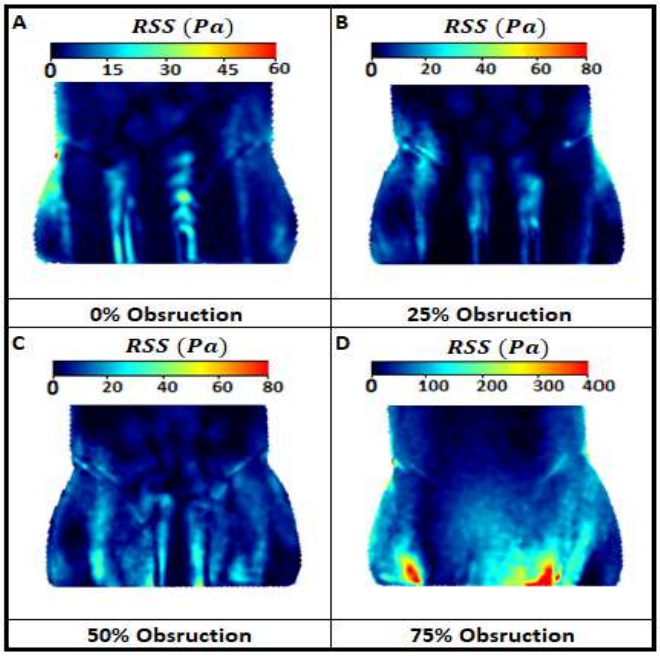


Fig. 4. Reynolds shear stress magnitude at peak systole and for four different severities of SAS.

Fig. 5(a) shows the relationship between Vpeak and the severity of SAS. For all cases, Vpeak was proportional to the severity of SAS. Additionally, Vpeak did not exceed the peak velocity magnitude of 3 m/s (possible valve stenosis) in cases of 25% and 50% severity of SAS. However, in case of 75% severity of SAS it is noticed that Vpeak exceeded the peak velocity magnitude of 4 m/s (significant valve stenosis) and reached up to 4.33 m/s. Fig. 5(b) shows the relationship between effective orifice area (EOA) and the severity of SAS. EOA was inversely proportional to the severity of SAS. Additionally, EOA did not fall below the EOA magnitude of 1.2 cm^2 in cases of 25% and 50% severity of SAS. However, in case of 75% severity of SAS it is noticed that EOA falls below the EOA magnitude of 1.2 cm^2 (possible valve stenosis) and reached down to the value of 1.10 cm^2. Fig. 5(c) shows the relationship between peak gradient (PG) and the severity of SAS. For all cases, PG was proportional to the severity of SAS. Additionally, PG did not exceed the PG magnitude of 36 mmHg (possible valve stenosis) in cases of 25% and 50% severity of SAS. However, in case of 75% severity of SAS it is noticed that PG exceeds the PG magnitude of 64 mmHg (significant valve stenosis) PG magnitude of 75 mmHg.

III. DISCUSSION AND CONLUSION

In all studied cases, velocity magnitude was proportional to the severity of SAS which is consistent with the principle of mass conservation and in good agreement with smadi and coauthors [9]. It is worth noting that by reaching 75% severity, the opening of valve leaflets is significantly reduced and led to

more lateral shear layers (towards the wall) instead of conventional central ones.

The magnitude and pattern of vortices are strongly related to the level of platelet activation, and earlier research have shown that this association is correlated with vorticity magnitude [10]. Moreover, and for all studied cases, regions with relatively high RSS magnitudes are coincided with regions that hold high vorticity magnitude which increases the likelihood of thrombosis and/or hemolysis.

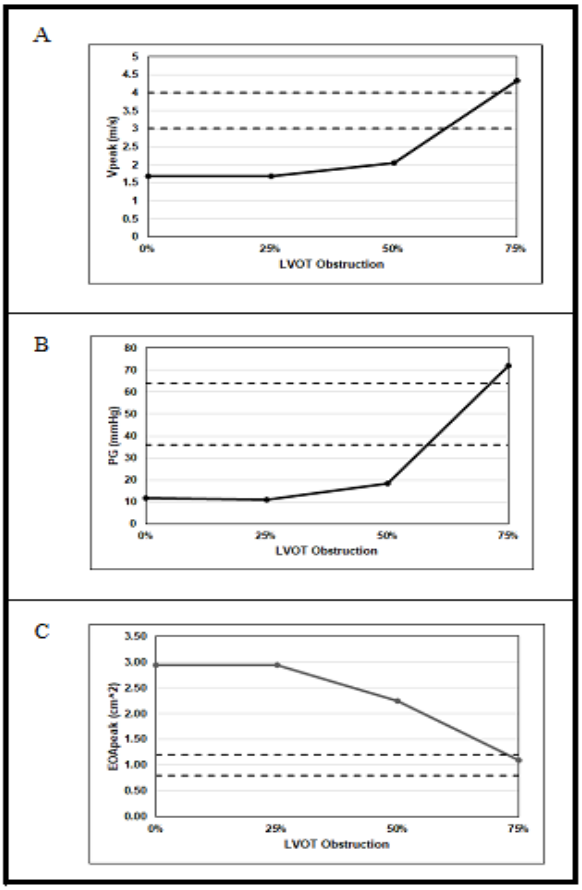


Fig. 5. Echocardiographic parameters measured using the PIV system. (A) Vpeak vs. the severity of SAS. (B) effective orifice area (EOA) vs. the severity of SAS. (C) peak gradient (PG) vs. the severity of SAS.

In general, the current study revealed that all cases with 75% severity of SAS compared to cases with inferior severity, showed dramatic increases and changes in all studied parameters (i.e. velocity magnitude and profile, vorticity dynamics, Reynolds shear stress, and Doppler-Echocardiographic parameters).

In conclusion, introducing discrete SAS resulted in greater escalation in velocity, vorticity, and RSS magnitudes. Which in turn might lead to higher risk of platelet activation and/or hemolysis. Moreover, 50% level of SAS severity might be considered as a cutoff value for optimal time of intervention, as 75% level of SAS severity showed a dramatic change in all studied parameters.

ACKNOWLEDGMENT

This work was fully supported by grant from the Deanship of Scientific Research at The Hashemite University.

REFERENCES

[1] D. S. Ezon, "Fixed subaortic stenosis: A clinical dilemma for clinicians and patients," *Congenit. Heart Dis.*, vol. 8, no. 5, pp. 450–456, Sep. 2013, doi: 10.1111/chd.12127.

[2] J. Aboulhosn and J. S. Child, "Echocardiographic evaluation of congenital left ventricular outflow obstruction," *Echocardiography*, vol. 32, no. S2, pp. 32–39, Jan. 2015, doi: 10.1111/echo.12181.

[3] J. A. Shar, K. N. Brown, S. G. Keswani, J. Grande-Allen, and P. Sucosky, "Impact of Aortoseptal Angle Abnormalities and Discrete Subaortic Stenosis on Left-Ventricular Outflow Tract Hemodynamics: Preliminary Computational Assessment," *Front. Bioeng. Biotechnol.*, vol. 8, Feb. 2020, doi: 10.3389/fbioe.2020.00114.

[4] K. S. Shapero and J. C. Chou, "Subaortic Stenosis With Elevated Aortic Gradients in a Pregnant Patient," *JACC Case Reports*, vol. 2, no. 1, pp. 131–134, Jan. 2020, doi: 10.1016/j.jaccas.2019.11.054.

[5] J. A. Shar, S. G. Keswani, K. J. Grande-Allen, and P. Sucosky, "Computational Assessment of Valvular Dysfunction in Discrete Subaortic Stenosis: A Parametric Study," *Cardiovasc. Eng. Technol.*, vol. 12, no. 6, pp. 559–575, Dec. 2021, doi: 10.1007/s13239-020-00513-8.

[6] F. M. Susin, S. Espa, R. Toninato, S. Fortini, and G. Querzoli, "Integrated strategy for in vitro characterization of a bileaflet mechanical aortic valve," *Biomed. Eng. Online*, vol. 16, no. 1, pp. 1–14, 2017, doi: 10.1186/s12938-017-0314-2.

[7] Y. Shi, T. J. H. Yeo, Y. Zhao, and N. H. C. Hwang, "Particle image velocimetry study of pulsatile flow in bi-leaflet mechanical heart valves with image compensation method," *J. Biol. Phys.*, vol. 32, no. 6, pp. 531–551, Dec. 2006, doi: 10.1007/s10867-007-9035-2.

[8] W. A. Zoghbi *et al.*, "Recommendations for Evaluation of Prosthetic Valves With Echocardiography and Doppler Ultrasound. A Report From the American Society of Echocardiography's Guidelines and Standards Committee and the Task Force on Prosthetic Valves, Developed in Conjunction," *J. Am. Soc. Echocardiogr.*, vol. 22, no. 9, pp. 975–1014, 2009, doi: 10.1016/j.echo.2009.07.013.

[9] O. Smadi, A. Abdelkarim, S. Awad, and T. D. Almomani, "Hemodynamic performance of dysfunctional prosthetic heart valve with the concomitant presence of subaortic stenosis: In silico study," *Bioengineering*, vol. 7, no. 3, pp. 1–15, 2020, doi: 10.3390/bioengineering7030090.

[10] A. Kheradvar, C. Rickers, D. Morisawa, M. Kim, G. R. Hong, and G. Pedrizzetti, "Diagnostic and prognostic significance of cardiovascular vortex formation," *Journal of Cardiology*, vol. 74, no. 5. Japanese College of Cardiology (Nippon-Sinzobyo-Gakkai), pp. 403–411, Nov. 01, 2019. doi: 10.1016/j.jjcc.2019.05.005.

A Method for the Contents Curation using Receiver Operating Characteristic (ROC) Curve

Hyun Jung Lee
Urban Policy Research
Goyang Research Institute
Goyang, Korea
hjlee5249@gmail.com

Euisin Kim
School of Business
Kyung Hee Cyber University
Seoul, Korea
eskim8@khcu.ac.kr

Mye Sohn
Dept. of Industrial Engineering
Sungkyunkwan University
Suwon, Korea
myesohn@skku.edu

***Abstract*—The Receiver Operating Characteristic (ROC) curve and Area Under the ROC curve (AUC) are used to determine the cutoff index for extracting a balanced information using the relationship between sensitivity and specificity in a confusion matrix. This method can reduce side effects like confirmation bias that can strengthen based on the user's preference in personalized recommendation systems. For the balanced curation of recommendation contents, it is important to identify the best cutoff index between two classes in the probability density function. This can be accomplished by adopting ROC curve to display the trade-off between personality and serendipity of the curated content at each cutoff. The decision making for the best cutoff index depends on the purpose of the curation. To do this, the hypothetical data are illustrated to plot the ROC curve on a discrete scale with five categories, similar to a binary with personality and serendipity for content curation in a personalized recommendation system. Through the hypothetical illustration, the best cutoff indices for the trade-offs between two classes are illustrated using the hypothetical data. For the balanced curation, the ROC is applied to determine whether the content belongs to personality or serendipity. It can be also applied in artificial intelligence-aided balanced learning. Further investigations are necessary to prove that the proposed method can provide appropriately curated contents according to each purpose by using the trade-off between the different features included in different classes. For the extraction of the balanced curated contents to reduce the confirmation bias, it is also necessary to determine the characteristics of the features for semantic relationships.**

Keywords—Contents, Curation, Confusion Matrix, ROC curve, AUC, Confirmation Bias, Personality, Serendipity

I. INTRODUCTION

Customized recommendations are actively used in various areas, such as recommendation systems, digital marketing, personalized advertising, personalization service of e-Learning, automatic recommendation of music or movie in YouTube, pushed news feed, personally curated video services, and so on. Especially, the robot-based automatic personalized recommendation systems have been actively used in online commerce, digital commerce, YouTube, Over the Top (OTT), publications, etc. Presently, the contents can be curated using methods that can consider the users' preferences. To date, various researchers have developed numerous algorithms, applications, and systems based on data analysis, machine learning, semantic web processing, data curation, statistics, and so on.

Although providing personalized and/or customized recommendations using contents, information, data, news, media, etc. is important to improve the performance of commerce, marketing, and so on efficiently and effectively, it is necessary to avoid foreseeable risks like confirmation bias, fake news, and so on. To date, some efforts have been made to prevent such risks, such as by providing a balanced information to the users to solve the illustrated problems. Among such methods, the one involving receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) is considered for the extraction of balanced information to provide a balanced view to the users. The ROC curve and AUC are applied to locate the best cutoff index to provide balanced contents and to reduce confirmation bias, which usually occurs in personalized and customized recommendation systems, especially in automatic recommendation systems based on artificial intelligence (AI) [5]. Further, the confusion matrix is used to determine the appropriate cutoff index, which can be a criterion for the golden ratio between the features of personality and serendipity, to extract the balanced contents.

In this study, we first overviewed the reported studies on confirmation bias, ROC curve, AUC, and information curation. Second, we introduced a probability density function and decision matrix as an overall architecture. Here, we discuss the area analysis of an ROC curve with a hypothetical illustration to identify the best cutoff index to curate balanced contents for recommendations. We also highlight the issues that require further detailed investigations.

II. LITERATURE REVIEW

A. Confirmation Bias

Automatic recommendation systems based on users' personalities and/or their needs have been developed to provide tailored information for different purposes. They are useful to improve user satisfaction and recommendation performance as well as to increase sales via online commerce, and other such applications, although some limitations do exit. First, such automatic recommendation systems provide a narrow view for the needed information, contents, media, products, etc., which can be usually generated and provided among the collected biased information on the Web. The popularized recommendation robots are mainly focused on providing tailored information, but the robots can also induce side effects like confirmation bias, which can be amplified according to different behaviors, such as user's information retrieval like sharing, searching, etc. [4]. The side effects can occur on online news, media, YouTube, marketing, OTT, etc. and can affect the users who are surrounded by biased information. Especially, digital confirmation bias has been intensified in the digital media era using digital contents [1, 4]. Filter bubble, which is a kind of side effect, has been intensified by personalization algorithms to skew or limit the information on online platforms that provide social network services (SNS). Above all, for a sound recommendation, providing balanced contents is crucial.

B. ROC Curve and AUC

The ROC curve and AUC are usually used to compare the difference between classes. The ROC as a probability curve and AUC as the probability rank a positive preference more highly than the negative preference. They are arranged from left to right in ascending order of logistic regression predictions as shown in Fig.1. In this graph, the y axis shows personality from 0 to 1, which are the thresholds in the curve, and the x axis shows the weight of the personality. If the threshold ≥ 0.5, then a positive probability is obtained, otherwise a negative probability is observed. Herein, the probability is indicated by the personality, and positive implies superiority of personality; other values indicate the absence of a strong personality.

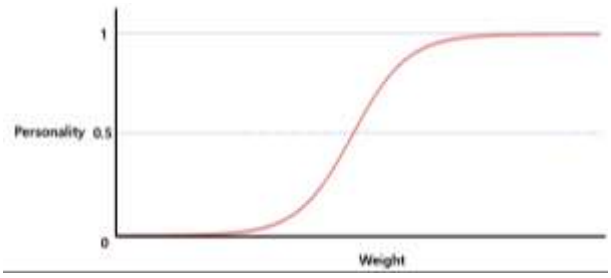


Fig. 1. Predictions from a logistic regression model.

The ROC curve and AUC methods are applied to provide balanced information and to mitigate confirmation bias, which can occur in the process of hyper-personalization. The application of an ROC curve and AUC to maintain a balance between the comparative classes is slightly different than measuring the diagnostic accuracy, such as sensitivity and specificity, in a confusion matrix. In general, they are used to separate two comparative classes, such as the discrimination of classes using accuracy. However, in this paper, the ROC curve and AUC are applied to find the best cutoff index to maintain the equilibrium between the positive and non-positive personalities according to users' preferences under the consideration of contexts.

C. Digital Content Curation

Curation is usually used for planning art exhibitions or galleries. In a digital environment, curation is adopted to maintain the balanced value of the information. Different types of curations are known, such as data curation, digital curation, social curation, content curation, etc. In addition, the curation method has been applied to automatically generate customized information for recommendation systems. To implement digital content curation, it is necessary to adopt methodologies such as content-based filtering using features, collaborative filtering based on the user's related information, demographic filtering based on the user's preferences, knowledge-based filtering based on the user's explicit inputs, and so on as well as statistics.

In general, personalized recommendation systems are focused on providing users' preferred information, including their preferences relevant contents, images, data, graphs, text, media and so on. Such systems strengthen the user's thinking by sing the gathered relevant, interesting, and meaningful information. It is true that curation-based systems effectively provide meaningful and valuable information to users. Therefore, it is widely used in a variety of field such as personalized marketing, news feed, advertisement, etc. which are based on the personality. In other words, it is difficult to have different view as serendipity from personality. In addition, these kinds of systems can strengthen the confirmation bias as a kind of side effect, which is generated by users' preferences according to the personality. Currently, robot advisors based on machine learning (ML) and artificial intelligence (AI) are popularly used to effectively provide customized information. Thus, automatic recommendation systems provide tailored information. However, such systems can also provide biased information to the users as well as reinforce biased thinking. For instance, the mass production of fake news can affect the users thinking capacity. The customized curation method has been improved according to users' needs and will be developed continuously according to the development of ML and AI. Therefore, it is important to consider how to provide balanced contents as well as reduce confirmation bias.

The lifecycle of recommendation can follow the principles of FAIR data, which are comprised of findable, accessible, interoperable, and reusable values [6,7]. Hence, the curation steps are defined as: find, select, editorialize, arrange, create, share, engage, track, seek, and evaluate [2]. Digital curation is conducted by human-driven and algorithmic techniques as well as a combination of both [2]. A digital curator gathers and selects relevant information and data, contents, such as videos, photos, audios, texts, etc., for one's own use. A digital curator is also called a data curator, content curator, information curator, etc. The curated contents can be used to curate the feeds containing news, multimedia resources, interesting posts, and so on. For digital curation, it is necessary to apply a mind-map, metadata, structure, semantics, syntactic, statistics, etc. to gather and extract relevant information, data, contents, and so on.

III. PROBABILITY DENSITY FUNCTION AND DECISION MATRIX

A. Binary Prediction

In recommendation systems, the user's satisfaction with the recommended service depends on the degree of reflection of personality. The individualized recommendation mainly focuses on the reflection of only the person's attention-based features. It is easy for the recommendation to fall into a narrow loop of thought, and it can also cause side effects, such as confirmation bias, fake news, filter bubbles, and so on. Therefore, it is important to focus on providing a balanced information to the users by reducing such side effects. In this study, we adopted the ROC curve and AUC methods for providing good information. In this case, the content was simply classified into binary features, viz. personality and serendipity, as follows.

$$Personality = \{pf_i | \, pf_i \; is \; i^{th} \; feature \; of \; personality, \; 0 \le i \le n, n \; is \; a \; number\} \quad (1)$$

$$Serendipity = \{sf_j | \, sf_j \; is \; j^{th} \; feature \; of \; serendipity, \; 0 \le j \le m, m \; is \; a \; number\} \quad (2)$$

Personality is specified by user's preference-based tailored information. Serendipity is specified by somewhat out of personal preference, although it may be interesting and fresh information to the user [3]. Therefore, personality is composed of the user's preference-based features, whereas serendipity is composed of features that are somewhat different from the personal preferences.

B. Decision Matrix

Binary variables, such as personality and serendipity, can be applied to digital curation for providing balanced information using the ROC curve and AUC methodology based on the probability density function shown in Fig. 2. In Fig. 2, the y axis indicates the frequency of the features like personality and serendipity, and the x axis represents the hypothetical predicted recommendation results for the binary characteristics; PP means True Personality, SS means True Serendipity, PS represents False Serendipity, and SP denotes False Personality. The shape of the graph for personality and serendipity is dependent on the cutoff index.

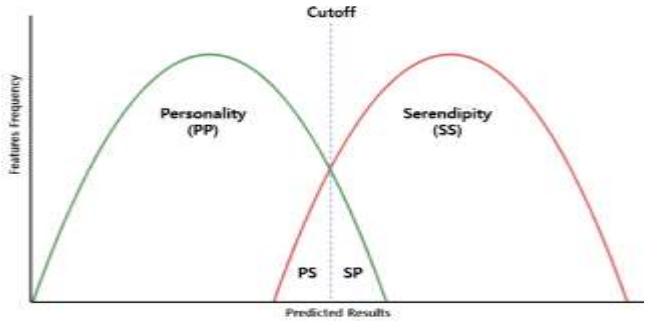


Fig. 2. Probability density function of hypothetical predicted recommendation results for two populations (personality and serendipity).

In this paper, it is assumed that the curated contents match the personality using a decision matrix. A confusion matrix is used to measure the performance of prediction through training in ML and AI, for comparing the predicted values with the actual ones. The matrix is often applied to clinical medicine for estimating and comparing the accuracy of competing diagnostic tests. Herein, the confusion matrix is applied to the decision matrix to show the curated recommendation results of a binary predicted class, according to each individual's actual preferences as follows:

TABLE I. DECISION MATRIX SHOWING CURATED RECOMMENDATION RESULTS OF A BINARY PREDICTED CLASS, ACCORDING TO EACH INDIVIDUAL'S ACTUAL PREFERENCES

Digital Curation for Recommendation of Predicted class	Characteristics Actual Class (User's Actual Preferences)		
	Personality (P+)	Serendipity (P-)	Total
Personality (R+)	PP(True Personality) True Positive(TP)	SP(False Personality) False Positive(FP)	PP+SP: Predicted Personality Sensitivity: PP/(PP+SP)
Serendipity (R-)	PS(False Serendipity) False Negative(FN)	SS(True Serendipity) True Negative(TN)	PS+SS: Predicted Serendipity Specificity: SS/(PS+SS)
Total	PP+PS: Personality	SP+SS: Serendipity	PP+PS+SP+SS: Sample size

In the confusion matrix shown in Table 1, True Personality (PP) as True Positive (PP) implies that the content is predicted as "Personality," and its actual class is "Personality," which is "True." False Personality (SP) as False Positive (FP) implies that the content is predicted as "Personality," and its actual class is "Serendipity," which is "False." False Serendipity (PS) as False Negative (FN) means that the content is predicted as "Serendipity," and its actual class is "Personality," which is "False." True Serendipity (SS) as True Negative (TN) indicates that the content is predicted as "Serendipity," and its actual class is "Serendipity," which is "True." Sensitivity as True Personality fraction is calculated as PP/(PP+PS), and specificity as True Serendipity fraction is calculated as SS/(SP+SS).

The predicted hypothetical recommendation results can be classified as binary features like personality and serendipity, and the corresponding results are summarized as in Table1. In the continuous predicted results, cutoff can be used to classify the results. As in Fig. 2, the personality is located below the cutoff threshold, and the serendipity is located above the cutoff threshold. In general, a probability density function is usually used to accurately classify sensitivity and specificity. Sensitivity correctly classifies the predicted result for the personality group as positive, and Specificity correctly classifies the test result for the serendipity group as negative. Therefore, the probability density function with binary characteristics, including personality and serendipity, is applied to curate digital contents for providing balanced information and for reducing confirmation bias as much as possible. It is necessary to prove

whether the providing balanced information can extract the positive results for the recommendation systems or not.

IV. AREA ANALYSIS OF ROC CURVE WITH ILLUSTRATION

A. ROC Curve with Illustration

A variety of decision making is possible depending on the cutoff threshold. For instance, depending on the cutoff, the classes predicted by the curation simulation can be classified into several classes, such as L1, L2, L3, L4, and L5. L1 is highly personality (more lax Serendipity), L2 is highly personality (lax Serendipity), L3 is moderate, L4 is highly serendipity (lax personality), and L5 is more highly serendipity (more lax personality) as shown in Table II.

TABLE II. RESULTS FROM HYPOTHETICAL DATA FOR AN ILLUSTRATIVE ROC PLOT

Digital Curation for Recommendation of Predicted class	Characteristics Actual Class		Decision rule (cutoff, C1-C4)	
	Serendipity	Personality	Sensitivity	1-Specificity
L1: more highly personality (more lax Serendipity)	24	158	C1=0.95	C1=0.73
L2: highly personality (lax Serendipity)	59	149	C2=0.82	C2=0.48
L3: moderate	103	95	C3=0.61	C3=0.31
L4: highly serendipity (lax personality)	137	112	C4=0.32	C4=0.12
L5: more highly serendipity (more lax personality)	149	71		
Total	472	585		

Table II shows hypothetical data for the ROC curve. Evidently, the decision results can be changed depending on the cutoff threshold value. For instance, cutoff 1 classifies the L1 category as personality and all the other categories are serendipity; cutoff 2 classifies L1 and L2 as personality and all the other categories are serendipity; cutoff 3 classifies L1, L2, and L3 as personality and all the other categories are serendipity; and cutoff 4 classifies L1, L2, L3, and L4 as personality, and L5 is the only category classified as serendipity. Thus, the personality features can be determined by the cutoff threshold value. That is, it is important to find the best cutoff threshold for recommending balanced contents to the users while minimizing the personalized information loss.

Based on the hypothetical example shown in Table II, the predicted recommendations are classified on a five-category discrete scale to create the ROC curve in the recommendation prediction. The meaningful interpretation of the AUC as the application of the ROC methodology in a recommendation system to determine the best cutoff indices are discussed in the next section. Therefore this kind of the cutoff index can be applied to determine the mixing rate of the features between personality and serendipity to mitigate the predictable confirmation bias.

B. Area Analysis of ROC Curve

The information curation accuracy depends on the AUC between 0 and 1, and a larger area implies a higher curation accuracy and the discriminatory ability of the accuracy. Contrarily, the aim of this study was to deduce the best cutoff reference index using the AUC to reduce the confirmation bias as the accuracy increases [5].

The cutoff index can be simply classified as strict, moderate, and lenient. Fig. 3 shows the strict cutoff; on the left side, two groups are completely classified using a probability density function, and on the right side, the area of the right graph (AUC) shows 1. A larger degree of correct classification of the two groups, results in a lower degree of overlapping between the distributions of the two groups. In this case, AUC is 1.0, and information curation is highly useful as illustrated in Fig. 3. AUC=1.0 shows that a recommendation result of 1 is a perfect test. It means that all the features in the recommendation are comprised of personality features, and the false positive rate is 0. In this case, it is assumed that the personalized recommendation is completed by perfectly reflecting the user's personality. Otherwise, the recommended contents do not contain any other feature of serendipity.

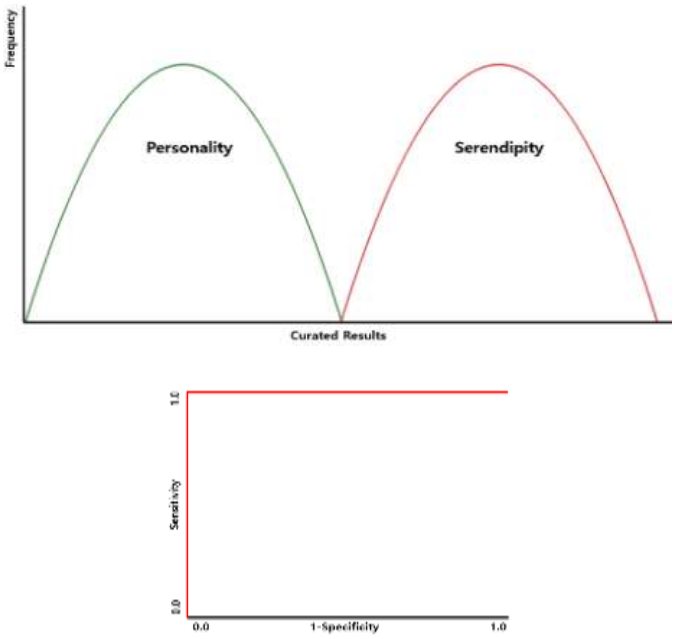


Fig. 3. Probability density function of hypothetical predicted recommendation for two populations (personality and serendipity) with a strict cutoff (AUC = 1).

Fig. 4 shows the moderate cutoff. In this case, the AUC > 0.5. As illustrated in Table II, the degree of overlapping between the two classifications can be different depending on the determined cutoff index. It can affect the curation of the recommendable information and/or contents as well as hinder the mitigation of confirmation bias, which can occur in preference-based automatic recommendation systems, like news feeds, filter bubbles, etc. Thus, it is important to determine the best cutoff index between the binary variables, depending on the application.

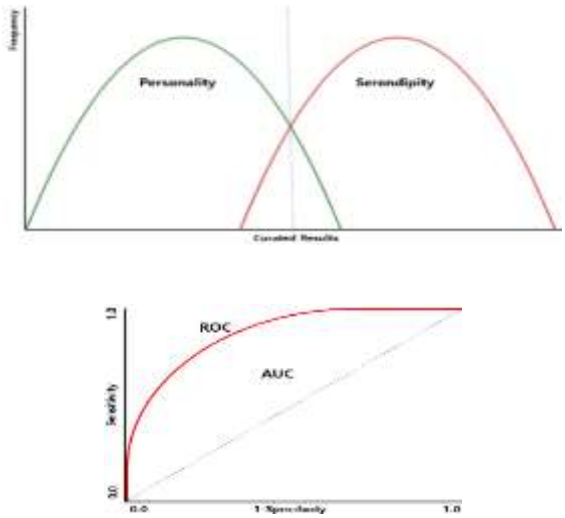


Fig. 4. Probability density function of hypothetical predicted recommendation for two populations (personality and serendipity) with a moderate cutoff (AUC > 0.5).

Fig. 5 shows a lenient cutoff. In this case, the AUC is 0.5, implying that the test results obtained from the two groups overlap and are non-informative. That is, the ROC curve shows a 45° straight line from the coordinates (0,0) to the coordinates (1,1), and PP(cutoff) = SP(cutoff) at all the personality test point cutoffs. Thus, it can be concluded that the curation/recommendation is non-informative as shown in Fig. 5. AUC = 0.5 is the case where the true personality rate and the false personality rate are the same. In this case, the diagnostic test result does not provide any useful information when judging information recommendation.

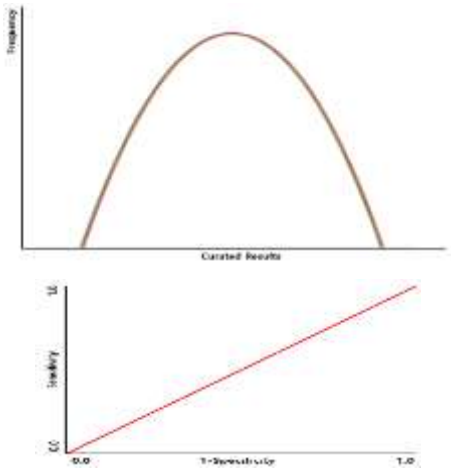


Fig. 5. Probability density function of the hypothetical predicted recommendation for two populations (personality and serendipity) with a lenient cutoff (AUC = 0.5).

In Fig. 6, AUC is 0, implying that the predicted recommendation has been completely reversed from personality to serendipity and vice versa. AUC = 0 is a situation, in which the serendipity-based content is completely misclassified as a personality-based content and vice versa for all the recommendation contents.

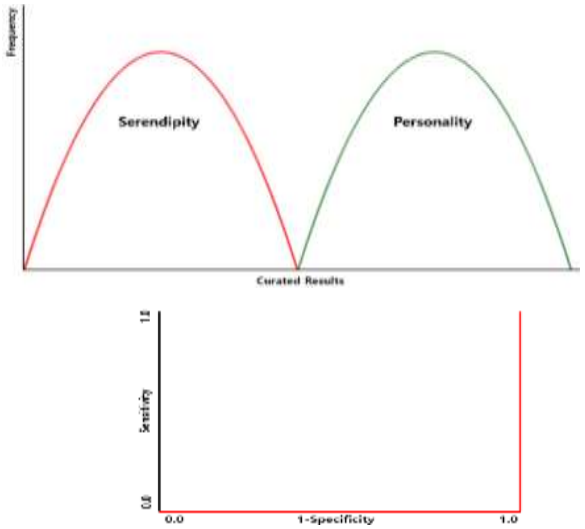


Fig. 6. Probability density function of hypothetical curated simulation results for two populations (personality and serendipity) with completely reversed (AUC = 0).

V. CONCLUSION

The relationship between personality and serendipity index derived using the ROC curve can be used to compare the characteristics of information curation according to the specific application and find the optimal reference point through various decision-making reference points [8]. For instance, in Table II, if the purpose of recommendation is to screen all the personality type information that is likely to match the personal preference, then high personality is required. Therefore, when the reference point is set to C1, the personality and serendipity are 95%, and 27%, respectively. In contrast, if the purpose of recommendation is to extract only the serendipity type information, then that is not included in the personal preference, if C4, which has a high serendipity reference point, is set, then the personality and serendipity are 32%, and 88%, respectively.

According to the popularity of the automatic personalized recommendation systems, they are ultimately centered on the improvement of customers' satisfaction, sales performance, marketing customization, personalized contents-based advertisement, customized news feed, and so on. It is more popular in automatic digital content recommendation systems based on ML and AI.

Unexpectedly, the personalized automatic recommendation systems exhibit some drawbacks, such as confirmation bias, fake news, feed bubbles, filter bubbles, and so on. Thus, in this study, we focused on providing balanced information and the mitigation of confirmation bias. The ROC curve and AUC are applied to determine the best cutoff index to extract a balanced information depending on the given recommendation purpose in the systems. As an illustration, we classified the data into five classes according to the cutoff index for the binary variables personality and serendipity depending on the probability density obtained using the hypothetical data. It is used to determine the

best cut off index to reduce the side effects from the personalized recommendation.

The proposed method can be applied to some areas related to automatic recommendation systems using ML and AI as follows (the details will be discussed in the further study):

1) Determine whether the Features of the Content are Personality or Serendipity

In general, a single content, such as an image, text, graph, etc., consists of multiple features. Here, it is a combination of binary features, such as personality and serendipity. To perform a satisfactory curation, it is necessary to clearly distinguish the features of the content. However, the features of the contents may dynamically be varied depending on their usage patterns or contexts. To perform a personalized feature classification, we will devise a learning mechanism based on a knowledge graph and the graph convolutional network.

2) Balanced Curation of Recommendation Contents

As illustrated before, for the balanced curation of information, contents, media, texts, news, etc., the proposed method using the ROC curve and AUC can be applied to determine the mixing rate of the contents. To determine the advanced mixing rate based on the context and semantics, it is necessary to consider the users' preference for the contents; for example, a rating, dwell time, or access route. In a future study, we intend to accurately identify the users' preferences by using the knowledge graph, which is widely used in recent recommendation systems.

To determine the characteristics of a single content, the ROC curve and AUC can be applied to classify the features of the content as binary. If the included features can be classified as binary in a single content, then the content should be classified as personality or serendipity. As mentioned previously, this method can be applied to make a decision if the feature belongs to personality or serendipity based on the context using the knowledge graph on the semantic web. For instance, the ROC curve and AUC can be used to determine if the content should be included in personality and serendipity according to features like pf_i and sf_i , included in the content.

As an illustration, the features are classified by the probability density function of personality and serendipity via the extraction of the best cutoff index. In personalized recommendation system, it can make a mitigation of the side effects like confirmation bias, filter bubbles, news feed bubbles, and so on. In addition, it is can be applied to make a decision if the feature can be belonged to personality or serendipity under consideration of context using knowledge graph on the semantic web.

3) Learning in AI

Learning and training datasets in AI, the ROC curve and AUC has been actively used for gathering of the dataset that is the most crucial step. Thus, the proposed method—determination of the best cutoff index—can be applied to generate the training dataset and to determine the mixing ratio of the features in the dataset for ML according for different purposes. Each training dataset can be differently constructed by each different ratio of features like personality and serendipity, that is comprised with homogeneous or heterogeneous data, even if they come from same data pool. When selecting a dataset from the data pool, the ROC curve and AUC can be utilized to determine which feature data set to select. The cutoff index can be used to determine which of the personality features and serendipity features to configure the data set with weight depending on the purpose. For instance, the data set can be generated by the type of strong personalized, balanced, and weakly personalized.

In this study, the ROC curve and AUC were applied to determine the best cutoff index to provide a balanced information and to mitigate side effects like confirmation bias across the spectrum of preferences, which come from the personalized recommendations. In further study, the proposed method will be analyzed via simulation of actual data to find the best cutoff index for the extraction of optimized balanced information, and the effectiveness and efficiency of reducing confirmation bias in automatic recommendation systems will be verified. In addition, the proposed method will be applied to recommend a balanced information using context-based contents.

REFERENCES

[1] D. H. Ahn, “A Study on the Confirmation Bias of Adolescents for Fake News,” Journal of Communication Science, vol. 20, no. 1, 2020, pp. 77-105.

[2] G. Grosseck, and C. Holotescu, “Scholarly Digital Curation in 140 Characters,” ISSA - International Conference of Applied Social Sciences, UVT Timisoara, 2012.

[3] J. H. Im, D. S. Chang, H. S. Choe, and C. Y. Ock, "Development of Personalized Media Contents Curation System based on Emotional Information," The Journal of the Korea Contents Association, vol. 16 no. 12, 2016, pp. 181-191.

[4] M. Kang, “Does YouTube Reinforce Confirmation Bias?: A Study on the Political Uses of YouTube and Its Effects,” Journal of Speech, Media and Communication Research, vol. 20, no. 4, 2012, pp. 261-288.

[5] S. I. Park, and T. H. Oh, “Application of Receiver Operating Characteristic (ROC) Curve for Evaluation of Diagnostic Test Performance,” Journal of Veterinary Clinics, vol. 33, no. 2, 2016, pp. 97-101.

[6] A. M. Tammaro, K. K. Matusiak, F. A. Sposito, and V. Casarosa, “Data Curator’s Roles and Responsibilities: An International Perspective,” LIBRI, vol. 69, no. 2, 2019, pp. 89-104.

[7] M. D. Wilkinson, M. Dumontier, J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” Scientific Data 3, Article number: 160018, 2016.

[8] M. H. Zweig, and G. Campbell, “Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine,” Clin Chem, vol. 39, 1993, pp. 561-577.

Zero-Shot Call Classification

Nevra Akbilek
Department of Industrial Enginnering
Engineering Faculty Sakarya University
Istanbul/Türkiye
nakbilek@sakarya.edu.tr
0000-0002-9525-1755

Yunus Akkaya
Database Assistant
Research Center
CMC_Turkey
Istanbul, Türkiye
Yunus.Akkaya@cmcturkey.com
0009-0000-0181-1401

Erçin Öztuncel
General Manager
Research Center
CMC_Turkey
Istanbul, Türkiye
Ercin.Oztuncel@cmcturkey.com
0009-0009-4335-5921

Kenan Türkyılmaz
Assistant General Manager
Research Center
CMC_Turkey
Istanbul/Türkiye
Kenan.Turkyilmaz@cmcturkey.com
0009-0002-3478-8222

Abstract— Traditional text classification methods required a training dataset for the training model. Nevertheless, finding data for the classification of some things in real-life problems is not easy. In this study, call recordings in a call center service were first converted into text and analyzed. Then, we classified the call center text datasets with not train datasets. Furthermore, we clustered by the k-means algorithm to analyze the result and compared outputs. So, from the conversations, the data of 20,000 Turkish phone conversations coming to a customer call center of a brand serving the electronic products sector, the topic determination was done using classification.

Text classification—zero-shot learning, clustering, deep learning, k-means clustering

I. INTRODUCTION

Text classification is crucial for natural language processing(NLP) practices like classifications of news, user, and question [1]. There are many supervised text classification methods, but they have a deficiency in arising unlabeled training data [2]. In recent years, especially GZSL has been used for classifying the images [3,4,5] and questions [6].

The critical issue here is how to learn helpful information for unseen classes from seen ones gradually. One approach is to assume the information about unseen classes to use generative models for creating examples and features for unseen areas [4,7,8,9]. Generative methods convert the GZSL Problem to a traditional supervised problem. Also, some works extend to using unlabeled data for unseen classes [10,11].

Traditional text classification methods required a training dataset for the training model. However, it is not easy to find data for classifying some things in real-life problems. In this work, we classified some call center text datasets with not train datasets. There is no generally accepted assessment area for zero-shot learning, and it is sometimes impossible to compare results from different studies. A study in 2020 defines a new benchmark to solve this task and make classification with varying types of zero-shot learning. In that paper zero-shot method was used for image classification and all results were benchmarked and systematically evaluated [12].

In another study for improving the effectiveness of zero-shot learning, they proposed a new approach named graph active zero-shot learning. Designed a new method-based k-center algorithm, and Laplacian energy is used. Usually, it is able used for image classification. Zero-shot learning can detect the relations of the classes, but the results need to be improved in terms of efficiency. At this point, they significantly increased the efficiency obtained with the new method they designed [13].

The traditional text classification method required a training dataset for the training model. However, it is not easy to find data for classifying some things in real-life problems.

In a study conducted in 2019, a document classification system based on zero-shot is being made. However, the work's approach differs from other one-shot or zero-shot learning algorithms. One-shot learning requires labeled data for training, and zero-shot learning test data differs from training data. So they call that cross-lingual zero-shot classification, and it is helpful for 100 languages, and they can classify a document in a language to another language. This method works by computing similarity with potential labels and documents [14].

In the zero-shot learning method, the point is to detect the relations amount of the classes, and to do that; we need transformers to get good performance.
The transformers used for zero-shot learning are usually pre-trained with a more extensive database, such as Wikipedia. So that it cannot be customized for a specific dataset [15].

When the calls were first received, they were in audio format. As the first step in this study, we had to convert the voices into text. To do that, we used the face recognition module in python. Speech recognition is an ability for machines to listen to voices from around and identify the words in a speech to transform them into a text form. We classified some call center text datasets with not-train datasets in this work using a zero-shot classification method. However, unfortunately, there is no satisfactory system to inspect and rate this method. For this study, as a solution for this problem, we made a clustering using a k-means algorithm, visualized the results, and analyzed the similarity between the outputs of the two methods.

II. MATERIALS AND METHODS

There is more than one common task for text classification, such as sentiment analysis, topic labeling, news classification, and question answering, and to do that, we need some algorithms. However, these algorithms required a dataset to teach the model. According to the amount of dataset, there is called n-shot learning. In this study, we classified using zero-shot learning to classify call center datasets with no labeled data for training the model. Zero-shot learning was the first time in a paper 2008 at AAAI 2008. However, that paper mentioned data less classification instead of zero-shot learning.

Zero-shot classification: Conventional machine learning algorithms require training and test data to train a model. However, today, when dealing with real problems in machine learning studies, there is not always enough data to use to train

the model. In such cases, classification can be made on limited or no training data using n-shot algorithms. For example, if we try to classify an image only with one image, it is a one-shot classification [3,13].

The working principle of this classification technique is based on finding certain features on the data to be classified, classifying the data with these features in common, and separating those that are not. There are two types. According to the training data;

Inductive zero throws: We have access to labeled data and known class information in this type. Classification is done on unknown data from a particular tag class.

Transudative zero shot: This type has both known and unknown tag definitions. It is not always possible to label every data in this way. This type is suitable for more straightforward use because it has definitions of unknown class properties. In this study, this method was used because the speech data did not have a label. According to test data;

Traditional zero-shot learning: The data to be classified in this model belong only to the invisible classes. Realistic can be said to be a little less valuable. However, we can improve the result by giving the model side information about the classes.

Generalized zero-shot learning: The data to be classified in this method can be new or known data. Here the model has to predict whether it is classifying new or test data [16].
In this study, it is clear which classes we will use to classify the calls made while the call is received, but since we do not have any labeled training data, traditional zero-shot classification was used.

K-means clustering: K- means is one of the most popular unsupervised learning algorithms. It is an algorithm for separating a given amount of data into their similarity. It is one of the most used clustering algorithms due to its ease and functionality. The point of this algorithm is to determine data that represents its cluster (k symbolizes the class number) and determine the other members of the cluster according to the distance of the selected data from the other data. The algorithm consists of 4 steps:
- Determine the center of the cluster
- Clustering the rest of the data according to the distance from the center
- Upgrading the centers of the clusters according to the new shapes of clusters
- Repeating steps two and step 3 until stabilizing the results.

Multinominal Logistic Regression: Logistic regression is a statistical method for classifying binary data for his study; we used multinomial logistic regression to see the consistency of the outputs of the zero-shot classifying. For this method, there are many formulas to describe the method. Nevertheless, the point of all of them is the set weights to optimize results.

III. IMPLEMENTATION AND DISCUSSION

For this study, we used call speech data as a text file taken from the call center. The aim, labeling the texts by 11 subjects we already have. There are some examples from the dataset we used for zero-shot classification and k-means clustering;

TABLE I. CUSTOMER TEXT ABOUT THE SPEECHES

Id	CallId	AgentText	CustomerText	AgentTex	CustomerTextTimes
4601	18636971	iyi günler	teşekkürler size i e beni	4.10 4.344	7.93 8.989 1
4602	18636988	****** be	alo merhaba iyiyim sağ	2.93 3.503	4.24 4.665 1
4603	18636999	iyi günler	model alo iyi günler efe	2.54 2.753	0.28 0.616 1
4604	18637004	iyi günler	e iyi günler **** bey ev	2.40 2.612	6.64 6.807 1
4605	18636969	iyi günler	*d alo he buyurun evet	9.30 9.519	1.90 3.528 1

One hundred call texts were used and labeled under 11 titles in the classification made. As a result, 100 call data; 'Dealer: 4', 'Info: 3', 'Exchange: 27', 'Returns: 2', 'Warranty: 18','Campaign: 0', 'Installation: 4', 'Service 12: ', 'Authorized Call: 3', 'Complaint: 0', 'Other: 27'.

Fig. 1. Zero-shot classification percentages

When we clustered the texts without any known label by the k-means algorithm, we got a result like in the graphic.

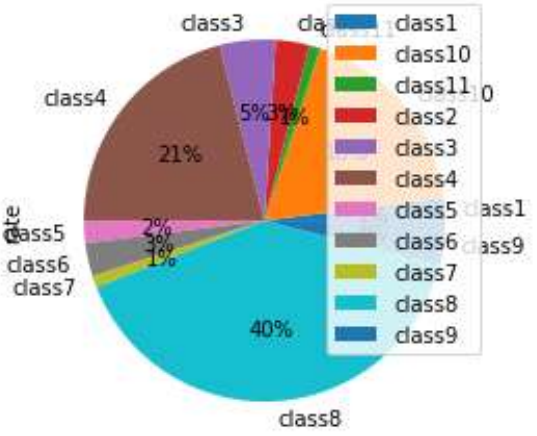


Fig. 2. Zero-shot classification percentages

When the density of the distributions is examined, the general similarity is noticed. If we compare the classes closest to each other as a percentage;

TABLE II. THE RESULTS OF K-MEANS

class8	=>	Other	40
class4	=>	Change	21
class10	=>	Guarantee	18
class3	=>	Service	5
class9	=>	Dealer	4
class2	=>	Institution	3
class6	=>	Information	3
class1	=>	Authorized Call	2
class5	=>	Return	2
class7	=>	Offer	1
class11	=>	Complaint	1

TABLE III. THE RESULTS OF ZERO-SHOT CLASSIFICATION

Other	27
Change	27
Guarantee	18
Service	12
Dealer	4
Institution	4
Information	3
Authorized Call	3
Return	2
Offer	0
Complaint	0

Table 2 shows that k-means clustering without the training dataset and labels. Table 3 shows the zero-shot classification results with labels but without the training dataset. Moreover, we match clusters that have the closest results.

If we compare according to the nearest percentiles and find the error rate;

Error rate = (∑ (k-means result - zero-shot result)/k-means) (1)

All error rates for eleven labels are given in Table 4.

TABLE IV. THE DIFFERENCE BETWEEN K-MEANS AND ZERO-SHOT RESULTS AS PERCENTAGE

Other	1.00%
Change	1.00%
Guarantee	0.00%
Service	0.50%
Dealer	0.00%
Institution	0.33%
Information	0.00%
Authorized Call	1.40%
Return	0.00%
Offer	0.29%
Complaint	0.33%
Total error rate	4.844

Table 4 shows the difference between k-means results and zero-shot results as percentages. The margin of error in Table 4 varies between 0 and 1.4 for each class. The total error was calculated as 4,844. In other words, there is a 95,156% similarity between the clustering study without titles and the zero-shot classification method, which gives titles without a training data set. Also, when we classified the zero-shot results to see consistency using multinomial logistic regression, we got a 0.38 F 1 score. In the same way, when we clustered with k means and evaluated them according to zero-shot, we got an f1 score of about 0.67.

IV. CONCLUSION

In this study, we worked on classifying some call center data without the training dataset by using zero-shot learning. After then, to determine the consistency of the results for that classification, we used multinomial logistic regression, and we got an F1 score: of 0.38. Moreover, we make a k-means clustering with no labels and train data to compare. We got an F1 score: of 0.67, and when we calculated the similarity rate, we got 0.95. According to this rate, we can say that the results we are getting from the zero-shot classification are self-consistent and very similar to the clustering with no labels.

REFERENCES

[1] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao (2021). Deep learning--based text classification: a comprehensive review. *ACM Computing Surveys (CSUR)*, *54*(3), 1-40.

[2] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C. P. Lim, X. Z. Wang, Q. M. J. Wu (2022). A review of generalized zero-shot learning methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[3] R. Socher, M. Ganjoo, C. D. Manning, & A. Ng (2013). Zero-shot learning through cross-modal transfer. *Advances in neural information processing systems*, *26*.

[4] Y. Xian, C. H. Lampert, B. Schiele & Z. Akata (2018). Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE transactions on pattern analysis and machine intelligence*, *41*(9), 2251-2265.

[5] W. Wang, V. W. Zheng, H. Yu & C. Miao (2019). A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-37.

[6] H. Fu, C. Yuan, X. Wang, Z. Sang, S. Hu, Y. Shi (2018, November). Zero-shot question classification using synthetic samples. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 714-718). IEEE.

[7] E. Schonfeld, S. Ebrahimi, S. Sinha, T. Darrell, & Z. Akata (2019). Generalized zero-and few-shot learning via aligned variational autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8247-8255).

[8] J. Zhang, P. Lertvittayakumjorn & Y. Guo (2019). Integrating semantic knowledge to tackle zero-shot text classification. *arXiv preprint arXiv:1903.12626*.

[9] C. Song, S. Zhang, N. Sadoughi, P. Xie, & E. Xing (2020). Generalized zero-shot text classification for icd coding. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 4018-4024).

[10] S. Rahman, S. H. Khan, and N. Barnes. 2019. Transductive learning for zero-shot object detection. In 2019 IEEE/CVF International Conference on Computer Vision, pages 6081–6090, Seoul, Korea (South). IEEE

[11] Z. Ye, Y. Geng, J. Chen, X. Xu, S. Zheng & H. Chen (2020, July). Zero-shot text classification via reinforced self-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3014-3024).

[12] Y. Xian, Student Member, IEEE, Christoph H. Lampert, Bernt Schiele, Fellow, IEEE, and Zeynep Akata, Member, IEEE, Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly ,2022

[13] Q. Wang, W. Wu, Y. Zhao, & Y. Zhuang (2021). Graph active learning for GCN-based zero-shot classification. *Neurocomputing*, *435*, 15-25.

[14] Y. Song, S. Upadhyay, H. Peng, S. Mayhew, & D. Roth (2019). Toward any-language zero-shot topic classification of textual documents. *Artificial Intelligence*, *274*, 133-150.

[15] A. Alcoforado, T. Palmeira Ferraz, R. Gerber, E. Bustos, A. S. Oliveira, B. M. Veloso, F. L. Siqueira, A. H. R. Costa (2022, March). ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling. In *International Conference on Computational Processing of the Portuguese Language* (pp. 125-136). Springer, Cham.

[16] X. Li, M. Fang, B. Chen (2021). Generalized zero-shot classification via iteratively generating and selecting unseen samples. *Signal Processing: Image Communication*, *92*, 116115.

Design and Analysis of a Highly Sensitive GeS-based SPR Biosensor for DNA Detection

Md. Saiful Islam
School of Engineering
Military Technological College
Muscat, Oman
saiful.islam@mtc.edu.om

Abbas Z. Kouzani
School of Engineering
Deakin University
Geelong, Australia
kouzani@deakin.edu.au

Shekhar Mahmud
School of Engineering
Military Technological College
Muscat, Oman
shekhar.mahmud@mtc.edu.om

Nasra Al Sharji
School of Engineering
Military Technological College
Muscat, Oman
Nasra.alsharji@mtc.edu.om

George Chen
Electro. &Comp. Sci. Engg.
University of Southampton
Southampton, UK
gc@ecs.soton.ac.uk

***Abstract*— In this paper, a germanium sulfide (GeS) based surface plasmon resonance (SPR) biosensor is designed and analyzed, consisting of fused silica prisms, gold, graphene, and GeS layers. The aim of this proposed SPR biosensor is to enhance the performance parameters including the accuracy of detection, sensitivity and quality factor by monitoring the recombination of cDNA and ssDNA on its sensing surface. The proposed structure is designed and numerically analyzed using FDTD technique through a Multiphysics simulation tool. A good agreement was found between the simulation results and those derived from the theoretical formula using Fresnel equations and matrix theory. As compared to recent publications, the proposed biosensor exhibits significant improvements in performance parameters. In addition, Furthermore, the FWHM is examined in relation to the number of graphene layers.**

Keywords—Biosensor, surface plasmon resonance, sensitivity quality factor, GeS.

I. INTRODUCTION

SPR-based biosensors are very effective tools which can be used for biomolecular interaction studies, chemical detection, and immunoassays. In this technique, a surface plasmon carries energy through the metal surface, creating transient electric fields that propagate down the plasmon-excited metal surface to the metal surface plasmon source and are recombined with the metal surface plasmon into another excited state that propagates through the metal surface again. As a label-free, fast and sensitive optical biosensing technique, SPR has been comprehensively researched during the past years. Due to their label-free detection protocol, SPR biosensors can be used in a wide range of medical and biological applications. The technique is widely used for a wide range of lab-on-a-chip sensors such as chemical sensing, biosensing, gas sensing, immunoassays, and food safety monitoring [1–3]. Attenuated total reflection, optical fibers, optical waveguides, and intensity measurements are among the available platforms for SPR sensing [4–6].

With the unique features of surface plasmon resonance biosensors (SPRBs) such as label-free detection, spectral tuning, rapid and direct analysis, local electric field enhancement, they are best suited for a wide area of biomedical applications. Among many of the approaches to improve sensitivity, It has been shown that silica nanosheets and transition metal dichalcogenides based SPR biosensors have improved sensitivity [7]. To improve the sensitivity of this sensor, a change in refractive index (RI) is introduced at the interface between detecting medium and sensor. Recent studies have investigated the effect of various adhesion layers including high refractive index germanium semiconductors and indium tin oxide transparent conductors, on SPR biosensors' performance [7], [8].

A few recent papers have reported the SPR graphene biosensor [9–11]. Qi Wang et al. [12] reported bimetallic SPR biosensor based on grapahene to improve the detection accuracy where silver was on the base of a prism. However, often the layer become silver oxide after the exposure of oxygen and thus creates a false detection. Very recently, Yue Jia et al. [13] reported a SPR biosensor with GeS metal layers. Although the detection accuracy is improved, sensitivity remained the same when compared with other reported works which is possibly due to the absence of graphene layer.

In this paper, a SPR biosensor is proposed which contains a gold layer at the base of prism and successive layer of graphene and (GeS) and showed superior performance when compared against other sensing layers such silicon nitride (Si_3N_4) and tungsten disulfide (WS_2).

II. PROPOSED BIOSENSOR

An illustration of 2-D representation of proposed SPR biosensor is illustrated in Fig. 1. The Kretschmann configuration is employed in our investigation of a prism against a gold thin film. And, As a layer of biological sensing, GeS is layered over a thin gold film followed by graphene. The corresponding properties of GeS sensing layer is considered provided that the sensing surface is functionalized with the receptor biomolecule. The movement of the resonance peak in the SPR optical spectrum is detectable due to adsorption of target biomolecules on the sensing surface. As shown in Fig. 2, the resonance peak shift is resulted by the interaction of target biomolecule indicates the detection of the target biomolecule.

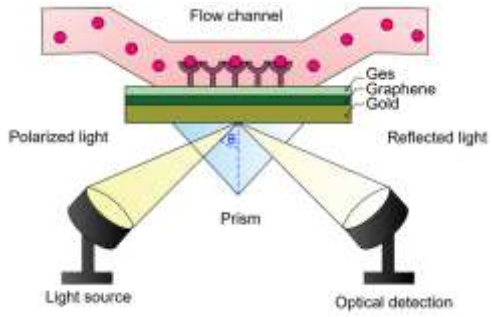


Fig. 1. Schematic representation of the proposed SPR biosensor with GeS combined on graphene thin layer in a basic Kretschmann configuration.

Through the prism, a TM polarized plane wave with a fixed wavelength of 632.8 nm is coupled to excite SPs. A fused silica glass coupling prism is used. The thickness of the gold thin film, graphene layer and GeS is selected as 53 nm, 0.335 nm and 0.52 nm , respectively. According to the Drude model, gold has a wavelength dependent complex valued RI [14]. The Sellmeier dispersion formula is used to derive the RI of fused silica-glass prisms[15] . The RI of GeS is a complex-valued function as obtained from [16].

The ssDNA is immobilized with a GeS sensing layer that is linked to the corresponding complementary DNA to facilitate recombine into dsDNA. When the target DNA (cDNA) binds to its receptor counterpart, the resonance changes, which results in a shift of SP dip as shown in Fig. 2. The initial RI considered due to the formation of dsDNA is 1.452 [27]. As a result of the formation of dsDNA at a density of 0.061 g/cm^3, we obtained an RI of 1.52 [17].

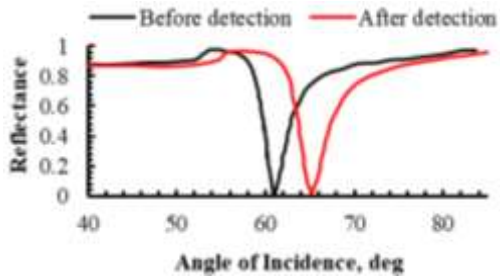
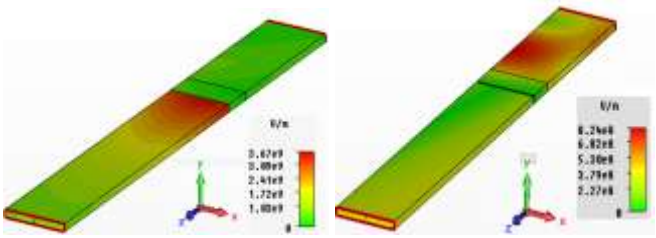


Fig. 2. Reflectance versus incidence angle curve to illustare the DNA binding activity through the shift plasmon dip.

Based on the well-established FDTD methodology, the proposed biosensor is designed and implemented. This simulation is validated by comparing the result with the standard method of calculating electromagnetic fields using Fresnel equations and matrix-based methods. To analyze performance parameters, The simulation includes the RI profile as well as various design parameters including thickness of graphene layer, GeS layer, and type of sensing material). Based on this method, each unit cell is considered separately and then the total response is calculated in greater detail. A 53 nm gold layer is formed first on the top of prism's base through extruding it on its top face. GeS is formed by extruding the top face of graphene with 0.52 nm and graphene is formed through extruding top of the gold layer with 0.335 nm. Fig.3 shows the numerical simulation of propagation of surface plasmons waves.



(a) (b)

Fig. 3. Numerical simulation illustrating the contour plot of absolute electric field: (a) with resonance and (b) without resonance.

III. THEORETICAL BACKGROUND

At the interface between the two media, an SPR system is well described by Kretschmann configuration and Maxwell's electromagnetic equations. For the SPs to be excited on the metal-dielectric interface, the following conditions must be met [18]:

$$k_{sp} = k_{ev} = \frac{2\pi}{\lambda}\sqrt{\frac{\varepsilon_M\,\varepsilon_{D,eff}}{\varepsilon_M + \varepsilon_{D,eff}}} = \frac{2\pi}{\lambda}\sqrt{\varepsilon_p}\,\sin\theta_{res} \qquad (1)$$

where surface plasmon propagation constant is known as k_{sp} and evanescent wave propagation constant is known as k_{ev}. Operating wavelength of excitation light is denoted by λ. A dielectric constant for a metal layer and a dielectric constant for a dielectric layer are determined by ε_M and $\varepsilon_{D,eff}$, respectively. θ_{res} stands for resonance coupling angle. In order to construct the SPR curve, Fresnel reflection coefficients are calculated using the following transfer matrix formalism [19], [20]:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = [M] \times \begin{bmatrix} E_{N-1} \\ H_{N-1} \end{bmatrix} \qquad (2)$$

where

$$[M] = \prod_{k=2}^{N-1} M_k = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \text{ with k= 2, 3, 4, 5.....N-1} \qquad (3)$$

with,

$$M_k = \begin{bmatrix} \cos\beta_k & (-i\sin\beta_k)/q_k \\ -i\,q_k\sin\beta_k & \cos\beta_k \end{bmatrix} \quad (4)$$

Lastly, the reflectance R_p for the TM-polarized light of our proposed biosensor is derived as:

$$R_p = \left| \frac{(M_{11} + M_{12}q_N)q_1 - (M_{21} + M_{22}q_N)}{(M_{11} + M_{12}q_N)q_1 + (M_{21} + M_{22}q_N)} \right|^2 \quad (5)$$

The complex RI of graphene thin film in the visible range is given below which is based on a recent experiments in Fresnel coefficient calculation [21]–[23]:

$$n_G(\lambda) = 3 + i\frac{5.446}{3}\lambda_0 \quad (6)$$

where λ_0 denotes the wavelength of vacuum in μm.

IV. RESULTS AND DISCUSSIONS

A. Reflectance

The SPR reflectance characteristics of the proposed structure is studied at different thickness of GeS sensing layer while the thickness of gold and graphene is kept constant at 53 nm and 0.335 nm respectively. The result of this investigation is shown in Fig. 4. It is obvious from Fig. 4 that the resonance dip shifted to higher resonance angle as the number of GeS layer is increased. This is clearly an indication of improved sensitivity.

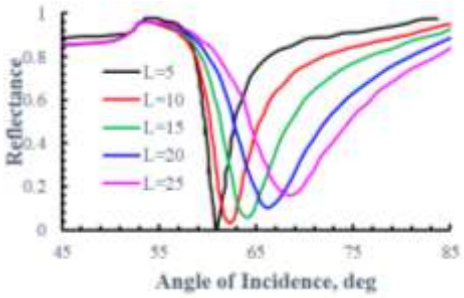


Fig. 4. Reflectance versus incidence angle curve with numbers of GeS layers for the proposed SPR biosensor.

B. FWHM

FWHM is an indication of the detection accuracy, which is investigated against the number of graphene layer. As shown in Fig. 5, FWHM increases steadily with the increase of the number of graphene layer. The slope of FWHM for the proposed GeS-based biosensor is steeper compared to Si_3N_4 and WS_2 structure. As number of graphene layer increases, the SPR curve becomes wider which degrades the detection accuracy. The proposed SPR biosensor works best with monolayer graphene sheet for better detection accuracy.

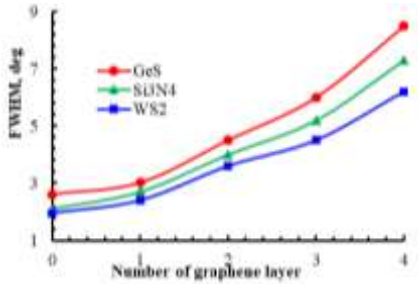


Fig. 5. Variation of FWHM for different number of graphene layers for the proposed SPR biosensor.

C. Sensitivity and Detection Accuracy

The sensitivity (S) and detection accuracy (D.A) are most crucial performance parameters, which are defined as:

$$S = \frac{\Delta\theta_{spr}}{\Delta n_s} \text{ [deg/RIU]} \quad (9)$$

$$D.A = \frac{\Delta\theta_{spr}}{\Delta\theta_{0.5}} \quad (10)$$

where, θ_{spr} and Δn_s are the resonance incidence angle and change of RI, respectively. $\Delta\theta_{0.5}$ is same as FWHM. Fig. 5 shows the variation of sensitivity and detection accuracy with different number of GeS sensing layers at 632.8 nm. As can be seen from Fig. 4 that GeS plays an important role to the improvement of sensing parameters. As L increases, an upward trend and downward trend are observed for sensitivity and detection accuracy, respectively. Although D.A degrades with the increase of L, it remains high compared to other biosensor structure.

D. Quality Factor

The quality factor (Q.F) is an important performance parameter of a biosensor, which is investigated against the number of graphene layer. Q.F is defined as:

$$Q.F = \frac{S}{\Delta\theta_{0.5}}\left(RIU^{-1}\right) \quad (11)$$

Fig. 5 illustrates the variation of Q.F against the number of graphene layers which decreases with the number of graphene layers. We obtained an optimum Q.F of 44 RIU^{-1}, 40 RIU^{-1}, and 34 RIU^{-1} for GeS, Si_3N_4 and WS_2 based biosensors.

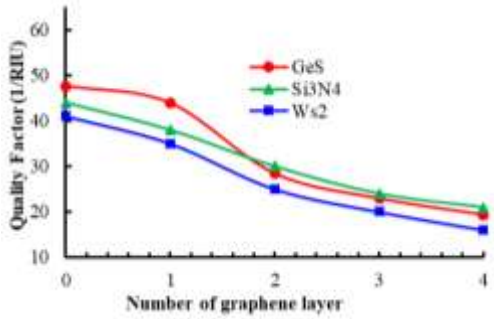


Fig. 5. Variation of quality factor for different number of graphene layers for the proposed SPR biosensor.

V. CONCLUSIONS

Biosensor based on SPR is designed and investigated in this study to enhance the sensing parameters including sensitivity, detection accuracy and, quality factor. The performance of the proposed sensor is compared with other recently published biosensors including Si_3N_4 and WS_2 based structures. It is found through the numerical modeling and simulation that the proposed GeS based SPR sensor has the superior performance against Si_3N_4 and WS_2 based structured SPR biosensors. The performance parameters are also investigated for different number of graphene and GeS layers. Investigation revealed that monolayer graphene with 10 layers of GeS provides optimized sensitivity and D.A.

REFERENCES

[1] R. D'Agata, N. Bellassai, V. Jungbluth, and G. Spoto, "Recent advances in antifouling materials for surface plasmon resonance biosensing in clinical diagnostics and food safety," Polymers (Basel)., vol. 13, no. 12, p. 1929, 2021.

[2] R. Gaur, H. M. Padhy, and M. Elayaperumal, "Surface plasmon assisted toxic chemical NO 2 gas sensor by Au/ ZnO functional thin films," J. Sensors Sens. Syst., vol. 10, no. 2, pp. 163–169, 2021.

[3] H. H. Nguyen, J. Park, S. Kang, and M. Kim, "Surface plasmon resonance: a versatile technique for biosensor applications," Sensors, vol. 15, no. 5, pp. 10481–10510, 2015.

[4] M. S. Islam and A. Z. Kouzani, "Variable incidence angle localized surface plasmon resonance graphene biosensor," in The 2011 IEEE/ICME International Conference on Complex Medical Engineering, 2011, pp. 58–63, doi: 10.1109/ICCME.2011.5876705.

[5] J. Dostalek et al., "Surface plasmon resonance biosensor based on integrated optical waveguide," Sensors actuators B Chem., vol. 76, no. 1–3, pp. 8–12, 2001.

[6] H. Fu, S. Zhang, H. Chen, and J. Weng, "Graphene enhances the sensitivity of fiber-optic surface plasmon resonance biosensor," IEEE Sens. J., vol. 15, no. 10, pp. 5478–5482, 2015.

[7] Q. Ouyang et al., "Sensitivity enhancement of transition metal dichalcogenides/silicon nanostructure-based surface plasmon resonance biosensor," Sci. Rep., vol. 6, no. 1, pp. 1–13, 2016.

[8] A. Gupta, S. S. Gaur, H. Sonawane, and D. S. Maske, "Numerical study of multilayer surface Plasmon resonance for biosensing application using III-V semiconductor GaSb," Mater. Today Proc., 2022.

[9] M. G. Daher, S. A. Taya, I. Colak, S. K. Patel, M. M. Olaimat, and O. Ramahi, "Surface plasmon resonance biosensor based on graphene layer for the detection of waterborne bacteria," J. Biophotonics, vol. 15, no. 5, p. e202200001, 2022.

[10] K. N. Shushama, M. Rana, R. Inum, and M. Hossain, "Sensitivity enhancement of graphene coated surface plasmon resonance biosensor," Opt. Quantum Electron., vol. 49, no. 11, pp. 1–13, 2017.

[11] M. S. Islam and A. Z. Kouzani, "Variable incidence angle subwavelegth grating SPR graphene biosensor," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 3024–3027, doi: 10.1109/EMBC.2013.6610177.

[12] Q. Wang, S. Cao, X. Gao, X. Chen, and D. Zhang, "Improving the Detection Accuracy of an Ag/Au Bimetallic Surface Plasmon Resonance Biosensor Based on Graphene," Chemosensors, vol. 10, no. 1, p. 10, 2021.

[13] Y. Jia, Y. Liao, and H. Cai, "Sensitivity Improvement of Surface Plasmon Resonance Biosensors with GeS-Metal Layers," Electronics, vol. 11, no. 3, p. 332, 2022.

[14] B. D. Gupta and A. K. Sharma, "Sensitivity evaluation of a multi-layered surface plasmon resonance-based fiber optic sensor: a theoretical study," Sensors Actuators B Chem., vol. 107, no. 1, pp. 40–46, 2005, doi: https://doi.org/10.1016/j.snb.2004.08.030.

[15] I. D. Rukhlenko, A. Pannipitiya, and M. Premaratne, "Dispersion relation for surface plasmon polaritons in metal/nonlinear-dielectric/metal slot waveguides," Opt. Lett., vol. 36, no. 17, pp. 3374–3376, 2011.

[16] J. Xie et al., "Optical properties of chemical vapor deposition-grown PtSe2 characterized by spectroscopic ellipsometry," 2D Mater., vol. 6, no. 3, p. 035011, 2019.

[17] K. Kim, S. J. Yoon, and D. Kim, "Nanowire-based enhancement of localized surface plasmon resonance for highly sensitive detection: a theoretical study," Opt. Express, vol. 14, no. 25, pp. 12419–12431, 2006.

[18] J. Homola, "Electromagnetic theory of surface plasmons," in Surface plasmon resonance based sensors, Springer, 2006, pp. 3–44.

[19] M. S. Islam and A. Z. Kouzani, "Simulation and analysis of a sub-wavelength grating based multilayer surface plasmon resonance biosensor," J. Light. Technol., vol. 31, no. 9, pp. 1388–1398, 2013.

[20] L. Wu, H.-S. Chu, W. S. Koh, and E.-P. Li, "Highly sensitive graphene biosensors based on surface plasmon resonance," Opt. Express, vol. 18, no. 14, pp. 14395–14400, 2010.

[21] M. Islam, A. Z. Kouzani, X. J. Dai, W. P. Michalski, and H. Gholamhosseini, "Design and analysis of a multilayer localized surface plasmon resonance graphene biosensor," J. Biomed. Nanotechnol., vol. 8, no. 3, pp. 380–393, 2012.

[22] R. R. Nair et al., "Fine structure constant defines visual transparency of graphene," Science (80-.)., vol. 320, no. 5881, p. 1308, 2008.

[23] M. Bruna and S. Borini, "Optical constants of graphene layers in the visible range," Appl. Phys. Lett., vol. 94, no. 3, p. 031901, 2009.

Floating Solar Photovoltaic as an ideal technology for distributed generation in developing countries: Pakistan Prospective

Majid Ali
AAU Energy
Aalborg University
Aalborg, Denmark
maal@energy.aau.dk

Muhammad Mashhood
IEEE
Punjab University
Lahore, Pakistan
mashhood786@hotmail.com

Hassan Zeb
IEEE
Punjab University
Lahore, Pakistan
hassanzeb.ieee@pu.edu.pk

Juan C. Vasquez
AAU Energy
Aalborg University
Aalborg, Denmark
juq@energy.aau.dk

Josep M. Guerrero
AAU Energy
Aalborg University
Aalborg, Denmark
joz@energy.aau.dk

Abstract: ***Countries around the globe, in particular, South Asia, are facing both energy and water shortage. In South Asia countries like Pakistan and India, this shortage is mainly because of poor electricity dispatch due to a lack of transmission system, higher line losses, and saltish water with no filtration facility. Secondly, with the rising temperature over the past fifty years, the evaporation rate has also increased, leading to higher evaporation losses, especially during summer. This research focuses on the potential of floating photovoltaic technology in Pakistan on various types of water bodies that will encourage distributed generation and reduce evaporation. Over the years, this technology has picked up pace across the globe, especially in Asia, and will prove vastly viable as it helps reduce evaporation, clean energy production, and shrink GHG emissions.***

***Keywords*—floating solar, solar, distributed generation, water conservation, GHG reduction**

I. INTRODUCTION

Floating solar photovoltaic (FSPV) is a nascent technology compared to conventional solar power plants and emerged in 2007 to substitute (conventional) land-based solar PV power plants for countries with growing populations and rising need for land for cultivation and real estate purposes [1]. The first floating small-scale solar plant was installed in 2007 in Japan, and later in 2013, the first utility-scale FSPV plants started their operation and have gained pace across the world and in Asia specifically. Countries like the UK, Singapore, Japan, South Korea, and China are in conquest to cut pressure on their land while contributing to the commitment strengthened in COP 21 in December 2015, Paris [2] to address climate change concerns and devise plans and policies to mitigate it. The range of capacities of operational utility-scale Floating Solar PV projects is from three (03) to hundred (100) MWp capacity and even larger ones are under construction in China, South Korea, and Japan [3]. With the growing With population that will reach three billion by 2050, issues related to the acquisition of land will be a significant barrier for Pakistan to meet its food, water, and energy demand by 2050. Hence, the Government of Pakistan (GoI) needs to explore the option of using the surfaces of water bodies, especially man-made reservoirs. With the ARE 2019, Pakistan has set to achieve 30% of its electrical generation from renewables, a high target since it stands at 4% as per the Energy Year Book 2019. The critical drivers toward deployment of FSPV systems in Pakistan are the resolution of the issues related to land, greater energy yield due to lower temperature over water bodies, reduction in water loss through evaporation, and opportunity to use existing infrastructure at the hydroelectric sites, reducing the investment making it FSPV more feasible in the competitive markets like Pakistan. The evolution of FSPV technology on a global level over the years concerning its installed capacity showing its acceptability and market penetration is illustrated in fig. 1.

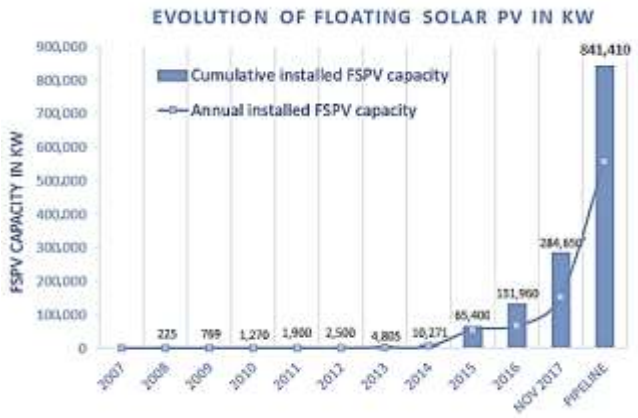


Fig 24. Evolution of FSPV Technology on a Global Level

A. Energy Mix of Pakistan

As of 2017-2018, electricity generation increased to 33,554 MW from 29,944 MW in 2016-2017, mainly due to four new thermal power plants in the national grid having a combined capacity of 2,799 MW. During the year 2017-2018, renewable energy also rose to 1,637 MW from 1,237 MW in 2016-17. The total renewable energy injected into the national grid from all sources, i.e., Solar, Wind, and Bagasse, is 3,857 GWh. Renewable energy makes up 2.9% of the total (electricity generated) 131,275 GWh and increased by 44.6% from the previous year. It is pertinent to note that hydel generation reduced by -13.2% in 2017-2018 [4]. As per the recent ARE

2019 Policy, Pakistan has set a target to achieve 15% of the total energy generated from renewables, further enhanced to 30% of the total mix by 2030. This target is over-ambitious, keeping in view that the installed RE capacity stands at 4% of the entire blend, excluding hydel [5].

TABLE 15. TREND OF THE CONTRIBUTION OF RENEWABLE ENERGY TOWARDS THE ENERGY MIX [4]

Year	*Power (GWh)*	*Total Power Exported (GWh)*	*Percentage of the Total Energy Mix (%)*
2014-15	802.00	107,408.00	0.75
2016-16	1,549.00	111,763.00	1.39
2016-17	2,668.00	123,614.00	2.16
2017-18	3,857.00	131,831.00	2.93

Therefore, to meet the target of 30% renewable energy in the total mix from the existing 4% is a challenging flight and requires huge investment and infrastructure development. However, FSPV is a technology that can use existing infrastructure and water bodies to serve as one of the best solutions to increase renewable energy contribution to the total mix in the short term.

B. FSPV for Pakistan

It is a nascent technology compared to land-based solar. In this setup, PV panels are designed and installed on water bodies such as reservoirs of hydro dams, effluent treatment plants, mining ponds, lakes, and lagoons [6]. There is a significant power generation potential for (Floating and land-based) Solar Photovoltaic systems in Pakistan, with average GHI values of over 1,800 kWh/m^2 [1]. Baluchistan receives the most significant amount of solar irradiation, with GHI levels of over 2,000 kWh/m^2, followed by Punjab and Sindh's southern regions. The least amount of irradiation is in mountain ranges located in the northern areas of Pakistan. With the hybrid model of PHES and FSPV, the intermittency of the solar can be handled [7]. The average GHI for each province of Pakistan is illustrated in fig. 2.

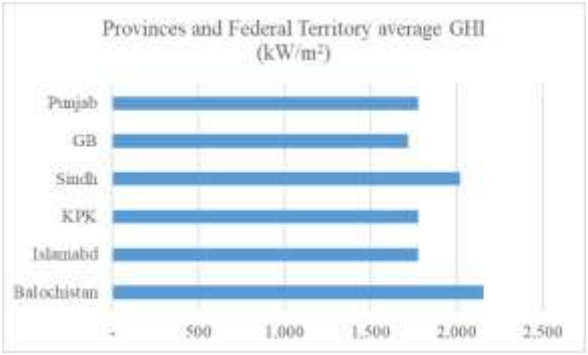


Fig. 2. Potential across Provinces, Federal Territory, and Average GHI

C. Benefits of FSPV:

FSPV has certain advantages over land-based solar, such as higher gains in energy production due to (comparatively lower) optimum temperature over water bodies than on land [9]- [6] and the FSPV is 1.5-2% more efficient than a conventional solar arrangement [10]. The next advantage is as FSPV is installed on water bodies, so it is land neutral [11]. It has come forward through research that with the current increase in population, there is an increase in demand and a reduction in drinking water supply; Pakistan is on the list of the world's most water-stressed countries [12]. With the installation of floating solar panels on water bodies, it provides a cover that contributes to the reduction in water evaporation which is highly lucrative for South Asian countries like Pakistan and India, where the evaporation rate is high in arid regions over recent years due to rise in temperature as shown in figure 3. SPG Solar claims that a reduction in evaporation rates with the installation of Floating Solar Panels is up to 70%, but no evidence has yet been put forward to back this claim [13].

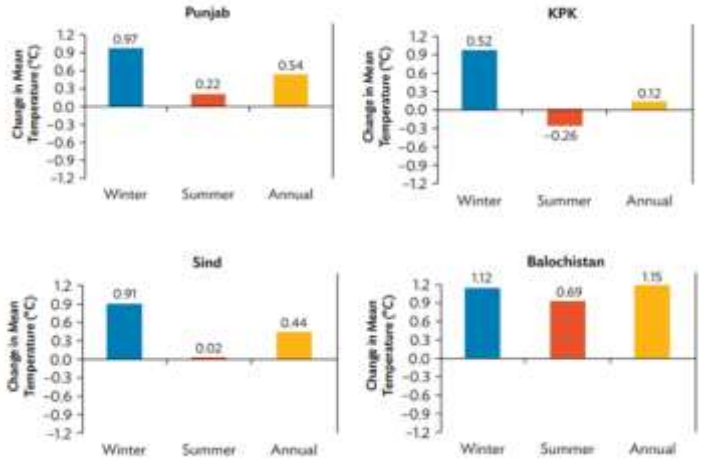


Fig 3. Change in Mean Temperature Across Pakistan [8]

In addition to all these advantages, in some instances, FSPV can share existing electrical infrastructure for power evacuation like dams and barrages [14]. It is a new source of revenue from existing redundant assets [6] FSPV cleaning is also convenient and economical compared to ground-mounted PV systems as it is located over the water source and faces less soiling loss because of less dust in the wind blowing over water bodies [1]. Mass production of platforms used for mounting made of potable water grade HDPE can make the FSPV more cost-effective than ground-mounted Solar PV. Therefore, with all these advantages and the utilization of waterbodies for the installation of FSPV makes sense, especially with the fast-growing population, the need for water, energy, and land will increase for food and infrastructure [15]. There is also important to note that Pakistan hosts 7,253 glaciers, which include 543 located in the Chitral Valley. According to various surveys and research, Pakistan has the largest reserve of glacial ice than anywhere on the planet outside the poles. All thanks to global warming, this ice is melting at an amplified rate [16]. In Pakistan, since 1960, there has been a rise in temperature. A trend of this rise is illustrated in fig. 3. With this rise in temperature, there is a rise in the melting of glacial ice which in the past led to catastrophic incidents that formed Attabad Lake and can lead to major floodings in the future. In order to mitigate the threat, Pakistan needs more water bodies to store that excess water. This offers an opportunity to construct waterbodies that can not only store water but also accommodate floating solar. This will create an extra avenue for renewable power generation and revenue while decreasing the loss of water because of evaporation and promoting distributed generation which will be reducing the stress on transmission lines, create employment opportunities and shift the trend of moving to major urban cities for better opportunities.

D. Challenges of FSPV:

In addition to these advantages over land-based solar, FSPV also possesses some added challenges that hinder technology growth. These are the unavailability of water body data, lack of bathymetric and Hydrographic surveys, and lack of guidelines for clearances required for the technology [6]. There are handy guidelines for land-based solar, but when it comes to FSPV, the guidelines are not available, and the ownership of water bodies is ambiguous among public agencies. It is essential to formulate FSPV component-specific standards and technical guidelines for designing such plants FSPV plants on water bodies' surfaces. It is challenging to deal with O&M issues and research and studies required to develop a bathymetric and hydrographic database for the water bodies across Pakistan. However, the installation of FSPV requires less time than land-based solar because, for conventional solar arrangement, the land needs to be prepared, and civil structure like foundations for panels is required [17]. In addition to the challenges of technological advancement, some challenges are related to installation and maintenance. FSPV for its installation needs parts like anchoring and mooring immersed in the water and needs regular inspection to ensure the FSPV plant's stable operation. As there is a high-humidity environment on the water bodies and thus impacts the insulation resistance of the system, it can sometimes drop significantly [18] Divers are needed to attend to these needs, which adds extra expense to the operation and maintenance cost. Also, faults and instances like a bird dropping, replacement of electrical parts, maintenance of cables, and wires are tricky things to handle and require trained plant personnel. Waterbodies hold high importance in the living species' settlements. Water bodies are used for multiple activities, such as fishing, drinking, agriculture, recreation, social events, aquatic life, and research. Further, these water bodies are wealthy in nearby amphibian greenery. The long-term impact of installing large-scale FSPV plants on nearby biodiversity is inadequately known, and there are no research information/studies accessible that can give precise data [21].

E. Databases Of Water Bodies in Pakistan

Potential assessment of floating Solar systems is vital to estimate the market size for FSPV in Pakistan. Promote a market case for FSPV technology, it is only possible if there is adequate opportunity for its possible utilization. This research provides an estimate for the potential assessment of floating soar systems in Pakistan and the province-wise breakup. The light of existing projects suggests that hydropower dams, irrigation dams, industrial ponds, drinking water ponds, barrages, and lakes are the most viable options.

In Pakistan, there is no detailed and credible survey available on water bodies across the country. No Government of Pakistan (GOP) database of waterbodies is available to date. The availability of such a database can result in better research and development studies. To address this issue, this study carried out a potential assessment of Floating solar and it is potential in synergy with pumped hydro storage to address the intermittency of renewables (solar) [19] Global Solar Atlas (GSA) was used to extract GHI Irradiance for each site [1]. The Global Solar Atlas provides a summary of solar power potential and solar resources globally. The World Bank Group provides a free service to carry out research and market study for the potential of solar projects across the world and 2014 report Irrigation system of Pakistan [20] wherein, over 200 hydro dams were listed. An extensive search on Google Earth further perfected this dataset to identify additional potential FSPV sites.

II. METHODOLOGY

In this study, the potential assessment of Floating Solar deployment in Pakistan has been estimated based on data collected from international databanks and a 2014 report titled "Irrigation system of Pakistan" [20] wherein over 200 hydro dams were listed. An extensive search on Google Earth. For the data on water bodies and dams, Wikipedia, the official website of WAPDA [22] data on dams, State of Industry

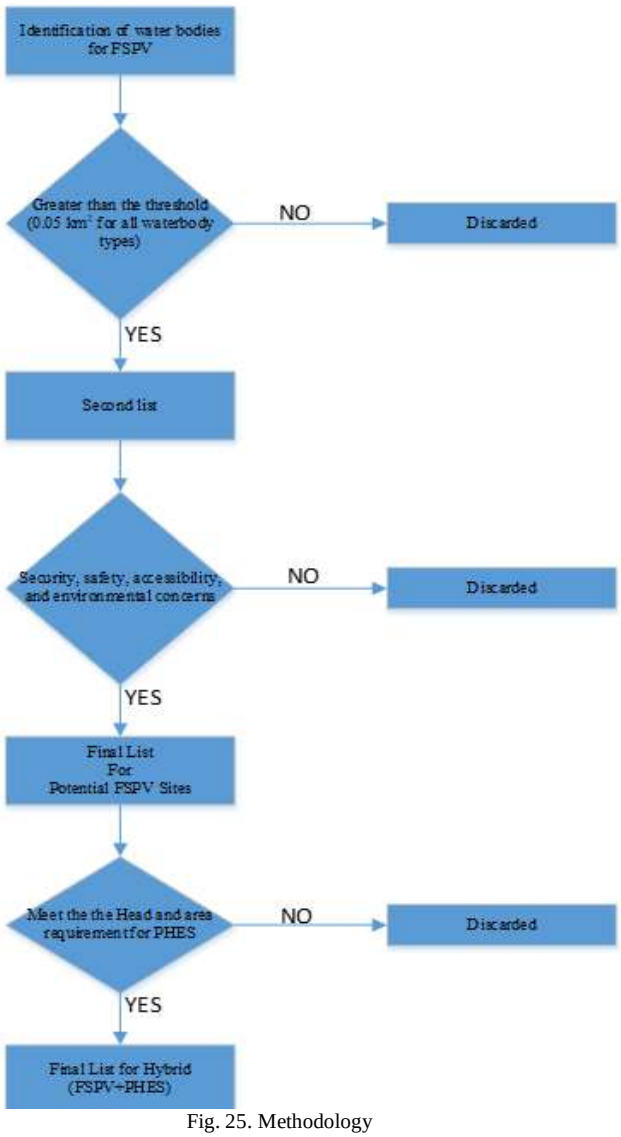


Fig. 25. Methodology

Report [23], Pakistan Energy Year Book [4]. This resulted in the preparation of the First list. For each of the sites in the First list, water ponds were identified on Google Earth, and for those greater than the threshold (0.05 km^2 for all waterbody types), a high-level assessment was carried out to take into consideration the security, safety, accessibility, and environmental aspects of setting up FSPV Systems. Additionally, conflict areas, national parks, flood or earthquake-prone zones, and too remote areas were excluded from further analysis. This activity resulted in the creation of a Second List.

A. Criteria Used for Potential Assessment

To assess the technical feasibility, the criteria are the presence of reservoir storage and capacity (m^2). The capacity is considered for the installation of an FSPV system. For sites, the area recorded is different from the entire surface area of the water body. The second parameter is the maximum depth of the reservoir or water body. The third and fourth technical parameter is the water level variation observed at the water body. The maximum water level is the difference between Maximum and Minimum Depth and the dead level or minimum depth of the water body. This gives the idea about the lowest level the water reaches during the year to accommodate the anchoring and mooring. The fifth parameter is the frequency at which the reservoir is emptied for cleaning or maintenance. The sixth parameter is to check if the water body is located at a dam/lake used for power generation. Whether the water body is used for power generation at present or not. If yes, grid/power off-taker availability is confirmed. The seventh parameter is the surface velocity to assess the possibility of splashes. The eighth parameter is to identify if the bathymetry of the water body is available. A bathymetry study provides detailed data on the water depth along the reservoir and the ground type. This parameter is linked with the mooring and anchoring of FSPV panels. The ninth parameter is the distance of the waterbody from the Nearest Substation/Grid Station. The tenth parameter is the voltage level at the nearest grid station: 500, 220, 132, and 11kV. The eleventh parameter is the maximum load of the substation/Grid Station to assess the grid station's maximum capacity to handle the electrical load. The twelfth parameter is the distance from Transmission Line (in km) to the nearest transmission/distribution line, and the thirteenth is the voltage level of that transmission line closest to the water body.

The fourteenth parameter of the sites was flood risk and assessed the possibility of flooding occurring. Similarly, the fifteenth parameter is the freezing risk, which identifies the possibility of freezing of water bodies around the year. The sixteenth parameter is the installed capacity of the hydropower power generation capacity (in MW) of the power producer, and the seventeenth parameter deals with the type of hydro turbine installed at the sites. For example, Francis Turbine, and Pelton Turbine. The eighteenth parameter dealt with the level of voltage at turbine output (Generation Side) at which electricity is generated.

B. Estimate of FSPV Power Generation Potential

The mapping of Google Earth identified sites, and regularly shaped polygons were drawn on each of the identified locations to estimate their power generation potential. Assuming an average energy density of 1,000W/m^2, every 0.001 km^2 area corresponds to a power generation potential of 1MW [24]. The shapes were drawn, keeping in mind the typical arrangements of FSPV systems, and it was ensured that sufficient areas were left from the edges of the water bodies to account for water level variations. List of sites with areas more significant than the threshold limit (0.05 km^2 for all waterbody types).

$$PFSPV=MW=1000W/m^2 \times \text{Area of water body covered for installation of FSPV } (m^2) \quad (1)$$

For measuring the reduction in evaporation, the average reduction in water evaporation for one square kilometer of water body covered is taken as 1.125 MCM Sq. Km [25] of covered area. Equation (2) is used to estimate the potential of evaporation reduction for each water body.

$$\text{Reduction in evaporation per year} = \text{Area of reservoir covered under FSPV} \times 1.125 \quad (2)$$

For each site in the Second list, several parameters such as flood/freezing risk, shadow masks, and distance to the grid were further assessed to identify non-suitable sites. From the Second List, **seventy-six** sites were **"eliminated"** due to the following reasons:

a) Max potential less than 5 MW (approx. 50,000 m^2 FSPV size): 30 sites
b) Bird sanctuary/designated wetlands: 4 sites
c) Dried out: 13 sites
d) Security (Border area): 01 site
e) No reservoir: 28

The remaining sites were further scrutinized based on the second list analysis based on the security, environmental, accessibility, and shading parameter. These parameters are used for preparing the final list.
Almost all the waterbodies that made it to the final list had GHI levels above 1,500 kWh/m^2 or more, representing good energy yield and better financial attractiveness of the FSPV project.

Fig. 26. Regular Polygon drawn at the location of Gandiali Dam on Google Earth.

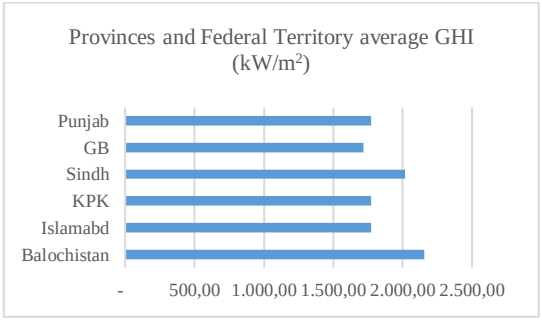


Fig 27. Potential across Provinces, Federal Territory, and Average GHI.

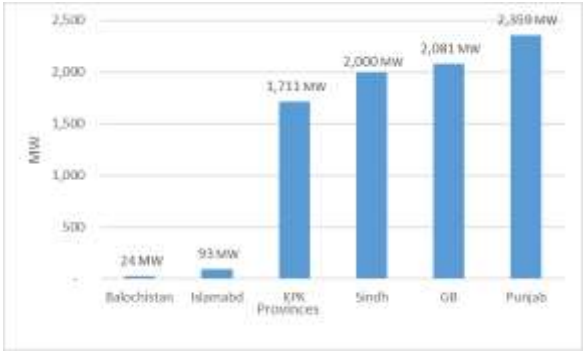


Fig 28. Distribution of Energy Potential of FSPV across Pakistan.

III. CONCLUSION

This research is an effort to develop a baseline for the potential assessment of FSPV across Pakistan. It has been estimated that a gross potential of 17.624 GW_P FSPV installations exists across Pakistan. Further, screening based on technical parameters, environmental, security, and accessibility reveals that there is 2.968 GW_P potential for FSPV in the 1700–2,200 kWh/m^2 range. This illustrates that to get a better return on investment and better financial viability, FSPV projects need to be installed on hydroelectric dams in provinces rich in solar energy.

REFERENCES

[1]. World Bank. (2019). Where Sun Meets Water FLOATING SOLAR MARKET REPORT. Washington DC: World Bank Group

[2]. BBC. (2015, December 13). COP21 climate change summit reaches deal in Paris. Retrieved from BBC: https://www.bbc.com/news/science-environment-35084374

[3]. Koshy, S. M. (2019, July 22). Government spots constraints on renewable energy projects. Retrieved from Down to Earth: https://www.downtoearth.org.in/blog/energy/government-spots-constraints-on-renewable-energy-projects-65755

[4]. Pakistan Energy Yearbook 2018. (n.d.). Pakistan Journal of Hydrocarbon Research.

[5]. GOP. (2019). Alternative Renewable Energy Policy 2019.

[6]. Devraj, M. A. (2019). Floating Solar Photovoltaic (FSPV): A Third Pillar to Solar PV Sector ? The Energy and Resources Institute (TERI).

[7]. Papapetrou, M. (2015, Ocotber 15). IndustRE project - Is industrial demand response complementary or competitive to pumped hydro storage? Birr, Switzerland: IndustRE.

[8]. Chaudhry, Q. (2009). Climate Change Indicators of Pakistan. Technical Report. No. 22. Islamabad: Pakistan Meteorological Department.

[9]. Zafri Azran Abdul Majid, M. R. (2014). Study on Performance of 80 Watt Floating Photovoltaic Panel. JOURNAL OF MECHANICAL ENGINEERING AND SCIENCES.

[10]. Luyao Liua, Q. W. (2016). Power Generation Efficiency and Prospects of Floating Photovoltaic Systems. 8th International

[11]. Shabbir, M. A. (2018). Prospects of floating photovoltaic technology and its implementation in Central and South Asian Countries. International Journal of Environmental Science and Technology.

[12]. Paul Reig, A. M. (2013). World's 36 Most Water-Stressed Countries.

[13]. SPG Solar, I. (2010). Evaporation in Reduction via Floatovoltaics Systems. Retrieved from Bureau of Reclamation: https://www.usbr.gov/lc/region/programs/crbstudy/1_Evaporation_Reduction_via_Floatovoltaics_Systems.pdf

[14]. HuzaifaRauf, M. S. (2019). Integrating Floating Solar PV with Hydroelectric Power Plant: Analysis of Ghazi Barotha Reservoir in Pakistan. Energy Procedia, 816-821.

[15]. Jo-Ellen Parry, D. H. (2016). The Vulnerability of Pakistan's Water Sector to the Impacts of Climate Change: Identification of gaps and recommendations for action. UNDP.

[16]. Craig, T. (2016, August 12). Pakistan has more glaciers than almost anywhere on Earth. But they are at risk. Retrieved from Washington Post: https://www.washingtonpost.com/world/asia_pacific/pakistan-has-more-glaciers-than-almost-anywhere-on-earth-but-they-are-at-risk/2016/08/11/7a6b4cd4-4882-11e6-8dac-0c6e4accc5b1_story.html

[17]. N. Krishnaveni, P. A. (2016). A Survey On Floating Solar Power System. IJCRME, NCFTCCPS.

[18]. Haohui Liu, A. K. (2019). The Dawn of Floating Solar—Technology, Benefits, and Challenges. Lecture Notes in Civil Engineering, vol 41. Springer, Singapore.

[19]. Aseem Kumar Sharma, P. D. (September 2016). Uninterrupted Green Power using Floating Solar PV with Pumped Hydro Energy Storage & Hydroelectric in India. International Journal for Innovative Research in Science & Technology| Volume 3 | Issue 04, 94-99.

[20]. Khalil, H. (2014). Irrigation system of Pakistan. University of Agriculture Faisalabad.

[21]. Tara Hooper, A. A. (2020). Environmental impacts and benefits of marine floating solar. Solar Energy, 1-4.

[22]. WAPDA. (2017, June 16). WAPDA DAMS. Retrieved from WAPDA: http://wapda.gov.pk/index.php/neelum-jhelum/itemlist/category/2-dams

[23]. NEPRA. (2020, June). State of Industry Report. Retrieved from NEPRA: https://nepra.org.pk/publications/State%20of%20Industry%20Reports/State%20of%20Industry%20Report%202020.pdf

[24]. Shahzada Adnan, A. H. (May 2012). Solar energy potential in Pakistan. Journal of Renewable and Sustainable Energy.

[25]. Victoriano Martínez Álvarez, A. B. (2006). Efficiency of shading materials in reducing evaporation from free water surfaces. Agricultural Water Management, 84(3):229-239.

Heterogeneous WSN Modeling: Packet Transmission with Aggregation of Traffic

Canek Portillo
Facultad de Ingeniería
Universidad Autónoma de Sinaloa
Culiacán, México
canekportillo@uas.edu.mx

Jorge Martinez-Bauset
Departamento de Comunicaciones
Universitat Politècnica de València
València, Spain
jmartinez@upv.es

Vicent Pla
Departamento de Comunicaciones
Universitat Politècnica de València
València, Spain
vpla@upv.es

Vicente Casares-Giner
Departamento de Comunicaciones
Universitat Politècnica de València
València, Spain
vcasares@upv.es

***Abstract*—The modeling and the performance analysis of a heterogeneous WSN transmitting in an APT (Aggregated Packet Transmission) mode is presented. With APT is possible to send more than one packet per cycle during the data transmission process. Packets are encapsulated in a unit of information called frame. The study considers the activity and procedures that occur during the data period of the transmission cycle. Results have been obtained and discussed for the following performance parameters: average packet delay, throughput and average power consumption.**

Keywords—aggregated packet transmission, heterogeneous WSN, wireless sensor networks, WSN modeling.

I. INTRODUCTION

One of the ways in which information is usually transmitted in WSN is by transmitting a single packet per cycle (SPT, Single Packet Transmission) [1], [2]; however, Aggregated Packet Transmission (APT) is also possible. Unlike SPT, in APT mode nodes transmit more than one packet (a batch) per cycle. Data aggregation, which is the process of combining multiple data packets into a single data unit called frame, is often used to improve energy efficiency in WSNs. This mechanism can help to reduce the number of transmissions and, consequently, it can help to diminish the consumption of energy [3]. Furthermore, data aggregation also helps to decrease the media access contention as well as the number of packets transmitted and, therefore, it can help to minimize the packet transmission delay [4]. Many data aggregation schemes that contribute to save energy, reduce packet delay and packet collisions have been proposed [3]-[5]. However, there are scarcely any analytical models for evaluating the performance of WSNs with traffic aggregation. There are some proposals related to packet aggregation schemes for WSNs [6]-[8], although these approaches are focused from a routing perspective and without considering any specific MAC layer protocol. Other MAC protocol proposals [9]-[11] integrate data aggregation in WSN, but these studies have been achieved mainly through simulations or based on tests with experimental prototypes. In [12], the authors have developed DTMC models to evaluate the APT scheme for a WSN whose MAC operates with duty-cycled (DC), but the study does not consider heterogeneous scenarios or node classes, nor any prioritization scheme. In [13] and [14], we have carried out a performance analysis of a heterogeneous WSN composed of different classes of nodes, operating with a MAC protocol governed by a synchronized DC, where there is prioritization and where nodes transmit in SPT mode. In the present work, a model with the characteristics mentioned above is developed, but which also expands the capabilities of the nodes to transmit with traffic aggregation.

The rest of the paper is distributed as follows: in section II, the network scenario is presented; in section II, the corresponding modeling of the system is shown; section IV explains how the performance parameters are obtained; section V deals with the numerical results; and finally, the conclusion is in section VI.

II. NETWORK SCENARIO

A. Network operation and assumptions

The network scenario considers the existence of two classes of nodes (N1 and N2) that send packets to a central cluster node called sink (shown in figure 1). This heterogeneous WSN has two classes of nodes. The nodes of class 1 have priority for accessing the channel, while nodes of class 2 can access the channel after nodes of class 1 have vacated the medium. A reference node (RN) is defined for each class. In general, the same assumptions are made as in [13], [14], except that in this model, the nodes can perform the aggregation of packet according to the packets they have in their queues. It is important to note that the packet aggregation capability applies to each class of nodes regardless of its priority. The sum of packets due to the packet aggregation results in a unit of information called frame. For practical reasons, the model defines a maximum frame size, F, in packets. When the RN, of any class, gains the access to the medium, it transmits this frame, and the number of packets in the queue of the RN is reduced according to the number of packets or the size of the frame sent. For example, if q is the number of packets in the queue, Q, of the node, and if $q \leq F$, when there is a successful transmission the queue of RN will be empty; on the contrary, if $q \geq F$, a frame with F packets will be transmitted, leaving $q - F$ packets in queue.

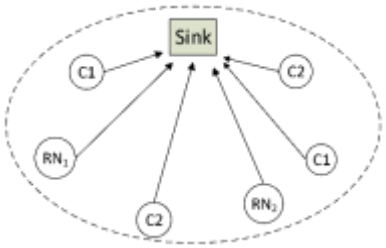


Fig. 1. Wireless sensor network in a heterogeneous scenario formed with different types of nodes.

B. Prioritization of access to the medium

The process considered here, in general, is the same as that described in [13], [14]. The main difference is that the transmissions made by the nodes, whatever their class, includes frames with packet aggregation, instead of a single packet. Nodes belonging to class 1 has priority to the transmission channel. In figure 2, the scheme of the transmission of a frame during the data period of a cycle can be observed. Note that the synchronization period has been omitted for any class of nodes. Also note that, as part of the MAC protocol, the CSMA/CA contention mechanism with the RTS/CTS/FRAME/ACK packet exchange is used. When cycle begins, just nodes of class 1 compete for access to the medium. Nodes of class 2 must wait until the contention window (W_1) of nodes of class 1 have finished. When the nodes of class 2 detect an available medium, because there is not any transmission in progress, the nodes of class 2 will attempt to access to the medium through the activation of the contention mechanism. But, if a node of class 2 detects a busy channel, they will return to a sleep mode to save energy and will wake up once anew in the next cycle. For cycles in which nodes of class 1 collide, nodes of class 2 are considered to detect the activity and will not contend.

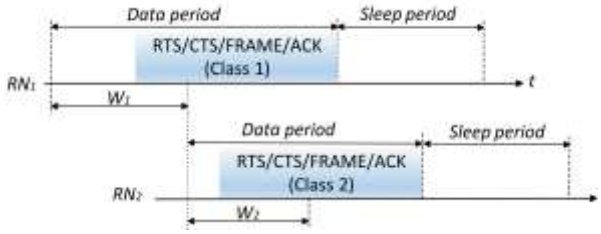


Fig. 2. Frame transmission scheme and the MAC protocol procedure.

III. SYSTEM MODELING

For facility in the explanation, in the following sections and particularly in expressions from (1) to (6), the notation is presented in a generic format, although it could represent both classes of nodes.

A. Access to the medium

Although explained in detail in [13], [14], it is convenient to expose some of the main model assumptions related to the access to the medium. Only the nodes that have at least a packet in its queue are capable of generate a backoff time. The time value is randomly selected from [0, W-1]. A successful transmission of a packet by the RN occurs when the other nodes that contend for the medium select greater backoff time values, compared with that selected by the RN. A collision or a failed transmission will take place when the backoff value obtained by the RN and the same value of at least one of the other nodes are coincident. Besides it must be the smallest value generated in the cycle. There are two possibilities if the resulting backoff time is not the smallest of all: (i) another node transmits with success; (ii) other nodes will collide their packets. All other nodes that could not transmit their packets go to an energy saving mode until the next cycle. Considering a cluster of nodes with packets in its queues, the variable k defines the nodes different from the RN, where: $0 \leq k \leq N-1$. According to the generic notation, N means the nodes of any class. Now, three probabilities can be established: $P_{s,k}$, $P_{sf,k}$ and $P_{f,k}$, which are defined as the probabilities when a packet is successfully transmitted, a packet is transmitted successfully or with collision, and a packet is transmitted with failure, respectively, when the RN and other k nodes contend for the access to the channel.

$$P_{s,k} = \sum_{i=0}^{W-1} {}^{1}/_{W} \left(W - 1 - {}^{i}/_{W}\right)^{k} \quad (1)$$

$$P_{sf,k} = \sum_{i=0}^{W-1} {}^{1}/_{W} \left(W - {}^{i}/_{W}\right)^{k} \quad (2)$$

$$P_{f,k} = P_{sf,k} - P_{s,k} = {}^{1}/_{W} \quad (3)$$

For class 1, these probabilities are calculated considering the reference node RN_1 and the corresponding contention window W_1. For class 2, RN_2 and W_2 are considered.

B. Classes and priorities

For the modeling of each class of nodes we use a two-dimension discrete-time Markov chain (2D-DTMC). These classes are represented by the reference nodes RN_1 and RN_2. Each chain models how the number of packets in the queue of the respective RN evolves over the time, as well as the number of nodes with packets in its queue of each class. The state of each 2D-DTMC is represented by (i,m). The probability of transition from state (i,m) to state (j,n) is represented by: $P_{(i,m)(j,n)}$. Where $i \leq Q$ represents the number of packets in the queue of the RN, and m is the number of nodes that have at least a packet to transmit, besides de RN, and $m \leq K$. For a better explanation and due to space limitations, the transition probabilities of both 2D-DTMC are shown in [15]. A fundamental part of the model is the implementation of the coupling between the two 2D Markov chains. For that reason, in the construction of the expressions for the transition probabilities that are developed for the 2D-DTMC of class 2, the parameter $R_{1,0}$ has been properly defined and incorporated. This parameter refers to the fraction of cycles in which nodes of class 1 have no need to use the channel, and its inclusion is important for the adequate coupling between both Markov chains. Another way to view this parameter is as the probability that there are not active nodes of class 1 in the WSN. At this point, is important to remark that due to the incorporation of the packet aggregation scheme, the expressions of the transition probabilities of each Markov chain are significantly modified. These changes are made through the F and α parameters. The first one was already mentioned in a past section and refers to the maximum number

of packets that can be aggregated in a frame; the second one is the number of packets that have been aggregated to the frame. The new expressions of the transition probabilities that we have obtained are presented in [16].

C. Solution of both coupled Markov chains

The set of linear equations shown in (4) have been used to solve each 2D-DTMC.

$$\boldsymbol{\pi P} = \boldsymbol{\pi},\ \ \boldsymbol{\pi e} = 1 \tag{4}$$

Where $\boldsymbol{\pi} = [\pi(i,n)]$is the stationary probability distribution, $\boldsymbol{P}$ denotes the matrix composed of the transition probabilities, and its different expressions are established in [15] and [16]. The parameter $\boldsymbol{e}$ refers to a column vector of ones. On the other hand, the average probability, P_s, of that the corresponding RN successfully transmits a packet in a random cycle, conditioned on the RN being active, is given by:

$$P_s = \frac{1}{G}\sum_{i=1}^{Q}\sum_{k=0}^{K}\pi(i,k)\cdot P_{s,k} \tag{5}$$

$$G = \sum_{i=1}^{Q}\sum_{k=0}^{K}\pi(i,k) = 1 - \sum_{k=0}^{K}\pi(0,k) \tag{6}$$

From (4) and (5), the stationary probability distribution is a function of P_s, $\boldsymbol{\pi}(P_s)$. There is a dependency relationship between P_s and $\boldsymbol{\pi}$, which enables the resolution of the set of (4), following an iterative fixed-point procedure that allows its solution to be determined, in this case: $\boldsymbol{\pi}$. To solve the second chain, it is necessary to have solved the first one, since that information is needed. However, in the process of coupling the Markov chains for both classes of nodes, the first chain is first solved with the iterative procedure to obtain the stationary distribution of the nodes of class 1 ($\boldsymbol{\pi_1}$). From $\boldsymbol{\pi_1}$, it is obtained the fraction of cycles in which the nodes of class 1 are inactive or the probability that the nodes of class 1 are inactive: $R_{1,0} = \pi_1(0,0)$. This parameter $R_{1,0}$ is fundamental in the formation of the transition probability matrix $\boldsymbol{P_2}$ of class 2, therefore, it can be established that the stationary distribution of the nodes of class 2 is also a function of $R_{1,0}$, that is: $\boldsymbol{\pi_2}(P_{s2}, R_{1,0})$. Finally, $\boldsymbol{\pi_2}$ is obtained with the mentioned iterative procedure. In the same way, $R_{1,0}$ must be considered for the determination of the performance parameters of nodes of class 2. This parameter allows the model to indicate that during the transmission of nodes of class 2, there are no active nodes of class 1 trying to transmit. Therefore, it is important for the correct operation of the protocol, especially in relation to the inclusion of the priorities of access to the medium.

IV. PERFORMANCE PARAMETERS

A. Throughput

For the determination of throughput, a conceptually significant incorporation is made; the accumulated number of packets by traffic aggregation and their transmission in a single frame is considered in the calculation. The throughput per node, η, is defined as the average number of packets that a node has successfully delivered in a cycle. The total throughput or system throughput, whose measurement unit is packets per cycle, is the addition of all individual throughputs due to each node, whatever the class it belongs. For class 1 nodes, the throughput per node is obtained with (7), while the total system throughput is determined with (8).

$$\eta_1 = \sum_{i=1}^{Q_1}\sum_{k=0}^{M_1}\alpha_1\pi_1(i_1,k_1)P_{s1,k1}, \tag{7}$$

$$Th_1 = N_1\eta_1 \tag{8}$$

$$\alpha_1 = \min(i_1, F_1) \tag{9}$$

Where α_1 represents the aggregated packets, i_1 refers to the packets in the queue of RN_1, and F_1 is the maximum number of packets that can be added according to the configuration set. For nodes of class 2, the throughput per node and the total throughput of the system are given by (10) and (11), respectively.

$$\eta_2 = \sum_{i=1}^{Q_2}\sum_{k=0}^{M_2}\alpha_2\pi_2(i_2,k_2)P_{s2,k2}\cdot R_{1,0}, \tag{10}$$

$$Th_2 = N_2\eta_2 \tag{11}$$

$$\alpha_2 = \min(i_2, F_2) \tag{12}$$

Where α_2 represents the aggregated packets, i_2 refers to the packets in the queue of RN_2, and F_2 is the maximum number of packets that can be added according to the configuration set. Note that for the calculation of the throughput for class 2, it is necessary to consider the inactivity of the nodes of class 1, through the parameter $R_{1,0}$, which is the stationary probability distribution of not finding active nodes of the class 1 (fraction of cycles where nodes of class 1 are idle).

B. Average paket delay

D is defined as the average delay experienced by a packet from its arrival at the queue of the node until it is successfully transmitted, and it is measured in cycles. For the determination of D, Little's law is applied. For class 1 nodes, the delay is calculated with the following expressions:

$$D_1 = N_{av1}/\gamma_{a1}\ ,\quad N_{av1} = \sum_{i=0}^{Q_1} i\pi_{i1}, \tag{13}$$

$$\gamma_{a1} = \eta_1,\quad \pi_{i1} = \sum_{k=0}^{M_1}\pi_1(i_1,k_1) \tag{14}$$

Where π_{i1} is the class 1 stationary probability of finding i_1 packets in the queue of the corresponding reference node of class 1, RN_1. $N_{av1}$is the average number of packets in queue of RN_1, and $\gamma_{a1}$is the average number of packets accepted by the queue of RN_1, which is equal to η_1. For class 2 nodes, the delay is calculated with the following expressions:

$$D_2 = N_{av2}/\gamma_{a2}\ ,\quad N_{av2} = \sum_{i=0}^{Q_2} i\pi_{i2}, \tag{15}$$

$$\gamma_{a2} = \eta_2,\quad \pi_{i2} = \sum_{k=0}^{M_2}\pi_2(i_2,k_2) \tag{16}$$

Note that the previous terms can be defined in a similar way as for those of class 1, only that class 2 must be considered in all parameters.

C. Average power consumption

For the determination of the average energy consumption, the accumulated number of packets by traffic aggregation and their transmission in a single frame is considered, as well. This consideration is a conceptually significant incorporation. The energy is calculated during the data period, and just the energy consumption due to the transmitter and receiver is considered in the study. The average energy that the RN consumes in a cycle during the data period can be determined by the following expression:

$$E_d = E_s^{tx} + E_f^{tx} + E^{oh} \quad (17)$$

Where, E_s^{tx}, E_f^{tx} and E^{oh} represent the terms of the energy consumed when the RN transmits with success, with failure and when it listens to the transmission of other nodes (overhearing), respectively. The E_s^{tx} consumption value is obtained with the following expressions:

$$E_s^{tx} = \sum_{i=1}^{Q} \sum_{k=0}^{M} \pi(i,k) P_{s,k} \left(P_{s,1}^{tx} + \alpha P_{s,2}^{tx} + P_{s,1}^{rx} + P_{s,2}^{rx}\right) \quad (18)$$

$$P_{s,1}^{tx} = t_{RTS} P_{tx} \; P_{s,2}^{tx} = t_{DATA} P_{tx} \quad (19)$$

$$P_{s,1}^{rx} = \left[t_{CTS} + t_{ACK} + 4D_p\right] P_{rx} P_{s,2}^{rx} = BT_{s,k} P_{rx} \quad (20)$$

Where t_{RTS}, t_{DATA}, t_{CTS} and t_{ACK} are the transmission times for the control packets used during the transmission process. P_{tx} and P_{rx} are the transmission and reception power levels, D_p is the one-way propagation delay, and $\alpha = \min(i, F)$ is the number of aggregated packets. $BT_{s,k}$ is the average backoff conditioned to a successful transmission of packets from the RN, when competing with k other nodes [14]. The factor α determines the number of packets that are added to the frame that is transmitted, in such a way that the greater the number of packets added, the greater the energy consumption when they are successfully transmitted. To determine E_f^{tx} and E^{oh}, the same procedure is carried out as that developed in [14]. To determine the average energy consumption per cycle for nodes of class 1, E_1, and for nodes of class 2, E_2, the following is expressions are used:

$$E_1 = E_{d1}, \quad (21)$$

$$E_2 = \left(1 - R_{1,0}\right) E_0 + R_{1,0} . E_{d2} \quad (22)$$

Where E_{d1} is the energy consumption during the data period for the nodes of class 1, and E_{d2} is energy that is consumed during the data period due to the nodes of class 2. $R_{1,0}$ refers to the stationary probability distribution of not finding active nodes of class 1. E_0 is the energy consumed by nodes of class 2 to wake up and detect if that medium is occupied.

V. NUMERICAL RESULTS

A. Parameter configuration and scenarios

From the developed models that are explained in section III, we have obtained analytical results which have been validated by simulation. To obtain simulation results, a discrete event simulator has been developed in C language, which simulates the WSN according to the network scenario explained in section II. It should be noted that the simulator previously developed for other related studies has been modified so that it can transmit with traffic aggregation, considering different possibilities of maximum queue size for any of the classes. It is important to note that the results that have been obtained analytically with the model, are totally independent of the results obtained with the simulator. In the following sections, the performance parameters results are presented. In the different figures, the simulation results are represented with markers only, while the results obtained analytically are represented with lines and markers. The analytical and simulation results perfectly match, confirming that the analytical model is highly accurate. We have obtained confidence intervals with a confidence level of 95%, however, as they are very small, they have been omitted from the figures for clarity. The parameter configuration is summarized in Table I.

TABLE V. PARAMETER CONFIGURATION

Parameter	value	Parameter	Value
Cycle duration (T)	60 ms	Propagation delay (Dp)	0.1 us
t_{SYNC}, t_{RTS}, t_{CTS} and t_{ACK}	0.18 ms	Slot time (ts)	0.1 ms
t_{DATA}	1.716 ms	Contention window (W)	128 slots
Data packet size (S)	50 bytes	Queue size (Q)	5 packets
Transmission power (Ptx)	52 mW	Reception power (Prx)	59 mW
Node number and scenarios	N1 = 5 (SC1 and SC2)	Number of packets per frame (F)	F1=F2= {2,5,10}
	N2=4N1=20 (SC2)	Packet arrival rate (packets/s)	λ_1 = {0.5, 1.0}
			λ_2 = [0.5, 4.5]

B. Average packet delay

In Fig. 3, the average packet delay is shown, which is measured in cycles. The scenario considers both classes of nodes, both transmission schemes (SPT, APT) and a packet arrival rate λ_1=0.5. D_1 and D_2 refer to the average delay of packets that each class has successfully transmitted, respectively.

As expected, class 1, being the priority class, experiences very low delay for both schemes (SPT, APT) and for the different sizes of F used in APT (F= 2, 5, 10). Consider that nodes of class 1 work with low load and have their queues empty most of the time. Therefore, when a packet arrives at its queues, it is transmitted almost immediately and with a very low probability of collision. It is also clear that for nodes of class 2 (the non-priority class), the impact of increased traffic and collisions is significant. Note that D_2 increases with λ_2, since the

fraction of colliding packets increases with λ_2, and more retransmissions are required to successfully transmit their packets. It is also observed that for APT scheme, lower values of D_2 are reached, when the value of F increases. This effect is very significant for the values of F= {5, 10}. The queue of the node empties faster when multiple packets are transmitted together, reducing contention for media access and, in consequence, also reducing the packet collisions.

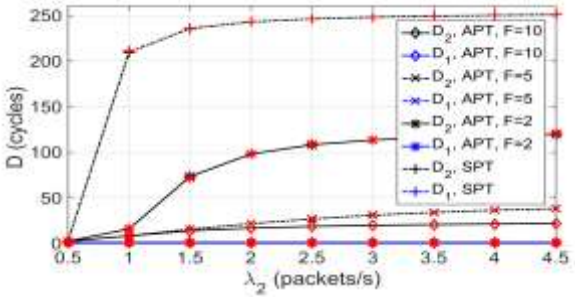


Fig. 3. Average packet delay for both classes and both transmission schemes.

C. Throughput

Figure 4 shows the throughput per node for both classes of nodes (class 1 and class 2). It also shows how class 2 (non-priority) benefits from the use of the APT scheme, obtaining higher throughput values.

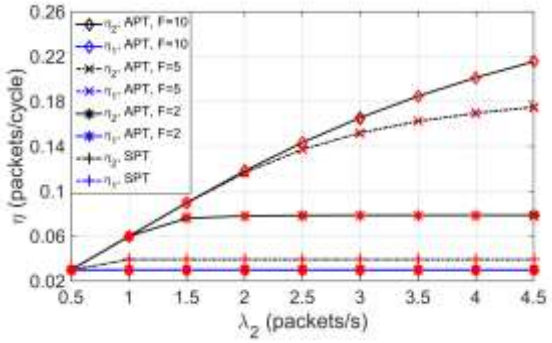


Fig. 4. Throughput per node for both classes and for both transmission schemes.

While in SPT scheme, class 2 reaches a maximum throughput limit (saturation) at λ_2 =1, when APT is used, these saturation limit values increase with F. Thus, for F=2 the throughput is doubled, and for F=5 and F=10 they do not reach any saturation point for the considered scenario. When there is saturation, all nodes have packets in their queues ready to be sent in almost every cycle. The throughput increases, not only because more packets per cycle are transmitted in APT, but also because the probability of a node successfully transmitting a packet also increases. In APT, the queue empties faster, therefore, the number of contending nodes per cycle decreases.

D. Average power consumption

In figure 5, the average energy consumption per cycle is shown. The scenario considers both classes of nodes as well as the two transmission schemes. The measurement unit used is the millijoule (mJ). The figure 5 also shows that for nodes of class 1, the energy consumed remains constant as λ_2 increases. This is due to the packet arrival rate λ_1 and the number of nodes N_1 are both constant values. For class 2 nodes, the packet arrival rate varies according to the values shown in table I, where the parameter configuration has been set: ($\lambda_2 \in [0,4.5]$ packets/s). However, the nodes eventually reach an activity limit, which has an associated power consumption limit. In addition, when the correspondences of figure 3 and figure 5 are analyzed, some relationships between the energy consumed and throughput can be inferred. For example, higher throughput values imply more transmissions, and, in consequence, more packet deliveries. The above, in terms of energy, also implies a greater activity by the nodes, and therefore, a grater energy consumption.

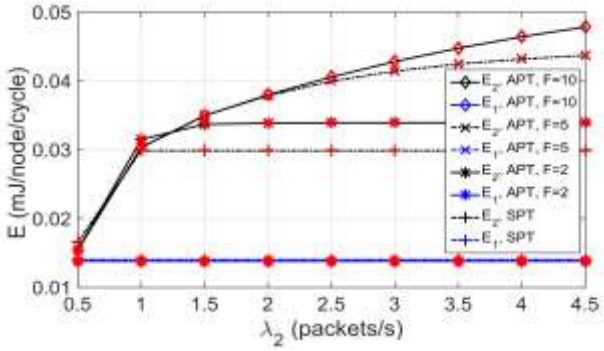


Fig. 5. Average energy consumption for both classes of nodes and both transmission schemes.

However, an important part of this power consumption is due to the node frequently incurring in overhearing (listening to other nodes). This occurs when the node loses the contention for accessing to the channel, but it had to listen to the channel during the backoff period to notice. The node knows if the channel is busy when it detects activity on the channel. With APT there is a higher power consumption per node per cycle compared to SPT. As the value of F increases, the power consumption reaches higher levels.

VI. CONCLUSION

We have carried out a study of the performance of a wireless sensor network composed of different types of nodes. For nodes, it also considers the assignment of priorities to access to the medium. Moreover, the single packet transmission (SPT) and the aggregated packet transmission (APT) schemes are included in the study, although the analysis is focused on APT. To achieve the previous mentioned, an analytical model has been developed for a MAC protocol of a WSN that operates with a synchronized duty cycled and that considers the heterogeneity of the nodes that make up the WSN. Furthermore, the access priorities, and the SPT and APT operation schemes are also considered in the model. Moreover, the analytical model has

been proven for different scenarios, obtaining results for the following performance parameters: throughput, average packet delay and average energy consumption. The validation of the analytical model has been done through discrete events simulations, which show accurate results. The analysis shows how, both types or classes of nodes, can be impacted when APT scheme is used, and especially how class 2, the non-priority class benefits from the APT transmission scheme. The above is particularly true when the nodes increase its traffic, allowing them to achieve a better performance than with SPT scheme.

ACKNOWLEDGMENTS

This work was supported in part by Grant PGC2018-094151-B-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe, in part by Grant 2014-0870/001-001 (EuroinkaNet) and in part by Grant DSA/103.5/15/6629 (SEP-SES).

REFERENCES

[1] W. Ye, Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communictions Societies, vol. 3, 2002, pp. 1567-1576.

[2] W. Ye, J. Heidenmann, and D. Estrin, "Medium access control with coordinated daptative sleeping for wirless sensor networks," in IEEE/ACM Transactions on networking, vol. 12, no. 3, pp. 493-506, June 2004.

[3] K. Akkaya, M. Demirbas, and R.S Aygun, "The impact of data aggregation on the performace of wireless sensor networks," in Wireless Comunications and Mobile Computing, vol. 8, no.2, pp. 171-193, 2008.

[4] M. Bagaa, Y. Challal, A. Ksentini, A. Derhab, and N. Badache,"Data aggregation scheduling algorithms in wireless sensor networks: sonlutions and challenges," in IEEE Communications Surveys Tutorials, vol. 16, no. 3, pp. 1339-1368, Third Quarter 2014.

[5] R. Rajagopalan and P.K. Varshney, "Data-aggregation techniques in sensor networks: a survey," in IEEE Comnunications Surveys Tutorials, vol. 8, no. 4, pp. 48-63, Fourth Quarter 2006.

[6] L. Galluccio and S. Palazzo, "End-to-end delay and network liftime analysis in a wireless sensor network performing data aggregation," GLOBECOME 2009 - 2009 IEEE Global Telecommunications Conference, 2009, pp. 1-6.

[7] M. Kamarei, M. Hajimohammaadi, A. Patooghy, and M. Fazeli, "OLDA: an efficient on-line data aggregation method for wireless sensor networks," in 2013 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications, 2013, pp. 49-53.

[8] B. Alinia, M.H. Hajiesmaili, A. Khonsari, and N. Crespi, "Maximum-Quality Tree Construction for Deadline-Constrained Aggregation in WSNs," in IEEE Sensor Journal, vol.17, no. 12, pp. 3930-3943, June 2017.

[9] Z. Li. Y. Peng, D. Qiao, and W. Zhang, "Joint aggregation and MAC design to prolong sensor network lifetime," IEEE International Conference on Network Protocols (ICNP), 2013, pp. 1-10.

[10] Z. Li, Y. Peng, D. Qiao, and W. Zhang, "LBA: Lifetime balanced data aggregation in low duty cycle sensor networks". In 2012 Proceedings IEEE INFOCOM, pages 1844-1852, March 2012.

[11] K. Nguyen, U. Meis, and Y. Ji, "An energy efficient, high throughput MAC protocol using packet aggregation," 2011 IEEE GLOBECOM Workshops (GCWkshps), 2011, pp. 1236-1240.

[12] L. Guntupalli, J. Martinez-Bauset, F. Y. Li, and M.A. Weitnauer, "Aggregated packet transmission in duty-cycled WSNs: modeling and performance evaluation," In IEEE Trans. on Vehicular Technology, vol. 66, no. 1, pp. 563-579, Jan 2017.

[13] C. Portillo, J. Martinez-Bauset, V. Pla, and V. Casares-Giner, "Modelling of S-MAC for Heterogeneous WSN," 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2018, pp.1-6.

[14] C. Portillo, J. Martinez-Bauset, V. Pla, and V. Casares-Giner, "Modeling of duty-cycled MAC protocols for heterogeneous WSN with priorities," in Electronics, vol. 9, no. 3, pp. 1-16, March 2020.

[15] C. Portillo, J. Martinez-Bauset, V. Pla , and V. Casares-Giner, "The state transition probabilities of the two 2D-DTMC," Technical note. [Online], Available:http://personales.upv.es/jmartine/public/2DDTMC.pdf

[16] C. Portillo, J. Martinez-Bauset, V. Pla , and V. Casares-Giner, "The State Transition Probabilities of the Two 2D-DTMC with traffic aggregation," Technical Note: [Online], Available: https://drive.google.com/drive/folders/1ezmxQxjrt410T_GglqUo6YWKgOEJKd1M?usp=sharing

Factors Impacting Ethical Behavior in South African Software Development Organizations

Robert Hans
Computer Science Department
Tshwane University of Technology (TUT)
Pretoria, South Africa
hansr@tut.ac.za

Senyeki Marebane
Faculty of Information and Communication Technology
TUT
eMalahleni, South Africa
MarebaneSM@tut.ac.za

Jacqui Coosner
Operations Department
Incus Data
Pretoria, South Africa
jjc@incusdata.com

Livhuani Nedzingahe
Research & Inovation Directorate
TUT
Pretoria, South Africa
NedzingaheL@tut.ac.za

***Abstract*— Some recent unsavory unethical behavioral conduct from some software practitioners has thrusted the software industry into a spotlight. This has resulted, amongst other things, in some researchers in the field attempting to understand existing ethical climates under which software practitioners operate. These ethical climates consist of factors which impact and dictate the ethical behavior of the practitioners. This study sought to establish which factors influence positive ethical behavior and which ones influence negative conduct from software practitioners' perspectives. Corporate processes, cultures and external factors were found to be positively influencing ethical behavior. On the other hand, people-related factors, including lack of knowledge about ethics, people not believing in the need for codes of ethics for software engineering, work pressures and unrealistic expectations from stakeholders were reported by the respondents of this study as having a negative impact on their ethical behavior. Software development organizations are urged to pay requisite attention to the factors that were found to impact ethical behavior negatively by the software practitioners. The number of respondents for the study somehow requires that any generalization of its findings be accepted with caution.**

Keywords—factors, impact, ethics, behavior, software, practitioners, South Africa

I. INTRODUCTION

The incidents of unethical behavior by some practitioners in software development have once more prompted a discussion on ethics related to the software development profession. The difficulty with the consequences of such unethical behaviors is that, apart from other losses, they negatively impact human lives. Some of the unbecoming unethical behavior include the emission scandal by Volkswagen [1], Uber's Greyball software used to evade law enforcement officers [2] and South Africa's Experian data breach of customer records [3]. Such unethical conduct not only tarnish the image of the software industry [4], but result in serious losses to business [5], and also loss of life, as was the case in Boeing [6] and THERAC 25's system failure [7].

Several studies, including [5], [8] have identified factors influencing ethical or unethical behaviour in the information technology field. Cronan, Leornard and Kreie [5] identified attitude (judgement whether the behavior is good or bad) and the subjective norm (perception of how one should behave) as some of the important factors that affect one's ethical behavioral intention. Ferrell and Gresham [9] mentioned four factors influencing the likelihood of ethical behavior, and these are: individual characteristics, significant others (people who wield influence), opportunity (e.g. code of ethics, organizational policies, etc.) and the ethical dilemma itself. Organizational culture has also been identified as one of the factors influencing ethical or unethical behavior [6]. In contrast, family influences, life experiences, education, religion, personal values and peer influences play a vital role in one's ethical conduct [10]. Even though several studies have been conducted on the factors influencing ethical behavior, Haines and Leonard [11] indicate that judgements and moral decision-making of individuals evolve with time and interaction with others; hence we find it important to conduct this study.

As part of a project to determine the ethical climate in the South African development environments, we collected data from software practitioners to establish factors they consider to be impacting ethical behavior in their organizational environments. This study reports on the data analysis of the collected data. To the authors' knowledge, there is no South African study that has been conducted in line with this study.

II. LITERATURE REVIEW

Reynolds [10] defines ethical behaviour as conforming "*to generally accepted social norms, many of which are universal*." In the context of software development, the ethical behavior of software practitioners should be aimed at serving the public, clients and users while acting to meet the employers' needs in line with the interest of the public [12].

According to Haines and Leonard [11] research on ethical decision-making focused on the examination of two elements: personal characteristics and the process of ethical decision-making. This is why Trevino's [8] four-component model and Rest's [13] and Ajzen's [14] theory on planned behavior models

of ethical decision-making focus on personal characteristics. Besides the above three foundational models, which are the successor to Kohlberg's [15] moral development model, there are several existing information and communication technology (ICT) ethical behavioral models, which are aimed at assisting software practitioners in making ethical decisions. These models include psycho-social model developed by Eining and Christensen, cited in Cronan, Leornard and Kreie [5], Banerjee, Cronan and Jones's [16] information technology (IT) ethical behavior intention model and an IT ethical behaviour model by Cronan, Leornard and Kreie [5]. However, Haines and Leonard [11] believe that consideration of external factors that influence individual beliefs and judgements is imperative besides personal characteristics and the ethical decision-making process. Organizational environments also contribute to ethical decision-making, which is why Bommer et al. [17] in their study on behavioral model, emphasizes the role played by organizational environment (specifically the professional and work environments) as having a strong influence on ethical decision-making.

An organizational ethical climate refers to a set of shared rules and policies on ethical values and standards that determines and influences the ethical conduct of members of the organization [18], [19]. An organization's ethical climate plays a critical role in promoting and fostering ethical behavior expected from employees [20] and decision-making in response to ethical dilemmas [21]. Baker, Hunt and Andrews [22] are in agreement with this assertion by indicating that an employee's ethical behavior is normally in line with its set ethical procedures, values and standards.

Management or leadership has a critical responsibility in establishing an organizational climate. According to Sims [23], managers are responsible for the establishment of a culture that supports the learning of personal values that promote and nurture ethical behavior in organisations. This may include managers (1) devising safe mechanisms for reporting unethical behavior and (2) showing commitment to ethics by disciplining those who conduct themselves contrary to the adopted codes of ethics, and (3) rewarding ethical conduct. Vitell and Hildago [24] showed that positive consideration of ethics by employees is influenced by organisational ethical culture, including managers' ability to enforce ethics. According to Dimitriou and Ducette [25] several studies show the critical role played by managers in influencing (positively or negatively) ethical/unethical behaviour of their subordinates in organizations. Deshpande, Joseph and Prasad [26] concur with this view and state that actions and conduct by managers set behavioral norms for their workforce. Therefore, ethical leadership sets a powerful tone and standards for subordinates' ethical behavior because such leaders influence their followers via modelling – as "targets of identification and emulation" [27].

As alluded above, several studies were conducted on the factors influencing ethical decisions in ICT-related environments. For example, Leonard and Cronan [28] examined environmental factors, moral obligation, and awareness of consequences that influence an individual's attitude toward ethical behaviour. The study found that environmental factors influence ethical behaviours, although there were varying results in terms of influences yielded by each environmental factor. It highlights that moral obligations and consequences emerged as high influencers of an individual's attitude consistently. On the male respondents it showed strong influence compared to their female counterparts, who appeared to be more amenable to being influenced by their peers.

A study by Charlesworth and Sewry [21] examined factors on the behaviour of employees in relation to principles in the codes of ethics in order to determine the ethical awareness of computing professionals. The study revealed that whilst professionals did not receive education on professional ethics of computing, the majority of the respondents showed awareness of ethics. However, their employer organisations were not aware of ethics. Furthermore, the *"abuse of confidential information'* and *"the tendency to produce poor quality work'* appeared as popular ways of violating ethics codes. This clearly demonstrates how the lack of an ethically tuned organisational climate can enable and sustain unethical behaviours of computing professionals.

Software development environments are naturally project-oriented. Therefore, there exist limits in terms of time and competing goals on the project, which subsequently exposes developers to unrealistic expectations from users and managers alike. In such environments, potential risks of not paying attention to ethics may arise, especially if there is no ethical leadership and culture that promotes an organizational ethical climate. Martin and Cullen [29] indicate that apart from individual factors or personal characteristics of professionals, the constitution of the environments they work has an influence on the ethical behavior of practitioners. Thus, an analysis of these factors requires research attention.

It is on these bases and the need for ethical decision-making in software development that we conduct this study to determine factors that software development practitioners consider as having an influence on ethical behaviors in their work environments. Even though several factors were identified in the reviewed studies, the majority of such research works were conducted outside South Africa. As a result, research gaps exist; hence this research wants to contribute to closing the gap.

III. RESEARCH METHODOLOGY

To address the aim and objectives of this study, a descriptive research design was applied in a quantitative approach. The data was collected from corporate participants within a software development environment in South Africa. The aim was to understand factors that impact on ethical behaviour of software practitioners in software development. A survey questionnaire was used to collect the data from an online platform. The link to the survey was shared with the participants through various methods, including emails and social media platforms, to participate in the study over a period of 12 months. The collected data formed part of a research project aimed at establishing ethical software engineering climate in South African software development environments. Hundred and three (103) participants responded to the call for participation, with all their responses usable for data analysis. The 103 were selected from an unknown population of practitioners, and a precision-based method was applied to calculate a representative sample from an

unknown population using a 10% margin of error and 95% confidence interval. Data was analysed using descriptive statistics (Frequencies and Percentage) to summarise participants' responses, Mean and Standard deviation to assess the average responses between participants that agree somewhat, agree strongly; disagree somewhat and disagree strongly. The one-way analysis of variance (ANOVA) was applied to assess if there are any significant differences between participants that agreed or disagreed with the statements regarding factors they considered to impact ethical behavior in their software development environments. A p-value of 0,05 was used to test the level of significance. The study analyzed the collected data using SPSS version 28.

IV. RESEARCH RESULTS

This section presents the results of the data analysis collected survey to establish factors considered facilitators or inhibitors of ethical behavior by software practitioners in the South African software industry.

A. Profile details of the participants

Table 1 provides a summary of the participants' profile details. The majority (84,31%) of the participants were male, while female participants formed 11,77% of the respondents. The remaining 3,92% of the respondents chose not to disclose their gender.

The age group of the participants was as follows: 18 – 29 years: 28,43%; 30 – 39 years: 32,35%; 40 – 49 years: 24,51%; 50 – 59 years: 12,75%; and 60 years and above constituted 1,96%.

The results in Table 1 show that 3,92% of the participants did not have matric qualifications, while 9,8% of them had it. Just over 31% possessed diplomas, whereas 28,43% of the respondents were degree holders. One (0,98%) participant had partially completed a master's degree. The remainder of the participants had post-graduate qualifications (24,51%) and 'other' (0,98%) qualifications. The majority (53,92%) of the respondents had more than 10 years' experience in the software development industry, while 20,59% and 25,49% of the participants had 6 to 10 and 0 to 5 years of experience in the software development industry, respectively.

The next subsection presents the data analysis results on factors which participants considered to be having an impact on their ethical behavior.

TABLE I. Profile details of the participants.

Profile variables	Description	Frequency	Percentage
Sex	Female	12	11,77%
	Male	86	84,31%
	Prefer not to say	4	3,92%
	Total	**102**	**100,0%**
Age group	18 – 29	29	28,43%
	30 – 39	33	32,35%
	40 – 49	25	24,51%
	50 – 59	13	12,75%
	60 and above	2	1,96%
	Total	**102**	**100,0%**
Highest level of formal education	Did not complete matric	4	3,92%
	Matric	10	9,8%
	Diploma	32	31,37%
	Degree	29	28,43%
	Partial masters degree	1	0,98%
	Post-graduate qualification	25	24,51%
	Other	1	0,98%
	Total	**102**	**100,0%**
Experience in software development industry	0 - 5 years	26	25,49%
	6 - 10 years	21	20,59%
	More than 10 years	55	53,92%
	Total	**102**	**100,0%**

B. Factors considered to impact ethical behaviour by South African software practitioners.

People don't know about ethics in software engineering: Table 2 shows that 31,07% of the participants strongly believed that people (i.e. other software project stakeholders, such as management, users, clients, etc.) don't know ethics in the software engineering field. At the same time, 43,69% of the respondents agreed somewhat with the statement, while 16,5% and 8,74% of the participants disagreed somewhat and disagreed strongly with the statement's sentiments, respectively.

People don't believe that we need a code of ethics for software engineering: With regard to this statement, 22,33% of the participants agreed strongly with the statement, whereas 33,98% of them agreed somewhat with it. However, 29,13% of the participants disagreed somewhat with the statement, while 14,56% disagreed strongly with it.

Our processes don't encourage stronger ethical behavior: According to Table 2, 22,33% of the participants agreed strongly with the statement, whereas 28,16% of the respondents agreed somewhat with the sentiments expressed by the statement. On the other hand, 14,56% of the participants disagreed strongly with the statement, whereas 29,13% disagreed somewhat with it.

There is too much work pressure due to lack of time, resources or skills: More than 24% of the respondents agreed strongly with the statement, while 42,72% agreed somewhat with it. However, 23,3% of the participants disagreed

Table II. Response distribution of participants

Factors they considered to impact on ethical behaviour in their software development environments	Agree somewhat		Agree strongly		Disagree somewhat		Disagree strongly		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
People don't know about ethics in software engineering.	45	43,69	32	31,07	17	16,50	9	8,74	103	100
People don't believe that we need a code of ethics for software engineering.	35	33,98	23	22,33	30	29,13	15	14,56	103	100
Our processes don't encourage stronger ethical behavior.	29	28,16	13	12,62	39	37,86	22	21,36	103	100
There is too much work pressure due to lack of time, resources or skills.	44	42,72	25	24,27	24	23,30	10	9,71	103	100
Management and/or users have unrealistic expectations.	41	39,81	29	28,16	21	20,39	12	11,65	103	100
We don't have control over external software developers.	30	29,13	21	20,39	34	33,01	18	17,48	103	100
Our corporate culture doesn't encourage good ethical behavior.	17	16,50	9	8,74	27	26,21	50	48,54	103	100
External factors, such as politics and the economy, affect our behavior.	30	29,13	17	16,50	25	24,27	31	30,10	103	100
Average	**33,88**	**32,89**	**21,13**	**20,51**	**27,13**	**26,33**	**20,88**	**20,27**	**103,00**	**100**

somewhat with the statement, while 9,71% disagreed strongly with it.

***Management and/or users have unrealistic expectations**:* The results in Table 2 indicate that 28,16% of the participants stated that they strongly believed (agreed) that management and/or users have unrealistic expectations. At the same time, 39,81% of them agreed somewhat with the statement. Conversely, 20,39% of the respondents disagreed strongly with the statement, and the remaining 11,65% of them disagreed somewhat with it.

***We don't have control over external software developers**:* Table 2 shows that 20,39% of the respondents agreed strongly that they had no control over external software developers, while 29,13% of them agreed somewhat with the statement. However, 33,01% and 17,48% of the participants disagreed strongly and disagreed somewhat with the statement, respectively.

***Our corporate culture doesn't encourage good ethical behavior**:* Table 2 shows that 8,74% of the respondents agreed strongly that their corporate culture did not encourage good ethical behavior. Sixteen and a half (16,5%) of the respondents agreed somewhat with the statement. Conversely, 26,21% of the respondents disagreed strongly with the statement, while the remaining 48,54% of them disagreed somewhat with it.

People don't believe that we need a code of ethics for software engineering: Regarding the responses to this statement, 16,5% of the participants agreed strongly with it, while 29,13% of them agreed somewhat with it. On the other hand, 24,27% and 30,1% of the participants disagreed somewhat and disagreed strongly with it, respectively.

Results in Table 3 show that, on average, 32,89% of participants agreed somewhat, while 20,51% agreed strongly that these factors could be considered to impact ethical behavior in their software development environments. This is compared to an average of 26,33% that disagreed somewhat and 20,27% that disagreed strongly. Thus, results between agreed somewhat, agree strongly, disagree somewhat and disagree strongly significantly differed across participants' responses between factors considered to impact ethical behavior in their software development environments (with $F= 3,09>2,94$; p-value = $0.043<0,05$). These results imply that within participants' responses, the majority of participants considered the following to be factors impacting negatively on ethical behavior in their software development environments: people not knowing about ethics in software engineering (74,75%); management and/or users have unrealistic expectations (67,96%); having too much work pressure due to lack of time, resources or skills (66,99%); and people don't believe that we need a code of ethics for software engineering (56,31%). In contrast majority of participants indicated that the following cannot be considered factors impacting negatively on ethical behavior in their software development environments: not having control over external software developers (50,48%); external factors, such as politics and the economy, affect our behavior (54,36%); processes that do not encourage stronger ethical behavior (59,22%); and corporate culture doesn't encourage good ethical behavior (74,75%). These results differed significantly between and within groups (Mean Square errors between groups = 284,1926 and Mean Square error within group = 91,8862; p-value = $0,043<0,05$). Thus, this study shows that factors considered to impact ethical behavior in their software development environments are: people not knowing about ethics in software engineering; management and/or users have unrealistic expectations; too much work pressure due to lack of

Table III: Summary One way analysis of variance

Groups	Count	Average	Variance	Standard deviation		
Agree somewhat	8	32,89	83,07	9,1140		
Strongly agree	8	20,51	57,75	7,5993		
Disagree somewhat	8	26,33	47,25	6,8736		
Strongly disagree	8	20,27	179,48	13,3970		
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F criteria**
Between Groups	852,58	3	284,1926	3,0928	0,0430	2,9467
Within Groups	2572,82	28	91,8862			
Total	3425,39	31				

time, resources or skills and that people don't believe that we need a code of ethics for software engineering.

V. DISCUSSION

The next discussion is on the research results of this study, starting with the results on the participants' profiles.

The majority (84,31%) of the participants were male, a determination, which should not be surprising to many given the history of the software industry. Male dominance of respondents is also observable in Charlesworth and Sewry's [21] study. In this study, the majority (60,78%) of the participants were between 18 and 39 years of age, thus showing that the South African software industry has somewhat young people. Furthermore, the industry has well qualified employees, with the overwhelming majority (86,28%) in possession of either a diploma or degree or post graduate degree. Another notable observation is that almost 54% of the participants had more than 10 years of experience in the software development industry.

The purpose of this study was to evaluate factors that practitioners considered as having an impact on their ethical behaviours in software development. The following discussion presents some observations on the results regarding the factors.

People don't know about ethics in software engineering: The lack of knowledge in software engineering ethics and lack of belief in codes of ethics by other stakeholders emerged as worrying factors. The results in Table 2 show that practitioners are reporting a significant lack of knowledge (74,6%) about ethics by other stakeholders in software development. Only 25,24% of the practitioners reported that people know about ethics. Software practitioners are expected to be ethically aware to competently deal with ethical dilemmas as they arise in software development. In handling ethical dilemmas, engagement with other stakeholders such as management, users, suppliers, and other non-technical colleagues may be required [30]. Therefore, such stakeholders should have some awareness or understanding of ethics and be inclined towards observing them and supporting the software development practitioners to deal with ethical dilemmas. The implications of these results are that lack of knowledge by some of these stakeholders may impede the addressing of ethical issues and subsequently, lead to the development of unethical software products. In as much as practitioners may be aware and have a willingness to deal with ethics if collaborating stakeholders lack knowledge of ethics, the efforts of practitioners may fail *to bring about ethical software*.

People don't believe that we need a code of ethics for software engineering: Bricknell and Cohen [30] state that an organizational ethical climate is enhanced by the adoption of standards of ethical behaviour alive in codes of ethics by the majority of employees. Although McKinney *et al*. [31] and Schwartz [32] consider codes of ethics useful devices for fostering ethical behavior, in this study, only 43,69% of the practitioners reported that people believe that there is a need for codes of ethics. As a result, 56,31% reported that people do not believe there is a need for a code of ethics. This reported lack of belief in codes may lead to behaviors inconsistent with principles outlined in the codes.

Therefore, the more people or other stakeholders believe that there is value in codes of ethics, the more they will refer to them for guidance when dealing with ethical dilemmas. As a result, the practitioners will start enjoying the support of other stakeholders in using the codes. However, as per the results of this study, it appears that software practitioners may be challenged to implement ethical code principles as the other stakeholders do not support them.

Our processes don't encourage stronger ethical behavior: Processes followed in carrying out various aspects of the software are also important in assisting in adhering to ethical standards. Therefore, having mechanisms to standards can strongly encourage ethical behavior. In this study, 59,26 % of the practitioners reported processes used in their places of work are supportive towards encouraging ethical behavior. Although 59,26% is the majority, it is worrying that the remaining 40,74% observed that their processes are not supportive of ethical behavior.

There is too much work pressure due to lack of time, resources or skills: Software development projects, like any other project, are time-managed, therefore they have clear expectations about delivery times. In addition, besides the fact that software projects are naturally pressure intensive due to limits of time, resources and skills, they are characterized by a continuous need for change [33]. As a result, such pressures from these factors make it prone to not adhering to the expected

ethical standards of behavior. The study results show that just over 24% of the respondents agreed strongly with the statement, while 42,72% of them agreed somewhat with it. However, 23,3% of the participants disagreed somewhat with the statement, while 9,71% disagreed strongly with it. In essence, the majority (66,99%) attest that their work exerts too much pressure on them. The implications are that pressures may make it difficult for practitioners to observe ethical principles.

Management and/or users have unrealistic expectations: Practitioners are responsible for delivering the different work products of the software project, therefore, the successful delivery of the project is in their hands. Therefore, managers and users are within their right to expect such deliveries from the practitioners. These expectations have to be in line with the project constraints such as time, budget and quality. The results in Table 2 indicate that majority of the participants agreed (28,16% agreed strongly and 29,81% agreed somewhat) that management and user stakeholders have unrealistic expectations. Seeing that majority of the respondents experience high levels of unrealistic expectations, it is likely that in trying to meet the unrealistic expectations practitioners may not prioritize ethics. Conversely, 20,39% of the respondents disagreed strongly with the statement, and the remaining 11,65% of them disagreed somewhat with it. Therefore, the observation ,in terms of lack of time, resources and skills as having high reported a negative impact on adherence to ethical behavior, the same applies to the expectations of management and users. Therefore, it is important that these factors are seriously considered to ensure that the pressure is lessened to allow practitioners reasonable space to adhere to ethics in their dealings with software development.

We don't have control over external software developers: As alluded above software development involves software developers. These developers may be allocated to the project on different terms, permanent or contracted from another company due to specialized skills or other structural arrangements. Although having external developers has several advantages, the fact that they may be independent, freelance or contracted from external sourcecomes with risks of lack of control by the mainstream practitioners. However, in both forms of employment, developers are bound (ought) to conduct themselves ethically. In this study, as shown in Table 2, 20,39% of the respondents agreed strongly that they had no control over external software developers, while 29,13% of them agreed somewhat with the statement. However, 33,01% and 17,48% of the participants disagreed strongly and disagreed somewhat with the statement respectively. This shows a nearly equal split of the responses to this statement. Having developers that you are not able to control may make them not feel committed to working according to the ethical standards but be committed towards the delivery of their allocated tasks because the risk of ethical accountability may be understood to rest with the mainstream software developers.

External factors, such as politics and the economy, affect our behavior: Consideration of external factors is critical because these factors do have an influence on the ethical conduct of employees. The study results show that minority respondents (16,5% agree strongly and 29,13% agree somewhat) support the statement under the test, whilst the majority of respondents (24,7% disagree strongly and 30,1% disagree somewhat) do not support the statement. Therefore, most of the respondents do not believe that external factors are considered to negatively influence their ethical behaviour. The implication of these results is that the practitioners are free to exercise their ethical competence without undue influence from these factors.

The preceding discussion presented factors which South African software professionals consider to be impacting their ethical behavior either positively or negatively. The following section provides some recommendations for organizations on the factors that were found to have unwanted ethical influence on software practitioners.

VI. CONCLUSION AND LIMITATIONS

The study found that practitioners believe that some factors considered in this study inhibit their ethical behaviors whilst other factors facilitate ethical behavior. Corporate processes and corporate cultures are reported as supporting ethical behaviour. Furthermore, external factors are reported not to negatively impact the ethical behaviour of practitioners. These findings are encouraging for organizations, clients and the public at large. In contrast, people-related factors, including lack of knowledge about ethics, people not believing in the need for codes of ethics for software engineering, work pressures and unrealistic expectations from stakeholders are reported by the respondents of this study as having a negative impact on their ethical behaviour.

Software development organizations are encouraged to nurture the factors that were identified as promoters of ethical behavior, but should pay particular attention to those factors that were found to be inhibitors of ethical conduct and behavior expected from the software practitioners. The ones that command urgent attention from organizations are work pressures and unrealistic expectations from stakeholders, because organizations have direct control over them.

Considering the number of respondents that participated in the study in relation to the size of the South African industry, generalization of this paper's findings would have to be done so with caution. However, the findings provide an opportunity for further investigation to be conducted on a broader scale.

ACKNOWLEDGMENT

The authors would like to express their gratitude to all participants of the study for their efforts and time – you made this study possible. Furthermore, they also express their appreciation to their institution for its support with resources.

REFERENCES

[1] R. Hotten, "Volkswagen : The scandal explained," *BBC News*, 2015. [Online]. Available: https://www.bbc.com/news/business-34324772

[2] J. C. Wong, "Greyball: how Uber used secret software to dodge the law," *The Guardian*, 2017. [Online]. Available: https://www.theguardian.com/technology/2017/mar/03/uber-secret-program-greyball-resignation-ed-baker

[3] A. Moyo, “Experian hacked, 24m personal details of South Africans exposed,” *ITWeb*, 2020. https://www.itweb.co.za/content/rxP3jqBmNzpMA2ye (accessed Aug. 12, 2020).

[4] C. Pratt and T. L. Rentner, “What’s Really Being Taught About Ethical Behavior,” *Public Relat Rev*, vol. 15, no. 1, pp. 53–66, 1989, doi: https://doi.org/10.1016/S0363-8111(89)80032-3.

[5] L. N. K. Leonard, T. P. Cronan, and J. Kreie, “What influences IT ethical behavior intentions - Planned behavior, reasoned action, perceived importance, or individual characteristics?,” *Information and Management*, vol. 42, no. 1, pp. 143–158, 2004, doi: 10.1016/j.im.2003.12.008.

[6] B. S. Cruz, D. Murillo, and O. Dias, “CRASHED BOEING 737-MAX: FATALITIES OR MALPRACTICE?,” *Global Scientific Journals*, vol. 8, no. 1, pp. 2615–2624, 2020, [Online]. Available: www.globalscientificjournal.com

[7] S. Gupta, A. Mishra, and M. Chawla, “Analysis and recommendation of common fault and failure in software development systems,” in *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings*, Jun. 2017, pp. 1730–1734. doi: 10.1109/SCOPES.2016.7955739.

[8] L. K. Trevino, “Ethical Decision Making in Organizations: A Person-Situation Interactionist Model,” *The Academy of Management Review*, vol. 11, no. 3, pp. 601–617, 1986, [Online]. Available: https://www.jstor.org/stable/258313

[9] O. C. Ferrell and L. G. Gresham, “A Contingency Framework for Understanding Ethical Decision Making in Marketing,” 1985. [Online]. Available: https://about.jstor.org/terms

[10] G. Reynolds, *Ethics in Information Technology*, Second Edition. Massachusetts: Thomson, 2007.

[11] R. Haines and L. N. k. Leonard, “Individual characteristics and ethical decision-making in an IT context,” *Industrial Management & Data Systems*, vol. 107, no. 1, pp. 5–20, Jan. 2007, doi: 10.1108/02635570710719025.

[12] R. E. Fairley and M. J. Willshire, “Why the Vasa sank: 10 Problems and some antidotes for software projects,” *IEEE Software*, vol. 20, no. 2. pp. 18–25, Mar. 2003. doi: 10.1109/MS.2003.1184161.

[13] J. Rest, *Development in judging moral issues*. Minnesota Press, 1992.

[14] I. Ajzen, “The Theory of Planned Behavior,” *Organ Behav Hum Decis Process*, vol. 50, no. 2, pp. 179–211, 1991.

[15] L. Kohlberg and R. H. Hersh, “Moral Development : A Review of the Theory,” vol. 16, no. 2, pp. 53–59, 1977, doi: 10.1146/annurev.ecolsys.3.

[16] D. Banerjee, T. P. Cronan, and T. W. Jones, “Modeling IT Ethics: A Study in Situational Ethics,” 1998.

[17] M. Bommer, C. Gratto, J. Gravander, and M. Tuttle, “A behavioral model of ethical and unethical decision making,” *Citation Classics from The Journal of Business Ethics: Celebrating the First Thirty Years of Publication*, vol. 6, no. 4, pp. 265–280, 1987, doi: 10.1007/978-94-007-4126-3_5.

[18] M. Dey, S. Bhattacharjee, M. Mahmood, M. A. Uddin, and S. R. Biswas, “Ethical leadership for better sustainable performance: Role of employee values, behavior and ethical climate,” *J Clean Prod*, vol. 337, Feb. 2022, doi: 10.1016/j.jclepro.2022.130527.

[19] S. Pagliaro, A. lo Presti, M. Barattucci, V. A. Giannella, and M. Barreto, “On the effects of ethical climate(s) on employees’ behavior: A social identity approach,” *Front Psychol*, vol. 9, no. JUN, Jun. 2018, doi: 10.3389/fpsyg.2018.00960.

[20] A. Newman, H. Round, S. Bhattacharya, and A. Roy, “Ethical Climates in Organizations: A Review and Research Agenda,” in *Business Ethics Quarterly*, Oct. 2017, vol. 27, no. 4, pp. 475–512. doi: 10.1017/beq.2017.23.

[21] M. Charlesworth and A. D. Sewry, “South African IT industry professionals ’ ethical awareness : an exploratory study,” in *Proceedings of SAICSIT 2004*, 2004, pp. 269–273.

[22] T. L. Baker, T. G. Hunt, and M. C. Andrews, “Promoting ethical behavior and organizational citizenship behaviors: The influence of corporate ethical values,” *J Bus Res*, vol. 59, no. 7, pp. 849–857, Jul. 2006, doi: 10.1016/j.jbusres.2006.02.004.

[23] R. R. Sims, “The institutionalization of organizational ethics,” *Journal of Business Ethics*, vol. 10, no. 7, pp. 493–506, 1991, doi: 10.1007/BF00383348.

[24] S. J. Vitell and E. R. Hidalgo, “The impact of corporate ethical values and enforcement of ethical codes on the perceived importance of ethics in business: A comparison of U.S. and Spanish managers,” *Journal of Business Ethics*, vol. 64, no. 1, pp. 31–43, 2006, doi: 10.1007/s10551-005-4664-5.

[25] C. K. Dimitriou and J. P. Ducette, “An analysis of the key determinants of hotel employees’ ethical behavior,” *Journal of Hospitality and Tourism Management*, vol. 34, pp. 66–74, Mar. 2018, doi: 10.1016/j.jhtm.2017.12.002.

[26] S. P. Deshpande, J. Joseph, and R. Prasad, “Factors impacting ethical behavior in hospitals,” *Journal of Business Ethics*, vol. 69, no. 2, pp. 207–216, Dec. 2006, doi: 10.1007/s10551-006-9086-5.

[27] M. E. Brown, L. K. Treviño, and D. A. Harrison, “Ethical leadership: A social learning perspective for construct development and testing,” *Organ Behav Hum Decis Process*, vol. 97, no. 2, pp. 117–134, 2005, doi: 10.1016/j.obhdp.2005.03.002.

[28] L. N. K. Leonard and T. P. Cronan, “Attitude toward ethical behavior in computer use: A shifting model,” *Industrial Management and Data Systems*, vol. 105, no. 9, pp. 1150–1171, 2005, doi: 10.1108/02635570510633239.

[29] K. D. Martin and J. B. Cullen, “Continuities and extensions of ethical climate theory: A meta-analytic review,” *Journal of Business Ethics*, vol. 69, no. 2, pp. 175–194, Dec. 2006, doi: 10.1007/s10551-006-9084-7.

[30] K. I. Bricknell and J. F. Cohen, “Codes of ethics and the information technology employee: the impact of code institutionalisation, awareness, understanding and enforcement,” *Southern African Business Review* , vol. 9, no. 3, pp. 54–65, 2005.

[31] J. A. McKinney, T. L. Emerson, and M. J. Neubert, “The Effects of Ethical Codes on Ethical Perceptions of

Actions Toward Stakeholders," *Journal of Business Ethics*, vol. 97, no. 4, pp. 505–516, Dec. 2010, doi: 10.1007/s10551-010-0521-2.
[32] M. Schwartz, "The nature of the relationship between corporate codes of ethics and behaviour," *Journal of Business Ethics*, vol. 32, no. 3, pp. 247–262, 2001, doi: 10.1023/A:1010787607771.
[33] F. P. J. Brooks, "silverbullet," in *No Silver Bullet – Essence and Accident in Software Engineering*, 1986.

Modeling and Simulation of a Hybrid Electromagnetic Accelerator

Erol KURT
Faculty of Technology
Department of Electrical and Electronics Engineering,
Gazi University
Ankara, Turkey
ekurt52tr@yahoo.com

Kayhan CELIK
Faculty of Technology
Department of Electrical and Electronics Engineering,
Gazi University
Ankara, Turkey
kayhancelik1923@gmail.com

Ahmet Yasir TEKSAR
Faculty of Technology
Department of Electrical and Electronics Engineering,
Gazi University
Ankara, Turkey
ayteksar@gmail.com

Hakan GORMUS
Faculty of Technology
Department of Electrical and Electronics Engineering,
Gazi University,
Ankara, Turkey
hakan.gormus523@gmail.com

Emin ÖZDEMİR
Faculty of Technology
Department of Electrical and Electronics Engineering,
Gazi University
Ankara, Turkey
eminozdemir016@gmail.com

***Abstract*—In this study, Modeling and Simulation of Hybrid Electromagnetic Accelerator was done. With the electromagnetic force acting on the projectile, the projectile accelerates in the rail launcher. In the coil gun, the projectile is accelerated by the coil acting as an electromagnet. A hybrid launcher is formed by the coaxial combination of the coil launcher and the rail launcher. The projectile velocity changes according to the magnitude of the excitation current given to the coil and the rails. In this study, magnetic analyzes and force analyzes were performed with ANSYS Maxwell simulation application.**

Keywords—Railgun, coil gun, hybrid electromagnetic accelerator, ANSYS Maxwell

I. INTRODUCTION

Since the beginning of the 20th century, research on electromagnetic launcher systems has been increasing and this interest has been expanding in recent years with the work of many research institutions [1, 2, 3, 4]. The main reason behind this situation is that the possibility of the using these systems in different sectors such as the defense and space industry[5].

An electromagnetic launcher is a kind of weapon, which has the capability of using electromagnetic force to accelerate the armature (i.e., projectile) to super high speeds. According to their structural design, the electromagnetic launchers are divided into 2 main parts: Railgun and reluctance varying gun. Within the present work, we aim to combine these two designs into a unique form as a hybrid launcher for the first time to our knowledge. Among them, railguns make use of a pair of conducting parallel rails and a sliding armature connecting the two rails to exploit the electromagnetic Lorentz force. The electrical current is injected to the single rail through the armature, then this current flows through the projectile holder and completes the circuit over the second rail. The current flowing through the parallel rails creates an immense magnetic field around the rail as well as the magnetic projectile holder. In addition, this magnetic field interacts with the current flowing through the holder to create a Lorentz force [5]. On the one hand, Electromagnetic Coil Launcher (EMCL) (i.e., reluctance varying gun) is a special launch device utilizes coaxial arrangement helix coils (HC) to generate the electromagnetic forces acting on the armature. Indeed, the coil gun has an elements of the launch coil, the high voltage capacitors, power source, trigger and gating system for discharging the energy to the launch coil and projectile. When the system is fired, the launch coil produces a high magnetic field that pulls the ferromagnetic projectile into the barrel. When the bullet reaches to the launch coil, the magnetic field is turned off which means that causing the bullet to eject from the muzzle at high speed.

According to literature, there exist some attempts to form a hybrid launcher. For instance, in Ref. [1], an electromagnetic launcher with a coil and an electromagnetic launcher with rails and as a new type of linear drive is considered as first hybrid electromagnetic launcher. The missile velocity was explored for the hybrid electromagnetic launcher with pneumatic assist by Domin et al. [1] in a different work. The system is made up of the module-P (Pneumatic), module-C (Coil) and module-R (Rail) which ones increase the velocity of missile by 29 %, 25.4% and 45.6% respectively. The projectile speed changes with the change of the module's parameters such as pressure and voltage. So, optimization can be made for the projectile speed over the values.

Whereas, in their study, this integrated construction is additionally equipped in an introductory pneumatic

module intensifying the meaning of the word "hybrid" and allows to introduce the developed full name: A hybrid electromagnetic launcher with the pneumatic assist according to their structure. As regards the pneumatic module, it is required for setting the introductory speed at the entry to the electromagnetic launcher [1].

There are different combinations denoted as "hybrid" in the literature. A novel simulation strategy for an electromagnetic launcher (EML) suggested for the aim of reducing complexity with a high accuracy by Yıldırım et al., [4]. The inductance and electromotive force (EMF) variations in the transient condition are modeled by using a finite element method (FEM) coupled with an electrical circuit simulation. To conclude, the proposed method showed that the transient inductance *L* and EMF calculations improved the simulation-experiment accuracy up to 3.49% in average.

In other work made by Cao et al. [6], the railgun, the magnetic field at vertical and horizontal axis are compared and analyzed which show that different position of the armature has influence on the magnetic field by. The analyses indicate that the peak value of the magnetic field should be in the middle of the railgun, so that, the shielding design can be accomplished protection of magnetic field inside the railgun.

In the work of Ciolini et al. [7]., the electromagnetic hardening on payloads of railgun is investigated with the help of experimental results and mathematical calculations. It is found that electrical shocks on the electronic devices such as DDR2-RAM and 7-SEG Display makes these devices self-destructed and unusable. In other paper, quasi-analytic simulation (QAS) and numerical simulation (NS) with 3D model are compared for electromagnetic gun's model by Orbach et al., [8]. QAS and NS have an error rate of 4% and 6% respectively, on the other hand, NS needs a longer time in the 3D model simulations than the QAS.

Three different railgun systems were investigated by Hundertmark et al., [9] for the calibration of the muzzle velocity of the masses in the desired range by using systems which have the various level of energy stored. The effect of different armature technologies (i.e., C-shaped armature and brush armature) were compared by Hundertmark et al., [2]. The C-shaped armature has an excellent performance with the efficiency of 41% than the other one, which has the only 23%. In a different work, Werst et al., [10] presents a simple, ultra-rigid, low-mass design providing 5 to 6 times the rail-to-rail structural rigidity of a conventional bolted, composite sidewall type EM gun construction. Their proposed design allowed the production of large, light, and structurally rigid rail rifles.

A co-simulation method of a compensated pulsed alternator (CPA) which was one of the power supplies for electromagnetic railguns, due to its high energy density was presented by Shumei Cui et al. [11]. The railgun and CPA models were implemented on the MATLAB/Simulink and FLUX2-D, respectively. The simulation results indicate that the optimum length of the barrel has the relation with trigger moment of alternator.

Asynchronous discharge of a multi-module superconducting pulsed power supply (SPPS), driving an electromagnetic rail gun, which reduced the amplitude and increased the pulse width was discussed by Falong Lu et al. [12]. The obtained results indicate that the increasing of the trigger delay time decreased the max armature current, max force, max acceleration, muzzle velocity and efficiency. Some methods were examined by Keshtkar et al., [13] to reduce the electromagnetic force on the rails which were a stepped laminated armature and laminated rails with different electrical specifications. It was found that change of the rail's resistivity decreases the maximum electromagnetic forces on the rails.

An animated mesh finite element/boundary element (FE/BE) hybrid scheme using the shift variant properties of the field distribution along the rail was proposed by Wang et al., [14]. This method solved the electric and magnetic fields directly in the conductor with the high computational speed and more advantageous than the COMSOL and EMAP3D programs. A super-fast hybrid launcher which combined a gas gun and coil gun was presented by Balikci et al., [15]. The optimization of the transition between the gas gun and the first section of the coil gun, and between successive sections of the barrel was realized by the software. The experimental and simulation results proved that it could be used in super-fast applications. The air-cooling and water-cooling methods were explored for high temperature rise problem during high energy discharge and filling in the electromagnetic launcher by Zhou et al., [16]. The analyses showed that air was more advantageous than the water for faster cooling of the battery of the electromagnetic launcher. A novel design and simulation of hybrid electromagnetic launcher (MFELS) combining the coil gun with the multipole field winding was proposed by Kondamudi et al. [17]. The proposed design had an advantage of the achieving the higher projectile velocities and force then the conventional coil gun.

In this paper, the novel design of the hybrid electromagnetic launcher is presented which has the high projectile speed due to the combination of magnetic force and Lorentz force by gathering the coil-gun and railgun launchers on the same axis. The designed system is the multi-stage coaxial hybrid launcher, which combines the coil and rail guns with the advantages such as large thrust, high efficiency. Therefore, it is suitable for faster the launching of the military bullets, which has the low-speed limit due to its large mass and it is thought that the designed system can be widely used in military applications due to these features.

II. THEORETICAL BACKGROUND

Electromagnetic force and magnetic flux density are used to move the projectiles in railgun and coil gun designs, respectively. In the railgun design, Fig. 1 represents design of railgun, the current flowing through the rails creates magnetic

fields and creates Lorentz Force, which keeps the projectile going all the way to the end of the rail. After current is applied to the rails, there is no reverse Electromotive Force (EMF) to slow down the projectile. Therefore, railgun is more advantageous than coil gun. In the railgun design, Lorentz force is given by:

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{1}$$

Here, the charge q, electric field E, instantaneous velocity u, and magnetic flux density B are denoted, respectively. As can be seen, we can say that the Lorentz force is formed due to electric-magnetic fields.

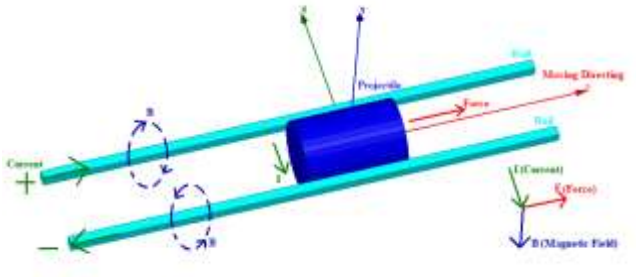


Fig. 1. Railgun Design Scheme

When current-carrying wires are parallel, they create a push-pull force against each other. This force is sometimes called the Laplace force, that:

$$\mathbf{F} = I\boldsymbol{\ell} \times \mathbf{B} \tag{2}$$

where the vector whose magnitude is the length of wire is l, and current is I. Since the rails are fixes. The Laplace force will act on each other, but the rails will remain fixed.

The magnetic component of the Lorentz Force ($q\mathbf{v} \times \mathbf{B}$) is responsible for the moving EMF. The ($q\mathbf{E}$) component is responsible for the electric force. In short, the Lorentz Force is created with the current flowing through the conductors and the railgun design is realized with this force. The greater the force applied to the projectile, the faster the projectile will move. The magnetic flux density B formula for a current carrying conductor is given by:

$$\mathbf{B}(r) = \frac{\mu_0 I}{4\pi r}\vec{\varphi} \tag{3}$$

given by the formula. where s is the perpendicular distance from the point on the armature to the axis of one of the wires. Note that $\vec{\varphi}$ between the rails is $\vec{z}$ assuming the rails are lying in the *x-y* plane.

When the electrical current flows through the windings, the magnetomotive force is formed by the multiplication of windings turn N and current I. The reluctance R is against the flux and maximizes by the non-magnetic media such as airgap. Therefore, the magnetic flux changes with the reluctance variance as given by Hopkins's law for magnetic circuits given [18],

$$NI = R\Phi \tag{4}$$

Here I, N, Φ and R are current, number of turns, magnetic flux, and reluctance, respectively. The reluctance here is given by,

$$R = \frac{1}{\mu_0 \mu_r}\frac{l}{S}, \tag{5}$$

In this formulation, permeability of vacuum $\mu_0 = 4\pi.10^{-7}$ H/m, relative magnetic permeability μ_r, cross-sectional area of the circuit S, and length of the magnetic circuit l.

In Table 1. The characteristic and geometric features of the system is given.

TABLE I. THE CHARACTERISTIC AND GEOMETRIC FEATURES OF THE SYSTEM

Properties	Value
Coil length (cm)	5
Number of turns (approximated)	1000
Coil's material	Copper
Coil inner radius (cm)	2
Coil outer radius (cm)	3
Magnet's material	NdFe30
Magnet radius (cm)	0.699
Shape of magnet	Regular Polyhedron
Rail's material	Copper
Rail length (cm)	10
Rail diameter (cm)	0.2
Distance between rails (cm)	1.4

III. RESULTS AND DISCUSSION

In Fig. 2(a), the model with railgun, reluctance variance coil and projectile are shown. Here the current path located just bottom of the projectile can move freely so that it can be accelerated by rail current from the left-hand side to the right. The meshed model is given in Fig. 2(b). In the meshing structure, the sides and corners have more tetrahedral mesh element due to the sensitive chances in value. In the electromagnetic modeling, especially the magnetic materials should have high meshing elements for the sensitive solution of electromagnetic formulation. Fig. 2(c) shows the rail gun, reluctance varying coil, projectile and current paths in rail and winding as formed in Ansys Maxwell package. According to Equation (1), the magnetic force over the rail part at the bottom of projectile drives the projectile from left to the right. This model enables to get benefit from railgun acceleration and the force ignited by the reluctance variance.

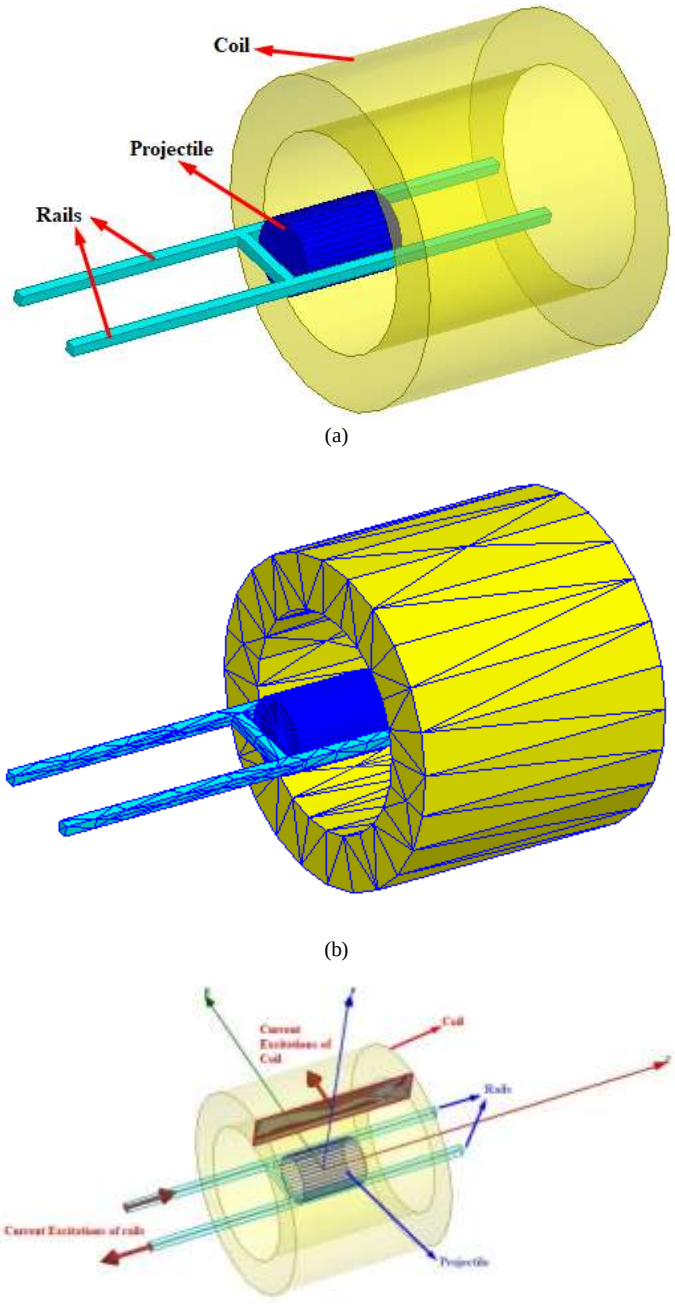


Fig. 2. (a) 3D model, (b) meshed structure of the hybrid model, (c) current excitations.

A coil is used in the coil gun design, Figs. 2(a, b, c) represent the design of coil gun, and when a moving object passes through the coil, the coil act as an electromagnet and accelerate the object. In order to accelerate the projectile moving along the central axis of the coil, the coils must be switched on-off and should occur in precisely times sequence. The acceleration of the projectile in the coil gun is provided by the magnetic flux density *B*.

In Fig. 3, the projectile is shown positioned on x-z plane. Due to the high magnetic field occurred in the middle regions of coil, arrows have higher values along the center of coil. For the current I = 100 A we have received B = 4.92 T for the winding turn N = 1000. The current flowing inside the windings causes high magnetic field in the inner part, whereas the flux lines are closed themselves at the outer part of the coil with relatively lower flux densities.

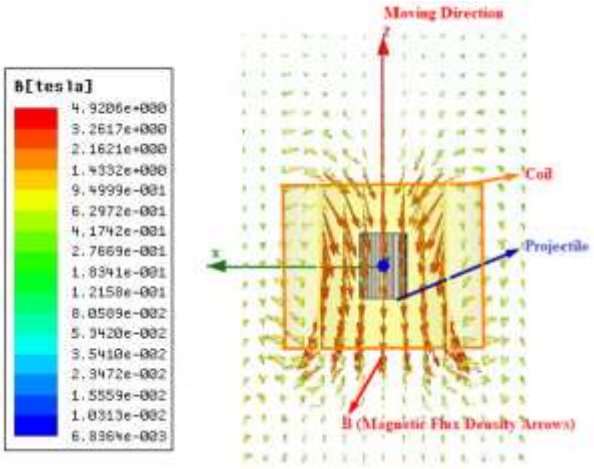
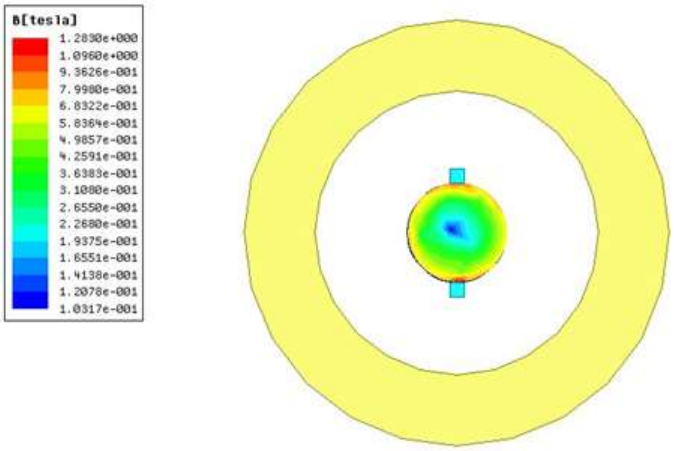
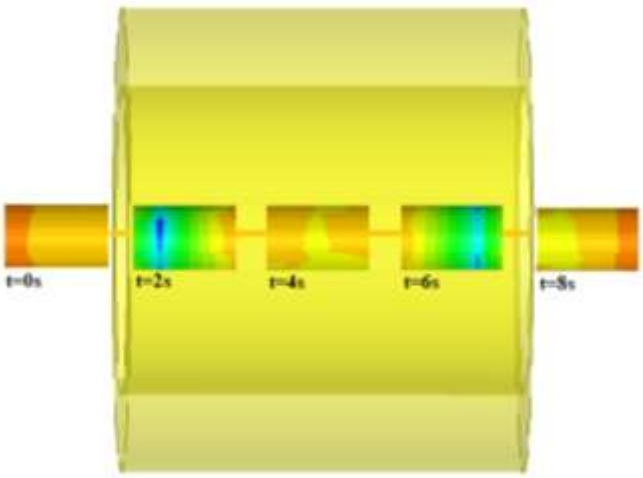


Fig. 3. Coil gun Design Scheme

In Fig. 4(a), the magnetic flux density over the projectile end is shown. The ferromagnetic projectile has high flux densities such as B = 1.2 T especially nearby the rail gun. The central region of the projectile has lower flux densities around 0.1 T. The outer circle denotes the inner and outer part of the coil. Note that both the effect of coil and railgun are added to the results of flux densities. In Fig. 4(b), the projectile is shown step by step for different time lags such as t = 0, 2, 4, 6, 8 s in 3D model. Since the projectile is about to enter the coil at t = 0 s, the flux density value is higher on the left of the bullet. As the bullet enters the coil at t = 2 s, the *B* value on its left side has decreased considerably. At t = 6 s and t = 8 s, similar flux density patterns occur on the right side of the projectile in that case. The flux density of 0.8 T is achieved at the middle of the coil, when the projectile is positioned at that location. On the one hand, the flux densities vary from 0.1 T to 0.79 T nearby the outer areas of the coil.



(a)



(b)

Fig. 4. (a) Magnetic flux of projectile, (b) Magnetic flux of projectile and t = 0, 2, 4, 6, 8 s

In Fig. 5, the magnetic flux density nearby the coil windings is shown on the *x-y* plane. The current of coil and rails current are adjusted 100A, 50A, respectively. The *B* value inside and outside the coil is found as $B = 1.44$ *T*. The maximum number of mesh elements are 10000 for the formation of the windings. The magnetic flux density on the projectile is $B = 2.07\text{x}10^{-2}$ *T* since it is outside the coil at t=0 s.

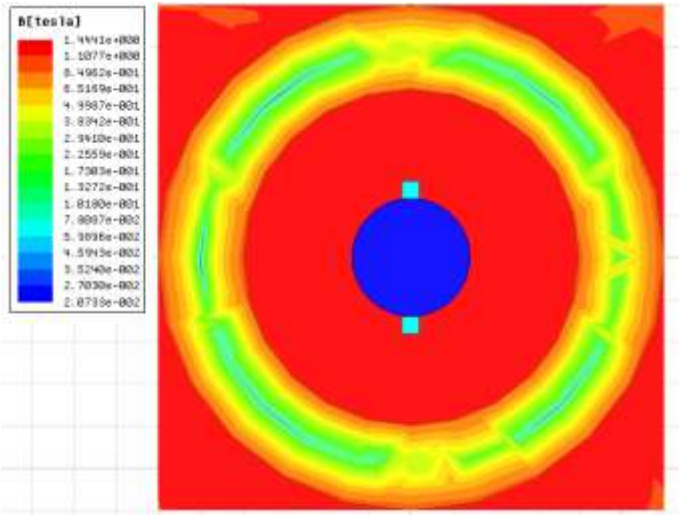


Fig. 5. Magnetic flux density distribution nearby the coil windings.

The mechanical force along the is *z* axis is shown in Fig. 6. The adjusted currents are 50A, 100A, and 150A, respectively. The animated 3D Model between 0s and 8s is given for this simulation. While the force value is 65.10 N for the current $I = 50$A in t = 7s, it is 127.88 N for $I = 100$A in t = 7s. Besides, the force reaches to 189.72 N for I = 150A in t = 6.75s. According to Eq. 2, the increase in current yields to an increase in force.

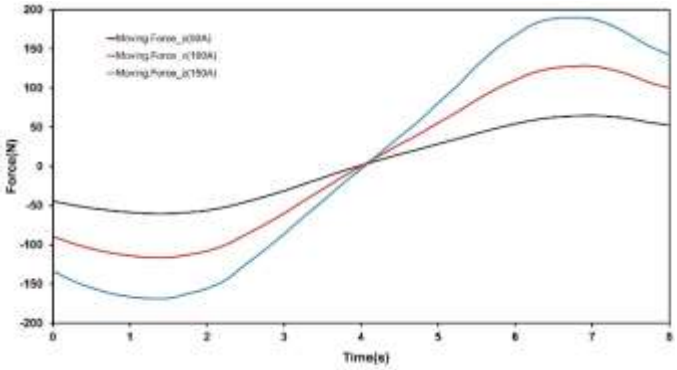
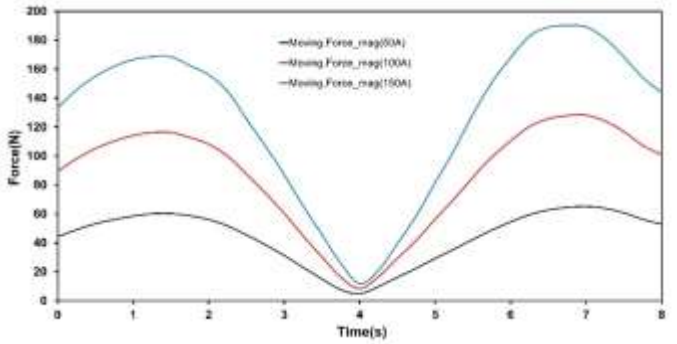
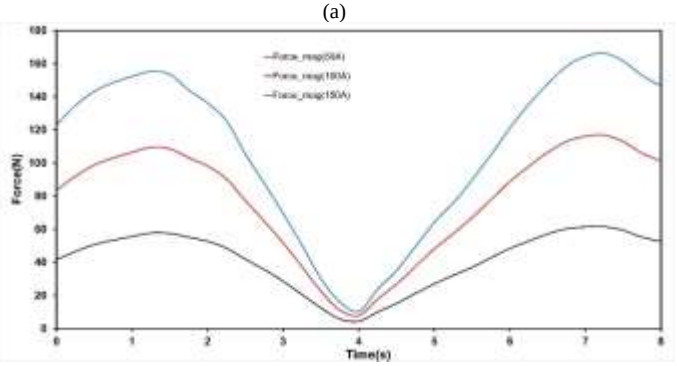


Fig. 6. Mechanical force component along z for current I = 50, 100, and 150 A.

In Fig. 7(a), the results of moving force for various coil and rail currents are given as $I = 50$ A, $I = 100$ A and $I = 150$ A, respectively. The projectile moves along the central axis of the coil between $t = 0$ s and t = 8 s. We have the force value 190.2 N for 150 A excitation, and it varies to other values for different currents. Note that the maximum moving force is obtained at the vicinity of the coil tips since the gradient of the magnetic energy denotes the force. Also note that at the right-hand side of the coil, the force is much higher due to the rail structure contribution. In Fig. 7(b), the magnetic force is plotted for $I = 50$ A, $I = 100$ A, $I = 150$ A, respectively. The projectile moving along the central axis of the coil faces with various magnetic forces along the coil axis. Note that this value is the magnitude of the field. For instance, it has 166.2 N for I = 150A in t = 7.25s.



(a)



(b)

Fig. 7. (a) Magnitude of Moving Force for coil and rail currents, $I = 50$, 100, 150 A in z axis, Maximum number of elements as 1000 in Maxwell package, (b) Magnitude of force for coil and rail currents, $I = 50$, 100, 150 A in z axis. Maximum number of elements as 1000 in Maxwell package

In Fig. 8, the force on the projectile moving along the center of coil between the rails has different maximum number of mesh elements from $t = 4$ s to $t = 8$ s under the constant current $I = 150$ A. When the projectile is at its $t = 4$s time, the force equals to zero, approximately. The moving force increases before the projectile leave from the tip. The maximum moving force is obtained at the vicinity of $t = 6.75$s for all maximum number of mesh elements. Note that the results do not change much for different meshing numbers, thereby the meshes with 1000 elements are sufficient for the design.

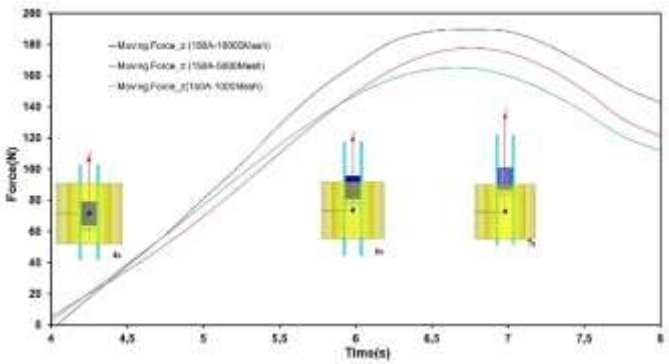


Fig. 8. Moving force for coil and rail currents, $I = 150$ A. in z axis. Maximum number of elements as 1000, 5000, and 10000 in Maxwell package

IV. Conclusion

In the present work, an electromagnetic gun in a hybrid structure is designed and flux structures and forces are discussed. It has been proven that the flux density value of 0.4 T exists at the tip of the projectile inside the coil, when it starts its move along the rails. In our electromagnetic analysis, we neglect the mechanical friction. A closer look to the model yields to the items below:

- By increasing the values in the mesh topology of the 3D Model, smoother flux density lines are formed along the coil axis.
- By increasing the excitation currents (i.e. coil and rail currents), the magnetic force acting on the projectile is increased up to 190 N for this small geometry.
- The increasing forces yield to high velocities at the right tip of the coil. In the coil systems, coil current should be increased to increase the magnetic field affecting the projectile.
- In same variables and same values, we had reached different graphics that has different number of meshing structure, and maximum value of graphic has different.

References

[1] K. Kluszczyński and J. Domin, "Hybrid electromagnetic launcher with pneumatic assist-influence of input supply data upon final velocity missile," in 16th International Conference on Research and Education in Mechatronics, 2015, November.

[2] S. Hundertmark, G. Vincent, D. Simicic and M. Schneider, "Increasing launch efficiency with the PEGASUS launcher.," IEEE Transactions on Plasma Science, vol. 45, no. 7, pp. 1607-1613, 2017.

[3] J. V. Parker, D. T. Berry and P. T. Snowden, "The IAT electromagnetic launch research facility," IEEE Transactions on Magnetics, vol. 33, no. 1, pp. 129-133, 1997.

[4] N. Tosun, H. Polat, D. Ceylan, M. Karagoz, B. Yıldırım , I. Gungen and O. Keysan, "A hybrid simulation model for electromagnetic launchers including the transient inductance and electromotive force," IEEE Transactions on Plasma Science, vol. 48, no. 9, pp. 3220-3228, 2020.

[5] N. J. Eckert, Review of Railgun Modeling Techniques: The Computation of Railgun Force and Other Key Factors, University of Colorado at Boulder, 2017.

[6] R. Cao, Z. Duo and M. Su, "Analysis of magnetic field waveforms of different launching stages of rail gun based on wavelet transform," IEEE Transactions on Plasma Science, vol. 47, no. 1, pp. 500-507, 2018.

[7] R. Ciolini, M. Schneider and B. Tellini , "The use of electronic components in railgun projectiles," IEEE Transactions on Magnetics, vol. 45, no. 1, pp. 578-583, 2009.

[8] Y. Orbach, M. Oren, A. Golan and M. Einat, "Reluctance launcher coil-gun simulations and experiment," IEEE Transactions on Plasma Science, vol. 47, no. 2, pp. 1358-1363, 2018.

[9] S. Hundertmark and O. Liebfried, "Options for an electric launcher system," IEEE Transactions on Plasma Science, vol. 47, no. 10, pp. 4433-4438, 2019.

[10] M. D. Werst, J. R. Kitzmiller, C. S. Hearn and G. A. Wedeking, "Ultra-stiff, low mass, electromagnetic gun design," IEEE transactions on magnetics, vol. 41, no. 1, pp. 262-265, 2005.

[11] S. Cui, Q. Liu and W. Zhao, "Simulation research of a CPA powered railgun system," IEEE Transactions on Plasma Science, 41(5), vol. 41, no. 5, pp. 1484-1487, 2013.

[12] F. Lu, Z. Yan, H. Deng and Y. Wang, "Asynchronous discharge of an eight-module superconducting pulsed power supply for driving an electromagnetic railgun," IEEE Transactions on Applied Superconductivity, vol. 29, no. 2, pp. 1-5, 2018.

[13] A. Keshtkar, A. Rabiei and L. Gharib, "Effect of armature and rails resistivity profile on rail's electromagnetic force and armature velocity," IEEE Transactions on Plasma Science, vol. 43, no. 5, pp. 1541-1545, 2015.

[14] G. H. Wang, L. Xie, Y. He, S. Y. Song and J. J. Gao, "Moving mesh FE/BE hybrid simulation of electromagnetic field evolution for railgun," IEEE Transactions on Plasma Science, vol. 44, no. 8, pp. 1424-1428, 2016.

[15] A. Balikci, Z. Zabar, L. Birenbaum and D. Czarkowski, "On the design of coilguns for super-velocity launchers," IEEE transactions on magnetics, vol. 43, no. 1, pp. 107-110, 2006.

[16] R. Zhou, J. Lu, G. Wang, X. Long and X. Zhang, "Thermal management of hybrid energy storage for electromagnetic launch," IEEE Transactions on Plasma Science, vol. 45, no. 7, pp. 1459-1464, 2017.

[17] S. C. Kondamudi, S. Thotakura, M. R. Pasumarthi, G. R. Reddy, S. M. Sathyaraj and X. X. Jiang, "A novel type coil-multipole field hybrid electromagnetic launching system," Results in Physics, vol. 15, no. 102786, 2019.

[18] V. Gies, T. Soriano, S. Marzetti, V. Barchasz, H. Barthelemy, H. Glotin and V. Hugel, "Optimisation of energy transfer in reluctance coil guns: Application to soccer ball launchers," Applied Sciences, vol. 10, no. 9, p. 3137, 2020.

A video dataset for substance abuse detection

Amin Khaksar Pour
Faculty of Computer Science & Information Technology
University of Malaya
Kuala Lumpur, Malaysia
amin.khaksar@gmail.com

Omid Haselforosh
Department of computer engineering
Institute for Higher Education ACECR - Khuzestan
Ahvaz, Iran
omid.hasel97@gmail.com

Nor Badrul Anuar
Faculty of Computer Science & Information Technology
University of Malaya
Kuala Lumpur, Malaysia
badrul@um.edu.my

***Abstract*— In recent years, access to adult video content such as pornography, nudity, violence, and substance abuse has been increasing due to the significant augmentation in social media applications. Whereas there is less specific control over using mobile phones, tablets, and computers, thus the underage kids are able to access adult video content via their gadgets. Consequently, video content detection automatically plays a crucial role to control the accessing inappropriate videos and it makes using social media more suitable for the underage kids. Short video content rating (VCR) is an automatic system for rating a video for the audiences age groups. Violence, pornography, nudity, profanity languages, and substance abuse are key actions in video for content rating. Since large-scale datasets are much needed for efficient machine learning or deep learning models, researchers have introduced some datasets to identify pornography and violence as components of video content rating. Nevertheless, lack of a video dataset is a shortage to develop a deep learning model for substance abuse detection. In this study, we created a substance abuse video dataset (SAVD) extracted from other action recognition datasets, 29 related movies, YouTube, and some free stock video footage websites such as Gettyimages.com, Pexels.com, Rnvato.com, istockphoto.com, storyblocks.com. In addition, we proposed five deep learning models with transfer-learning approach for base model, then train and test the models using this dataset. Finally, we compare the models according to their accuracy.**

Keywords— Video content rating (VCR), substance abuse video dataset, substance abuse detection, deep learning, computer vision, artificial intelligence

I. INTRODUCTION

With the growth of social media, access to videos with inappropriate content comprises substance abuse has also increased. Besides, recent studies show that substance abuse and other high risk behaviours are increased due to media exposure [1]. To cope with these issues, social media organize some restricted rules for accessing teen users. For example, Facebook, Twitter, WhatsApp, Instagram, Snapchat, Musically, Skype, and define 13-year-old as the minimum age to create a profile. However, Ofcom [2] has reported that 77% of the 8-11 and 89% of the 12-15 year-old children use YouTube and 69% 12-15 year-old children have at least one profile in social media applications. These reports illustrate that the policies of social media age restriction are not enough to control accessing of the children to video adult content.

In this regard, Khaksar Pour et al. [3] proposed an automatic VCR that determines whether a short video is suitable for younger groups by investigating key adult contents such as nudity, pornography, violence, substance abuse, and profanity. [4, 5]. In the VCR system, a deep learning model is trained to detect these key contents automatically and then rate videos according to the audience's ages.

Scholars have done a lot of research on pornography and violence detection in video, and several datasets have been created and used to train machine learning models for this purpose. To illustrate, researchers published some datasets for nudity and pornography detection such as the NPDI or pornography-800 [6] and the Pornography-2k [7] as well as some datasets for violence detection such as the Hockey Fight [8], the Violence in Movie [9-12] , and The Violent Flows [13]. However, the lack of a substance abuse video dataset is shortage for training deep learning models for the substance abuse detection process. In this study, we filled this gap and collected video clips with drug abuse actions (drinking, inhaling, injecting, and smoking) from 29 related movies, YouTube and some free stock video footage websites. Likewise, we collected video clips with No-drug actions from existing action recognition video datasets. Afterwards, we trained five deep learning models by this dataset, and evaluated the model outputs as well.

II. PREVIOUS WORKS

The video content rating (VCR) would be an automatic system for rating a short video published in social media to determine audience age groups [3]. In this system, some components including pornography, nudity, violence, and substance abuse are investigated by computer vision capabilities. Detecting these components in a video is a form of action recognition.

In video content analysis or video content rating, features are extracted from data such as frames, audio, and text, and some facts are automatically identified from the video using machine learning methods. Several video content rating applications are as follow:

A) Substance abuse detection: for the time being, there is not any exactly related study for substance abuse detection in video, except that some actions relevant to drug abuse such as drinking, and smoking are seen in some existing action recognition research.
B) Nudity and Pornography detection: Some researchers' efforts are to detect adult content in the video. Pornography detection research categories are including skin detection for detecting nudity, image descriptor-based methods, video-based methods, deep learning-based methods, and child sexual abuse detection methods [14].
C) Violence detection: In video content analysis efforts, violent event detection has received the most attention from researchers. Due to the variety of invasive events, it is difficult to provide a single definition of it and

requires a complex interpretation [15]. Generally, the same techniques as other computer vision programs, such as action recognition, object recognition, surveillance, etc., are used to detect violent content in video [16].

D) Profanity detection: for offensive dialogue detection in the video, it is possible to use the audio aspect of the video to convert the speech into text, and then the profanity detection can be done from the text.

A large-scale dataset is highly required for training deep learning models efficiently. As a result, for video classification efforts, scholars create several video datasets. According to a video dataset used for machine learning algorithms, after model training, an action is detected. For the time being, several pornography and violence video dataset was published by scholars, however, there is a lack of substance abuse video dataset to train machine learning and deep learning models. Researchers have introduced datasets to identify pornography and violence as components of video content rating. Each dataset contains several video clips in different classes.

Avila et al. [6] introduced a pornography dataset is known as NPDI or pornography-800 dataset. This dataset contains 800 videos including 400 Non-pornographic (that 200 easy and 200 difficult distinguishing positions such as 'beaches', 'wrestling', and 'swimming') and 400 pornographic videos, for detection. In a more complete effort, Moreira et al. [7] have published pornography-2k dataset including 1000 non-pornographic and 1000 pornographic videos, totally 140 hours and each clip duration is from 6 second to 33 minutes. The dataset is containing professional and amateurs in different genres from live-action to cartoons and various behaviours and ethnicities. This dataset covers difficult positions of non-pornographic video such as "beach", "wrestling", "swimming", "sumo", etc.

Nievas et al. [8] created the hockey fight dataset for evaluating models of violence detection. It includes 1000 hockey game videos from the National Hockey League (NHL). It includes videos of non-violence and violence in the ice hockey sport. All videos have 50 frames of 720×576 pixels that labelled as " non-fight" or " fight", and just two or very few athletes were displayed. A real-world video dataset with 246 crowd video clips (123 violent and 123 non-violent clips) is introduced by Hassner et al. [13] . The duration range of the clips is from 1.04 to 6.52 seconds, and the average length of the clips is 3.60 seconds. In another study, the Violence in Movies dataset included 200 video clips published by Demarty et al.[9-12]. In the dateset, 100 video clips display a person-on-person fight extracted from action movies and 100 of them are non-fight videos selected from other action recognition datasets. All video clips have a 360*250 pixels average resolution and 50 frames. Soliman et al [17] presented a real-life violence situations dataset with 2000 videos contains 1000 violence videos and 1000 non-violence videos. They collected the videos from YouTube, violence videos are many real street fights situations in several environments and conditions. On the other hand, non-violence videos are collected from many different human actions like sports, eating, walking …etc. Table 1 illustrates the video datasets published for pornography and violence detection as it is described before.

III. SUBSTANCE ABUSE

Substance abuse, also known as drug abuse, refers to the use of legal or illegal substances, usually for the purpose of psychoactive effects on the brain. United Nations Office on Drugs and Crime (UNODC) [18] reports that there are around 275 million drug users in the world, and this issue has been increasing recently, especially among young and adolescence people. Drug abuse causes long-term damage to the body and increases the risk of HIV and hepatitis B and C infections. Drug abuse has various reasons such as curiosity and peer pressure especially among adolescents and young people, as part of religious practices or rituals, recreational intention, as a tool to gain creative inspiration, and sometimes it is the use of drugs to treat and reduce pain, but then it becomes addictive.

Alcohol, tobacco, cocaine from coca, opium and opioids from poppy plants, hashish or marijuana from cannabis, and synthetic drugs such as heroin, ecstasy and LSD are example of drugs. There are several methods for abusing drugs, such as oral administration in pill form, injection into the arteries, inhalation of the substance in the form of smoke, or inhalation of the substance for absorption into the blood vessels through the nose.

IV. SUBSTANCE ABUSE VIDEO DATASET

For creating substance abuse video dataset (SAVD), we collected clips from YouTube, and some free stock video footage websites such as Gettyimages.com, Pexels.com, Rnvato.com, istockphoto.com, storyblocks.com. In addition, we utilize 29 movies (see Table 2) to extract drug abuse actions. All the movies are related to substance abuse therefore they are useful for our purpose to extract substance abuse activity videos. It needs to be emphasized that we extracted many video clips comprises substance abuse actions from each movie. Furthermore we extracted No-drug clips from available action recognition video dataset comprises HMDB51[19], UCF101[20], and Real-Life Violence Situations Datasets[17].

SAVD dataset includes four drug abuse action classes (drinking, inhaling, injecting, and smoking) and one No-drug action class. We have considered the following for substance abuse actions in the videos. (See Fig. 1 that shows the sample frames of drug abuse actions and no-drug actions)

Table 1 Pornography and violence detection datasets

Action type	Dataset Name	Total videos	Number of videos in each class	Video lengths
Pornography detection	NPDI	800	400	-
	Pornography-2k	2000	1000	from 6 seconds to 33 minutes
Violence detection	The Hockey Fight	1000	500	50 frames
	Violence in Movie	200	100	50 frames
	The Violent Flows	246	123	1.04 to 6.52 seconds
	Real Life Violence Situations Dataset	2000	1000	

- Drinking alcohol: How people drink and their actions before and after drinking alcohol.
- Drug inhalation (inhaling): How to use cocaine inhalation and their actions before and after inhaling.
- Drug injection (injecting): how to inject heroin and their actions before and after the injecting.
- Smoking: includes how to smoke flowers, cannabis, and cigarettes and their actions before and after the smoking.
- No-drug: does not include any of the above and includes a normal action in the video such as sport, exercise, eating, talking, playing music, etc.

Table 2 List of movies that utilizing for create substance abuse video dataset (SAVD) – all movies have sequences related to substance abuse actions
* We extracted many video clips comprises substance abuse actions from each movie

A Star is Born 2018	American Gangster 2007	The faculty 1998
Christiane F. 1981	Blow 2001	Filth 2013
Heaven knows what 2014	In Bruges 2008	Léon: The Professional 1994
The Panic in Needle Park 1971	Fear and Loathing in Las Vegas 1998	Requiem for a Dream Directors Cut 2000
Scarface 1983	London 2005	Thirteen 2003
Drugstore Cowboy 1989	Sorry to brother you 2018	The Basketball Diaries 1995
Spring Breakers 2012	Barfly 1987	Leaving Las Vegas 1995
Boogie nights 1997	Gia 1998	Pulp Fiction 1994
Trainspotting 1996	Rough night 2017	Traffic 2000
The Great Gatsby 2013	Sideways 2004	

Fig. 1 Examples of drug abuse (drinking, inhaling, injecting, and smoking) and No-drug frames from the substance abuse video dataset (SAVD)

Due to the large number of clips, it took a long time for the clips to be extracted manually, so we wrote a Python code to do some of our works automatically. In this code, Adobe Premiere Pro software[1] is controlled using some APIs. The method is that first a clip of the movie is made in Adobe Premiere; then the file path, the output file format, storage path, and the start and finish frame numbers is stored in a Notepad file according to the related class. Therefore, it will be easy to access the clips. Table 3 illustrates the statistics of Substance Abuse Video Dataset (SAVD) that shows the dataset includes 2060 clips in 4:05:00 hours and classified into five classes. As it can be seen in the table, we collected 394 clips from YouTube and other websites with duration 1:36:47 as well as we collected 665 clips from the 29 selected movies that mentioned in Table 2, with duration 1:04:56 for substance abuse activities (drinking, inhaling, injecting, and smoking). On the other hand, we collected 1001 clips with duration 1:23:17 from other existence datasets for no-drug actions.

V. Experiment

All clips in the SAVD dataset are categorized into five classes: drinking, inhaling, injecting, smoking, and No-drug. In this section, we illustrate the classification model outputs to detect substance abuse actions in the video by using different deep learning architectures. To do this, we utilized TensorFlow and Keras libraries in Python. Also, we ran the models on a laptop by intel core i7 9th Gen, 32GB RAM, and NVIDIA GEFORCE RTX 2060 GPU. Deep learning models are trained by the SAVD dataset, and the performance of the models is compared.

Table 3 Substance Abuse Video Dataset (SAVD) statistics

		Drinking	Inhaling	Injecting	Smoking	No-drug	Total
YouTube and websites	Clips	139	36	60	159	0	394
	Duration	0:29:31	0:06:26	0:20:42	0:40:08	0	1:36:47
29 movies	Clips	158	93	67	347	0	665
	Duration	0:15:01	0:07:32	0:12:06	0:30:17	0	1:04:56
Other datasets	Clips	0	0	0	0	1001	1001
	Duration	0	0	0	0	1:23:17	1:23:17
Total clips		297	129	127	506	1001	2060
Total duration		0:44:32	0:13:58	0:32:48	1:10:25	1:23:17	4:05:00
Min Clip Length		0:01	0:01	0:01	0:01	0:01	-
Max Clip Length		1:01	1:01	4:56	1:38	2:59	-
Frame rate		24					
Resolution		320*240					

[1] Adobe Premiere Pro is a timeline-based video editing software application developed by Adobe Inc

First, we extracted the 40 keyframes from each video clip, labelling them according to their action classes (drinking, inhaling, injecting, smoking, and No-drug), besides the frames are categorized into train and test data. Then, 5000 frames were sampled from each class. We trained and validated the deep learning models by the train data, then evaluated the models by the test data. We ran the SAVD dataset in a CNN model includes 9 layers shown in Fig. 2. Furthermore, we modeled four transfer learning model that have a base model on the ImageNet dataset [21] and transfer the obtained weights to the model. The base models that are used in this research are Xception [22], MobileNet [23], VGG19 [24], ResNet50 [25]. (see Fig. 3)

In the transfer learning models (Xception, MobileNet, VGG19, and ResNet50), feature maps with shape $7 * 7 * 1024$ were utilized, from the base model layers in ImageNet as the models' input. In all models, the vector dimension of memory is set to 1024. The coefficient of weight decay is set to 0.0001 and in all experiments, the optimizer is the Adam algorithm with an initial learning rate of 0.0001.

Table 4 reports our experimental results for 200 epochs in the five models on the SAVD dataset and 5000 frames selected for train and test data. It can be seen in Table 4, ResNet50-Dense model has the best accuracy with 99.38%.

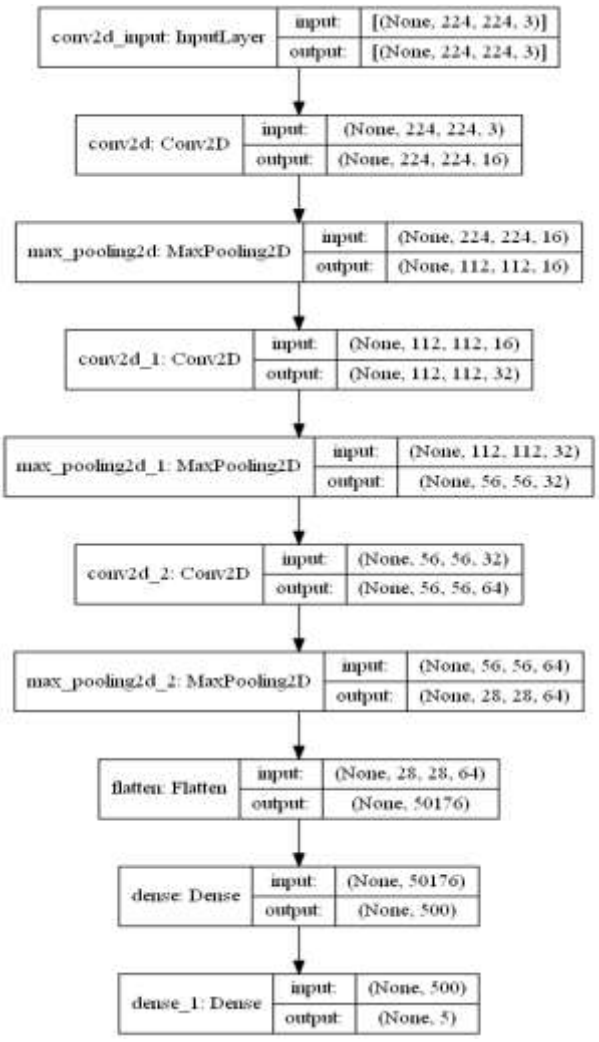


Fig. 2 CNN-9-layer model

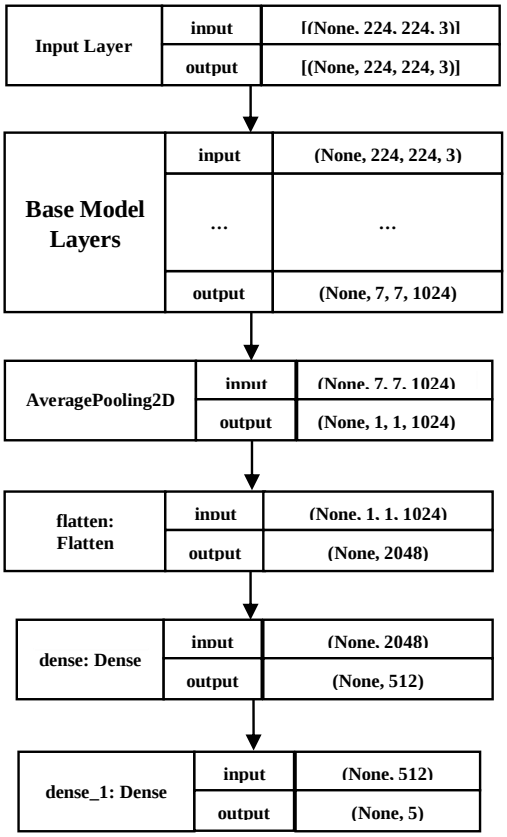


Fig. **3** The transfer learning models
Note: Base Model Layers shows the layers of the base models that in this case are Xception, MobileNet, VGG19, or ResNet50

Table **4** Models' performance comparison on SAVD dataset

Proposed Model	**Accuracy (%)**
CNN-9-layer	98.89%
Xception-Dense	75.00%
MobileNet-Dense	91.93%
VGG19-Dense	97.80%
ResNet50-Dense	**99.38%**

VI. CONCLUSIONS

Automatic video content rating (VCR) as a computer vision subject, monitor the social media clips. Substance abuse, pornography, nudity, violence and profanity languages are the key actions of adult content in the videos. In this paper, we concentrated on substance abuse detection in the videos. We created a substance abuse video dataset (SAVD) extracted from other action recognition dataset, YouTube, and some free stock video footage websites such as Gettyimages.com, Pexels.com, Rnvato.com, istockphoto.com, storyblocks.com and 29 related movies. It includes 2060 clips in 4:05:00 hours and classified into five classes (drinking, inhaling, injecting, and smoking and no-drug). We proposed five deep learning models that four of

them are applied transfer learning from Xception, MobileNet, VGG19, and ResNet50 pretrained models in ImageNet database and a CNN model includes 9 layers. The proposed deep learning models are trained in the SAVD dataset, and as we described in previous section, ResNet50-Dense model gives the best accuracy.

REFERENCES

[1] S. Villani, "Impact of Media on Children and Adolescents: A 10-Year Review of the Research," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 40, no. 4, pp. 392-401, 2001.

[2] Ofcom, "Children and parents: Media use and attitudes report 2018," 29 January 2019 2018.

[3] A. Khaksar Pour, W. Chaw Seng, S. Palaiahnakote, H. Tahaei, and N. B. Anuar, "A survey on video content rating: taxonomy, challenges and open issues," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24121-24145, 2021/07/01 2021, doi: 10.1007/s11042-021-10838-8.

[4] I. Motion Picture Association of America and I. National Association of Theatre Owners, "CLASSIFICATION AND RATING RULES," 2010.

[5] TV Parental Guidelines Monitoring Board. "TV Parental Guidelines." http://www.tvguidelines.org (accessed 03 Feb, 2019).

[6] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, "Pooling in image representation: The visual codeword point of view," *Computer Vision and Image Understanding*, vol. 117, no. 5, pp. 453-465, 2013, doi: 10.1016/j.cviu.2012.09.007.

[7] D. Moreira *et al.*, "Pornography classification: The hidden clues in video space-time," *Forensic Sci Int*, vol. 268, pp. 46-61, Nov 2016, doi: 10.1016/j.forsciint.2016.09.010.

[8] E. B. a. S. Nievas, Oscar Deniz and Garcia, Gloria Bueno and Sukthankar, Rahul, "Hockey Fight Detection Dataset," *Computer Analysis of Images and Patterns*, pp. 332--339, 2011. [Online]. Available: http://visilab.etsii.uclm.es/personas/oscar/FightDetection/. Springer.

[9] C.-H. Demarty, C. Penet, G. Gravier, and M. Soleymani, *A Benchmarking Campaign for the Multimodal Detection of Violent Scenes in Movies*. 2012, pp. 416-425.

[10] C.-H. Demarty, B. Ionescu, Y.-G. Jiang, V. Lam, M. Schedl, and C. Penet, *Benchmarking Violent Scenes Detection in movies*. 2014, pp. 1-6.

[11] C.-H. Demarty, C. Penet, M. Schedl, B. Ionescu, V. Lam, and Y.-G. Jiang, "The MediaEval 2013 Affect Task: Violent Scenes Detection," vol. 1263, 10/17 2013.

[12] C.-H. Demarty, C. Penet, M. Soleymani, and G. Gravier, "VSD, a public dataset for the detection of violent scenes in movies: design, annotation, analysis and evaluation," *Multimedia Tools and Applications*, vol. 74, 05/01 2014, doi: 10.1007/s11042-014-1984-4.

[13] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent Flows: Real-Time Detection of Violent Crowd Behavior," presented at the 3rd IEEE International Workshop on Socially Intelligent Surveillance and Monitoring (SISM) at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Rhode Island, June 2012, 2012. [Online]. Available: www.openu.ac.il/home/hassner/data/violentflows/.

[14] A. Gangwar, E. Fidalgo, E. Alegre, and V. González-Castro, "Pornography and Child Sexual Abuse Detection in Image and Video: A Comparative Evaluation," *8th International Conference on Imaging for Crime Detection and Prevention*, 2017.

[15] P. C. Ribeiro, R. Audigier, and Q. C. Pham, "RIMOC, a feature to discriminate unstructured motions: Application to violence detection for video-surveillance," *Computer Vision and Image Understanding*, vol. 144, pp. 121-143, 2016, doi: 10.1016/j.cviu.2015.11.001.

[16] T. Zhang, W. Jia, C. Gong, J. Sun, and X. Song, "Semi-supervised dictionary learning via local sparse constraints for violence detection," *Pattern Recognition Letters*, vol. 107, pp. 98-104, 2018, doi: 10.1016/j.patrec.2017.08.021.

[17] M. M. Soliman, M. H. Kamal, M. A. E.-M. Nashed, Y. M. Mostafa, B. S. Chawky, and D. Khattab, "Violence Recognition from Videos using Deep Learning Techniques," in *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 8-10 Dec. 2019 2019, pp. 80-85, doi: 10.1109/ICICIS46948.2019.9014714.

[18] United Nations Office on Drugs and Crime (UNODC), "World Drug Report 2021."

[19] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "A Large Video Database for Human Motion Recognition," presented at the Proceedings of the International Conference on Computer Vision (ICCV), 2011.

[20] K. Soomro, A. Roshan Zamir, and M. Shah, " UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild," presented at the CRCV-TR-12-01, November, 2012, 2012.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, *ImageNet: a Large-Scale Hierarchical Image Database*. 2009, pp. 248-255.

[22] F. Chollet, *Xception: Deep Learning with Depthwise Separable Convolutions*. 2017, pp. 1800-1807.

[23] A. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 04/16 2017.

[24] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv 1409.1556*, 09/04 2014.

[25] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*. 2016, pp. 770-778.

A New Sorting Algorithm for Integer Values (Array Sorting Algorithm)

Hesham N. Elmahdy
Information Technology Department Faculty of Computers and Artificial Intelligence, Cairo University,
Giza, Egypt
ehesham@fci-cu.edu.eg

Abstract: Sorting algorithm is an old classic practice. Beginners in programming practice it to test their understanding of programming concepts. This paper presents a new sorting algorithm, which the author coins as Array Sorting Algorithm (ASA). It works only if the elements to be sorted are positive integers. Most database systems use positive integer numbers as primary keys. Other important data fields, such as IDs, bank accounts, and aviation tickets use positive integers. none Comparison-based Sorting. The real challenge when dealing with ASA is the repeated elements to be sorted, which the present study tackles using a two-dimensional array. The second column is used as a counter, in order to show the repetition of each element. The complexity of the founded sorting algorithms was bubble sort, selection sort, and insertion sort algorithms. Their complexity is O(n * log n). ASA's complexity varies between O(n), best-case, and O(2*n), worst case. This depends on whether or not the value of the elements is repeated. Another advantage of ASA is its simplicity in understanding and implementation. The none Comparison-based Sorting algorithms prevail the lower bound of O(n*log n) of the comparison-based sorting algorithms because they do not use comparison sort..

Keywords: Sorting algorithm, bubble sort, selection sort, insertion sort, merge sort, the heap sort, randomized quick sort and quick sort

I. INTRODUCTION

Sorting is an old classic problem of reordering items that may be an array or a list of integers, floating point numbers, or strings. Sorting may be in ascending or descending order. Each sorting algorithm has a variety of strategic ideas in computer science. These strategies are comparison vs. non-comparison, iterative versus recursive, divide and conquer, best/worst/average complexity, and randomized ones. Comparison-based Sorting Algorithms are bubble sort, selection sort, insertion sort, merge sort, quick sort, and random sort. The none Comparison-based Sorting Algorithms are counting sort, and radix sort. The complexity of bubble sort, selection sort, and insertion sort is O (n^2). The complexity of merge sort, heap sort, and quick sort is O (n*log(n)).

The lowest complexities of the founded sorting algorithms are merge sort, heap sort, randomized quick sort, and quick sort. Their complexity is O (n*log(n)). The none Comparison-based Sorting algorithms prevail the lower bound of O(n*log n) of comparison based sorting algorithm because they do not use comparison sort.

N.B.: Complexity is the main factor of running time.

Figure 1 shows the trends of complexity vs. number of elements (retailored from [1]).

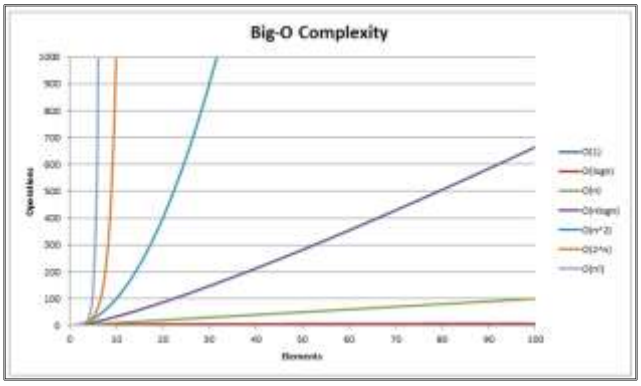


Fig. 1. The trends of complexity vs. number of elements

In this paper, Section II summarizes the literature review and background, Section III discusses the new algorithm, Section IV discusses the experiment evaluation. Section V presents the extension of the new algorithm (Negative Integers, Long Integers, and Text Elements), and finally Section VI offers conclusions and future work.

II- Literature Review and Background:

Zhi-Gang Zhu concluded that it is crucial to master sorting algorithm, as part of data structure courses. The sorting algorithm is used frequently in program design. It is recognized that sorting algorithms is highly significant. Tutors employ computer technology in the process of comparing different sorting difficulties, such as consummate analysis of algorithm indicators, ranking test results, etc., based on the growth of science and technology. [2]. Algorithm complexity is a critical metric when evaluating programs. ASA introduces the lowest complexity of the declared sorting algorithms. Computer Science departments would require their students be trained in implementing sorting algorithms.

Sepahyar et al. demonstrated that time complexity and memory complexity are important considerations for all algorithms, particularly sorting algorithms. Using the appropriate sorting algorithm for data can potentially reduce time and memory usage. The sorting problem receives a lot of attention because efficient sorting is necessary for optimizing other algorithms as well. A sorting algorithm is typically composed of two nested loops that determine the algorithm's complexity. However, other factors, such as the number of data

and data types, also play an important role. As a result, by employing the proper sorting algorithm, time and memory can be efficiently used. In this paper, various sorting algorithms were used for various data types, in order to determine the best use of time and memory for these algorithms. This means that knowing the kind of dataset can facilitate the selection of the most appropriate method. Figure 2 shows the importance of sorting searching and algorithms [3].

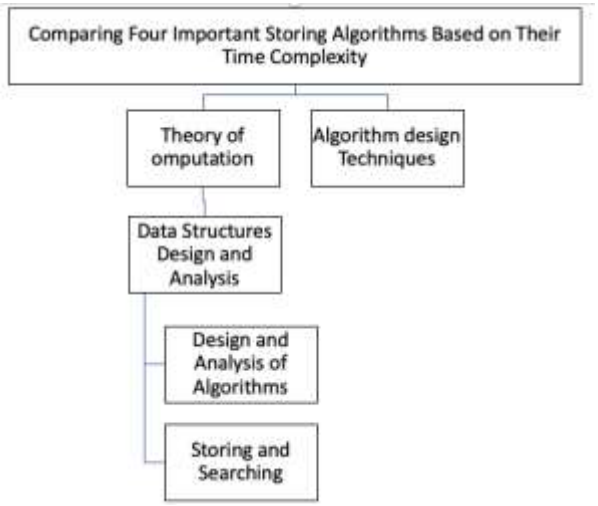


Fig. 2 The importance of sorting searching and algorithms

Dlamini et al. conducted a meta-analytical comparison of the energy efficiency of quick sort and merge sort algorithms, by combining results from the experiments performed on various computing systems and testing, whether there is a significant difference in the energy consumption of both sorting algorithms [4].

As mentioned above, this new algorithm works only on positive integer numbers. An integer number is written without a fractional component; the word is derived from the Latin word “integer,” which means "whole." Integer numbers include positive and negative numbers. For example, integers are 17, 9, 0, and -1024, -16, -1958, whereas 3.1428, 22/7, and $\sqrt{5}$ are not. Positive integers represent main data fields in real life, e.g. National IDs, phone numbers, bank account numbers, aviation numbers, etc. Most database systems use primary keys as positive integer numbers. Figure 3 shows the representation of all numbers (Tailored from [5]). Integers can be compared to discrete, evenly spaced points on an infinite number line.

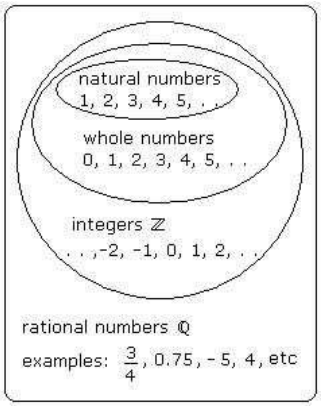


Fig. 3. the representation of all numbers.

Limitations of the existing sorting algorithms were addressed, and a new quadratic sorting algorithm was developed. The present study employed the notion that an unsorted data sequence can be viewed as a collection of disjoint sorted sequences of data items [5].

This section discusses the most familiar sorting algorithms and concludes with a comparison between them.

II.1 Comparison-based Sorting Algorithms:

Sorting Algorithms are bubble sort, selection sort, insertion sort, merge sort, and quick sort.

II.1.1 Bubble Sort Algorithm:

Figure 4 shows an example of the first round of steps of the bubble sort algorithm, for example, starting from the integer numbers 5, 3, 8, 4, 6, 7, 5, 7.

Fig. 4. An example of the Steps of the bubble sort algorithm

The complexity is O (n * log n).

II.1.2 The Selection Algorithm:

Figure 5 shows an example of selection sorting algorithm.

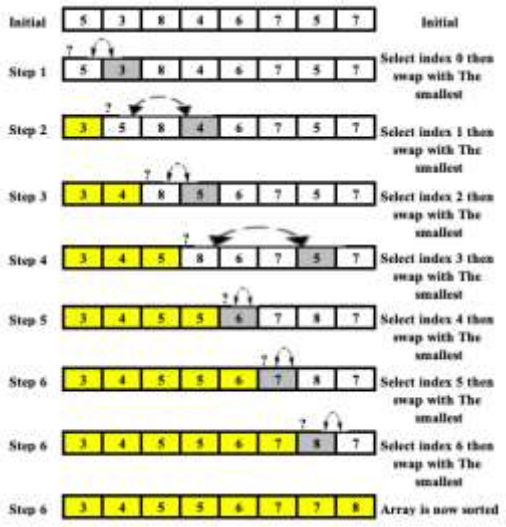


Fig. 5. An example for selection sorting Algorithm

II.1.3 The Insertion Algorithm:

Figure 6 demonstrates an example of insertion sorting algorithm.

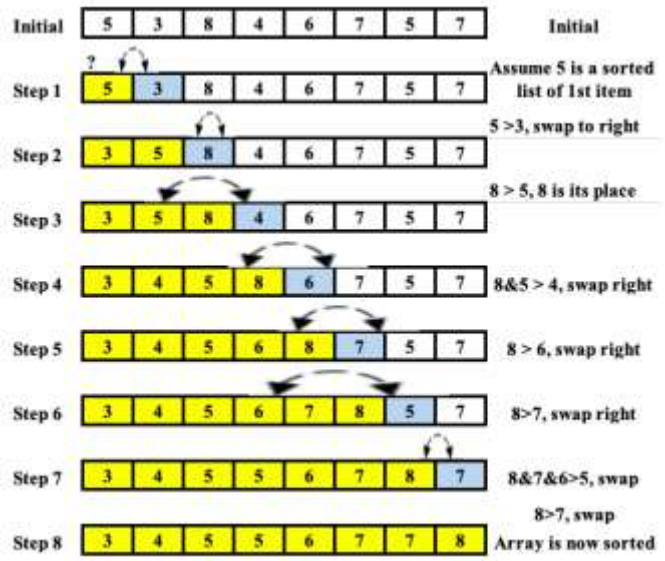


Fig. 6. An example insertion sorting Algorithm

II.1.4 The Merge Algorithm:

Figure 7 shows an example of insertion sorting algorithm.

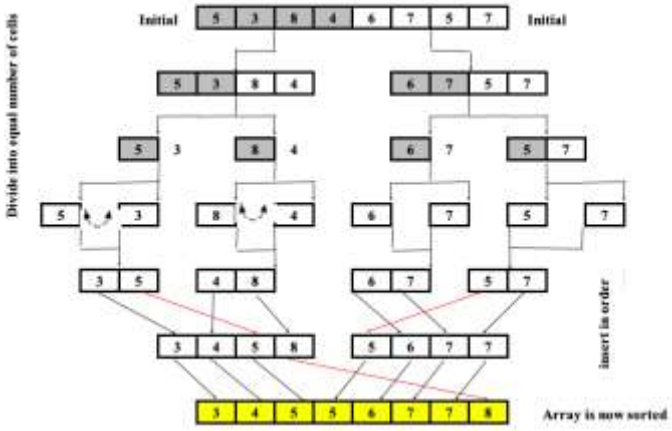


Fig. 7. An example for merge sorting Algorithm

Merge sort showed to be very fast, not losing to its faster competitors. However, this speed might be partially due to the slight optimization used in the study; i.e., the function did not need to allocate the secondary array each time it was called. This was definitely a strong option if the extra memory requirement was not a problem. [6]

II.1.5 The Quick Algorithm:

Quick sort revealed to be undependable. The vanilla version behaved as expected, but the optimized version behaved considerably differently. This highlighted how difficult it was to correctly implement quick sort.[6]

Figure 8 is an example of quick sorting algorithm.

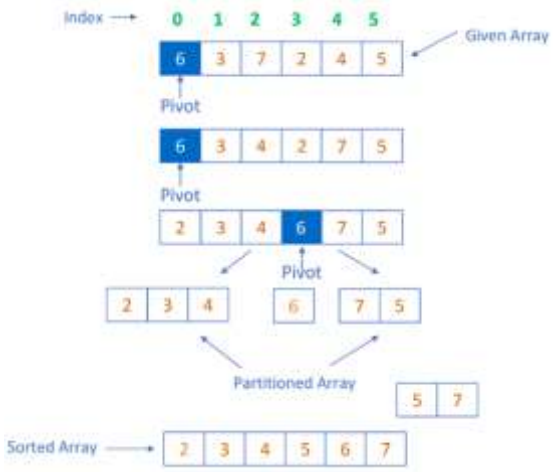


Fig. 8. An example for quick sorting Algorithm

II.2 Algorithms:

The none Comparison-based Sorting Algorithms are counting sort and radix sort.

II.2.1 Counting Sort Algorithm:

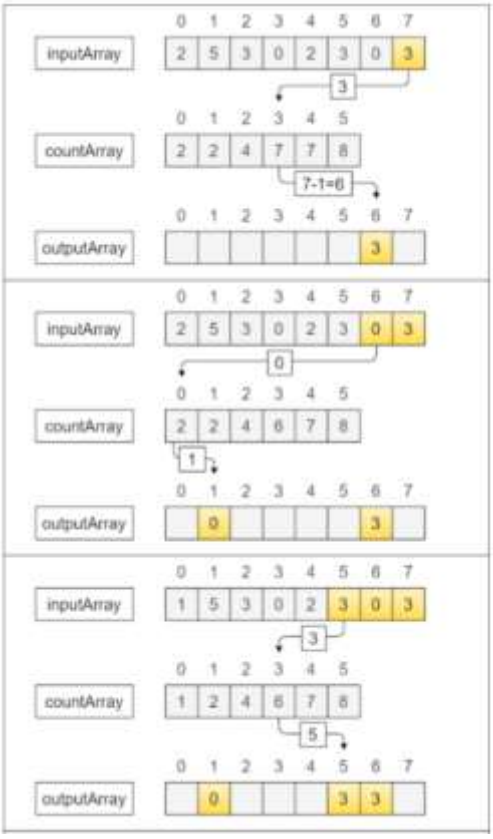


Fig. 9. An example of the sequence of the counting sort algorithm (A).

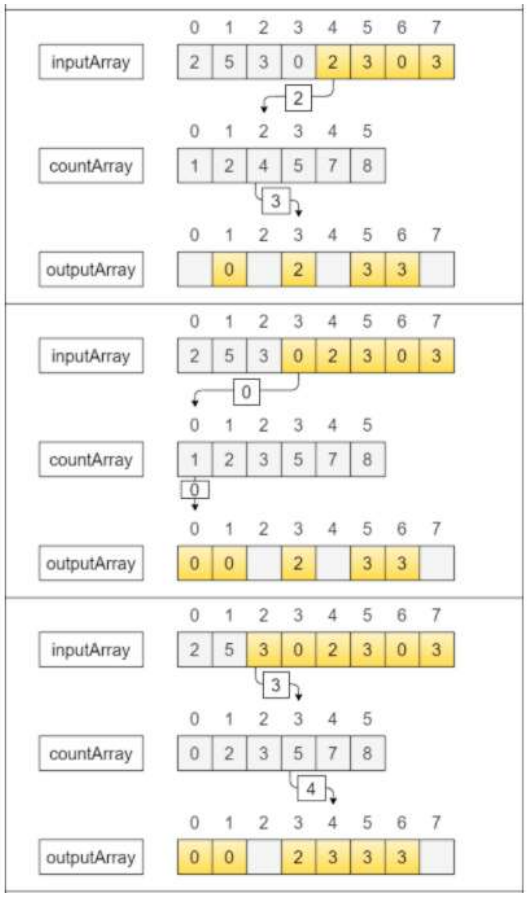


Fig. 9. An example of the sequence of the counting sort algorithm (B).

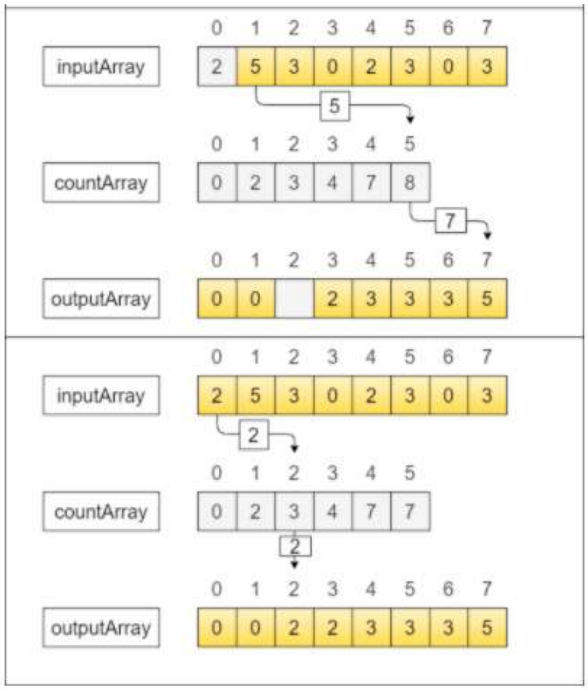


Fig. 9. an example of the sequence of the counting sort algorithm (C).

Vikram Gupta invented Counting sort algorithm, which is a sorting algorithm that sorts an array's elements by counting the number of times each unique element appears in the array. Each unique element's count is stored in an assistant (count) array. Sorting is done by mapping the count, as an index of the assistant (count) array. Figure 9 (A, B, & C) shows an example of the sequence of the counting sort algorithm. Because array indices begin with 0, counting sort only works with positive and negative whole integer [6].

The Counting sort algorithm has a complexity space of O(n). It increases with the increase of the O-range of elements, and vice versa.

II.2.2 Radix Sort Algorithm:

Stehle et al. discussed A Memory Bandwidth-Efficient Hybrid Radix Sort on GPU. Radix sort is a sorting algorithm that groups individual digits of the same place value, before sorting the components. Afterwards, it arranges the components in an ascending or descending order. This process continues until the last significant location is reached. Figure 8 shows an example of the radix sorting algorithm. This algorithm starts with completing all the numbers, so they are three place numbers by adding zeroes to the left of each number, then sorts numbers ascendingly, according to the ones place (least significant digits – ones place). Figure 10 shows an example of the radix sorting algorithm [7].

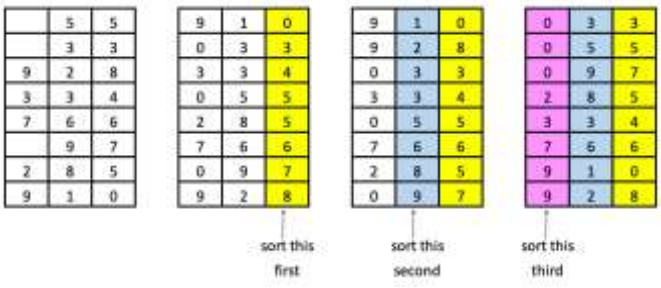


Fig. 10. An example of the radix sorting algorithm.

III- The Array Sorting Algorithm (the New Algorithm)

This paper presents a new sorting algorithm (ASA). This algorithm worked only when the sorted elements were integer numbers. It allocated a two-dimensional array and set the value of each element in the index that equaled the value of that element. The second column of the array, initially set as zero, represented the number of element repetitions. Therefore, the index of each element equaled the value of the element itself. Finding the value zero in the second column's slot indicated that this element did not exist. On the one hand, the value of any element in the second slot meant that the element existed, and it was in its right ascending order. On the other hand, if the value of the second slot was greater than one, this number would indicate the number of repetitions of that element. Hence, these cases should be considered when printing the final sorted output.

Before starting, the maximum value of the elements was checked. The maximum number declared the array size, after incrementing one. The first step was to assign the element value in the corresponding index of the element. The second step was to assign the value to the array element being pointed by that index. The third step was to increment the current

counter to the related column of that element. Upon finishing the assignment of all elements, the elements with a zero counter were removed.

Figure 11 shows an example of the array algorithm.

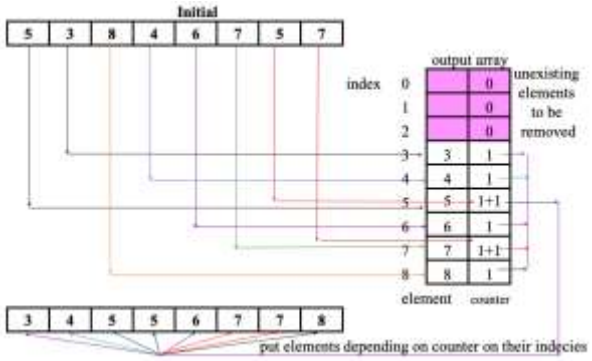


Fig. 11. An example of the ASA.

III.1 Pseudo-code for the ASA

```
1- Find the largest element in the array.
2- Declare a two-dimensional array [Large+1,2]
3- Set all elements of the two-dimensional array
   to be zeroes.
4- Counter=0
5- Input the elements to be sorted.
6- For i inside [Large+1]
        Array[element, counter++]=element
7- Remove element with counter=0
8- Print array
   ========
N.B: in case of non-repeated elements, one was
declared one-dimensional array without counter
```

IV- Evaluation of The Array Sorting Algorithm

ASA's complexity varied between O(n) and O(2*n). This depended on whether or not the value of elements was repeated. Another advantage of ASA was its simplicity in understanding and implementation because it executed three steps for each element. Figure 12 shows the complexity of the presented array sorting algorithm compared to the old ones. The complexity of the array sorting algorithm, represented on the dashed black line, was almost less than one third of the complexity of the best-known ones. As the number of elements to be sorted increased, the performance of ASA improved.

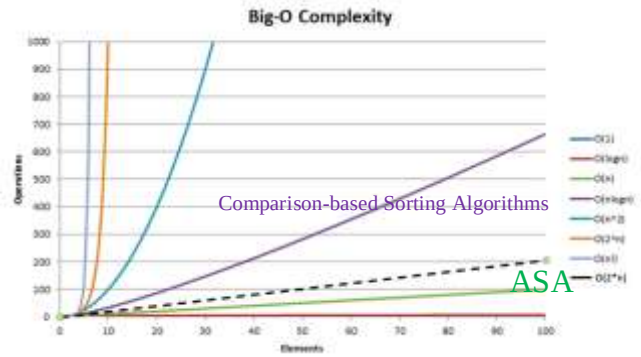


Fig. 12. The complexity of ASA

V. Extension of the ASA:

V.1 Treating Negative Integer Numbers:

The existence of negative elements required separating the negative numbers. Those numbers were converted to be positive, and another array was declared. The new element, such as ASA, was handled, except for setting the value as negative into positive index value. Finally, this array was concatenated before the other ordered positive array.

V.2 Treating Long Integer Numbers:

To improve the space complexity in case of long integers, the smallest element was found, then its value was subtracted from all elements before starting the algorithm. After finishing the algorithm, the value that was subtracted from each element was added in its position.

V.3 Treating Text elements:

Characters as constituent of text would be transferred to its ASCII code. ASCII stands for the American Standard Code for Information Interchange. Since ASCII code consisted of positive integer numbers, ASA was applied.

VI. Conclusions:

This paper presented ASA as a new sorting algorithm. This algorithm was so simple and easy to implement. ASA's complexity varied between O(n), best-case, and O(2*n), worst case. This depended on whether or not the value of the elements was repeated. ASA works with positive integer numbers only. The main motivation behind using positive integer numbers is their existence in most used information. In addition, a solution to negative integer numbers was presented. A road map to solve text elements was also pointed.

VII. Acknowledgement

Author is gratitude to prof. Adel M. Belal, the vice president of the *Arab Academy for Science Technology and Maritime Transport* (AASTMT). Prof. Belal has given the author to be involved in teaching a course in AASTMT. That course was the Introduction to Computers (CS111). Teaching such a course is a great fun, especially when dealing with great current students (most of them). It was a golden chance to return to early knowledge in computer science. Teaching to beginner students is crucial challenge. Thanks to Prof. Mohamed Sharkawy who was the first reader to this paper. His reviews were valuable and were considered. He promised to cooperate on string elements and floating-point elements sorting array.

References:

[1] https://www.hackerearth.com/practice/notes/big-o-cheatsheet-series-data-structures-and-algorithms-with-thier-complexities-1/ Last visit Aug. 29, 2022.

[2] Zhi-Gang Zhu, "Analysis and Research of Sorting Algorithm in Data Structure Based on C Language," Journal of Physics: Conference Series, The Proceedings of the 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP), pp:1-5, 20-22 March 2020, Suzhou, China, June 2nd., 2020

[3] Soheil Sepahyar, Reza Vaziri and Marzieh Rezeal, "Comparing Four Important Sorting Algorithms Based on Their Time Complexity,"ACAI 2019: Proceedings of the 2019 2nd International Conference on

Algorithms, 2019, ISBN: 978-1-4503-7261-9, pp: 320–327, Sanya, China, Dc. 20-22, 2019 https://doi.org/10.1145/3377713.3377808
[4] Gcinizwe Dlamini, Firas Jolha, Zamira Kholmatova and Giancarlo Succi, "Meta-analytical comparison of energy consumed by two sorting algorithms," Information Sciences, Vol 582, pp: 767-777, ISSN 0020-0255, Jan 2022, https://doi.org/10.1016/j.ins.2021.09.061
[5] https://www.pinterest.co.uk/pin/19844054578235813/ last visit Aug. 29, 2022.
[6] https://docs.microsoft.com/en-us/cpp/cpp/integer-limits?view=msvc-170 last visit Aug. 29, 2022.
[7] Vikram Gupta, Visualizing, Designing, and Analyzing the Merge Sort Algorithm: A complete analysis for the Merge Sort Algorithm," https://levelup.gitconnected.com/visualizing-designing-and-analyzing-the-merge-sort-algorithm-cf17e3f0371f last visit Aug. 29, 2022.

General References

[8] Elias Stehle and Hans Arno Jacobsen, "A Memory Bandwidth-Efficient Hybrid Radix Sort on GPUs," The Proceedings of the 2017 ACM International Conference on Management of Data, May 2017 pp: 417–432, Chicago, IL, USA, May 14-19, 2017 https://doi.org/10.1145/3035918.3064043
[9] Jehad Hammad, "A Comparative Study between Various Sorting Algorithms," IJCSNS International Journal of Computer Science and Network Security, VOL.15 No.3, pp: 11-16, March 2015.
[10] Anand Zutshi and Dipanjan Goswami, "Systematic review and exploration of new avenues for sorting algorithm," International Journal of Information Management Data Insights, vol. 1, no. 2, pp: 1- 7, ISSN 2667-0968, Nov. 2021

About the Author:

Prof. Hesham N. Elmahdy is the Ex-vice dean of the Faculty of Computers and Artificial Intelligence (FCAI) Cairo University (CU). Hesham got his B.Sc. in Automobile Engineering in the Military Technical College in 1981 with honor degree with the highest GPA over all graduates in Engineering Faculties in Egypt. He got a diploma in Computer Science and Information Systems in The Institute of Statistical Studies and Researches (ISSR) CU in 1984. He got the first M.Sc. degree in Computer Science (Artificial Intelligence) in ISSR CU in 1992. He got the second M.Sc. in Computer Science in Faculty of Engineering University of Mississippi USA in Aug. 1996. He got his Ph.D. in Computer Science in Faculty of Engineering University of Mississippi USA in Dec. 1997. He was awarded an Associate Professor degree in the department of Information Technology FCAI CU in 2006. He was awarded a Professor Degree in the department of Information Technology FCAI CU in 2011. He was selected as the Chair of the department of Information Technology FCI CU from Nov. 2014 to Jan. 2017. He was designated as the vice dean of the Faculty of FCAI CU for Society Services and Environment Development. He was awarded many national and international prizes and distinguished medals. Hesham was selected as THE BEST INFORMATION TECHNOLOGY PROFESSOR OF AFRICA by THE AFRICA EDUCATION LEADERSHIP AWARDS | 12th DECEMBER 2012 | MAURITIUS. Hesham has been nominated to get “The King Faisal International Prize for Islamic Studies,” 1993. Hesham has been included in the 2006-2007 (9th.) Edition of Who’s Who in Science and Engineering. Hesham has been included in the Outstanding Scientists of the 21st. Century, Cambridge, UK, 2007. Hesham has been nominated to CU Prize in Computer Science in 2007. Hesham has been included in the 2009 (26th.) Edition of Who's Who in The World. Hesham was awarded the prize of “The Best Innovative Ideas to Develop CU” in August 2009. Hesham got the Medal of the Professor of the Year 2011 from CU Club of the Faculty Members. Hesham got the Medal of the Professor of the Years 2011 and 2012 from Cairo University Club of the Faculty Members. In 2018 the Learning News site selected Hesham as one out of a hundred of “Africa’s Movers and Shakers in Corporate Online Learning. In Jan 2019 Hesham was honored the Shield Honor of the Engineering Syndicate as one of the Pioneers of Mechanical Engineer in Egypt. His recent research direction is: IOT, Cloud Computing, Big Data Analytics and eLearning. (ehesham.cu.edu.eg).

Automatic Metal Workpiece Measurement System Using Machine Vision

Haiming Gan
School of Information and Communication Guilin University of Electronic Technology
Guilin, China
ganhaiming@mails.guet.edu.cn

Kun Yan
School of Information and Communication Guilin University of Electronic Technology
Guilin, China
yk5702@guet.edu.cn

Zhi Li
School of Information and Communication Guilin University of Electronic Technology
Guilin, China
19022201022@mails.guet.edu.cn

***Abstract*—Widely used metal workpieces require production inspection to ensure quality. Dimensional measurement is an important part of quality control. By measuring the appearance size of the metal workpiece to determine whether it meets the standard. Traditional methods mainly use physical tools such as rulers, micrometers, and vernier calipers, which have a limited measurement range and low efficiency, and the measurement results are easily affected by subjective factors. With the development of computer science technology and image processing technology, the application potential of machine vision-based recognition and measurement methods has gradually increased. It can realize non-contact, real-time automatic measurement of the measurement target. However, the performance evaluation of machine vision measurement systems for large-scale applications is insufficient. Therefore, this paper takes circular and polygonal metal workpieces as the research objects, designs a high-precision and high-stability visual measurement system, and evaluates the system performance.**

Keywords—geometric measurement, machine vision, automatic inspection

I. INTRODUCTION

In mass production, in order to ensure the quality, it is usually necessary to perform process inspection on the produced metal parts, such as contour integrity, roundness, angle, length, width and other indicators [1]. Traditional methods use measuring equipment or manually to complete multiple measurements, and then take the average or median value as the final measurement result. Due to artificial factors, traditional methods have large errors and are not conducive to mass production [2]. With the continuous update of manufacturing technology and the continuous iteration of computer technology, measurement and detection methods based on machine vision are also booming, and their detection efficiency and measurement accuracy have greater advantages than traditional methods.

In recent years, many scholars have proposed some machine vision measurement methods. Te-Hsiu Sun et al. used a machine vision-based approach to study the measurement of electrical contacts of tiny components. They used a feature screening method to extract the top and bottom contours of the electrical contacts, respectively. Particle swarm optimization was used to improve measurement accuracy. The resulting errors are 0.112 mm and 0.497 mm [3]. Giuseppe Di Leo et al. designed a high-precision automatic measurement system for mechanical parts based on machine vision, with an average error of 0.02 mm [4]. Hong Yun et al. used machine vision measurement technology to measure micro parts, first smoothing was used to remove image noise, and then Canny algorithm was used to locate edges to remove false edge interference. The measurement accuracy is within 20 um [5]. Li Min et al. proposed a measurement method based on Hough line detection and least squares curve fitting for the measurement of black crystal panels, but the accuracy is low [6]. Liu Bin et al. used statistical template information to accurately locate the edge, and established a local measurement coordinate system to measure the size of the screen printing template, but the computational complexity was comparatively higher [7].

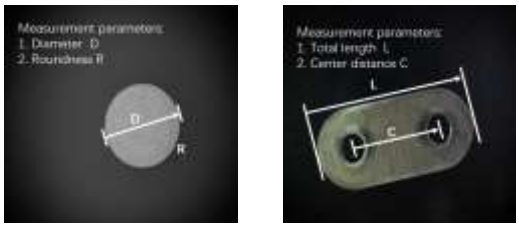
Metal workpieces are used as the research object in this paper to explore a high-precision and high-efficiency measurement method. The measurement steps of metal workpieces based on machine vision are divided into three steps. First, an image acquisition device is built, and industrial cameras are used to take pictures of the target to be measured. Secondly, image preprocessing, feature screening, data analysis and other operations are performed to obtain the required feature parameters. Finally, the automatic measurement results are obtained after calculation. The measurements presented in this paper are tested in large-scale production. Fig. 1 shows the measurement objects and measurement parameters used in this paper.

II. SYSTEM

The structure diagram of our proposed measurement platform for metal workpieces based on machine vision is shown in Fig. 2. It consists of a PC (denoted by (1) in the Fig. 2), a control module (denoted by (2)), an industrial camera (denoted by (3)), a photoelectric sensor (denoted by (4)), a speed collector

This work was supported by (NSFC 62101147) from National Science Foundation of China, Research Enhancement Award from Guangxi Province (2020GXNSFAA159146, AA21077008), (ISN22-10) from State Key Laboratory of Integrated Service Networks and (2021YCXS032) from GUET graduate student innovation program.

(denoted by (5)), a turntable (denoted by (6)), a motor driver (denoted by (7)), and a reject device (denoted by (8)).



(a) Workpiece 1 (b) Workpiece 2

Fig. 1. Two types of detection objects.

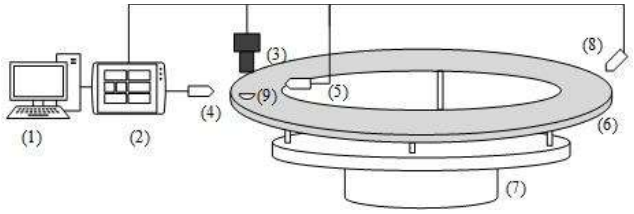


Fig. 2. Platform structure of the measurement system.

A control module is designed to serve as the core of the whole system. Both a laser beam sensor and a tachometer are connected to the control module. First, the laser beam sensor detects whether the workpiece is transferred to a turntable. The tachometer is used to collect the speed of the turntable, which is controlled by a motor driver. Using the speed of the turntable, the workpiece can be accurately positioned and the operation of the entire platform can be stabilized. When the laser beam sensor captures the workpiece to be inspected on the turntable, it converts the optical signal into an electrical trigger signal and transmits it to the control module. The trigger signal will be processed by the control module to drive the industrial camera to take photos. Then, the photo of the workpiece to be inspected will be fed to an industrial computer. After receiving the image, the industrial computer uses the measurement algorithm to process the image, and determine whether the workpiece is defective according to the product quality index. The processing result will be sent to the control module, which will drive the rejection device to sort the products.

A. Control Modul

The structure diagram of the control module is shown in Fig. 3. During initialization, the control module monitors the speed of the turntable and controls the turntable motor driver to stabilize the state and speed of the turntable. After initialization, the laser beam sensor detects the presence of the workpieces. When a workpiece is present, the laser beam sensor can capture the variation of the laser signal. The optical signal is converted into an electrical trigger signal, which is relayed by the control module to trigger the camera. The camera takes pictures and sends the pictures to the industrial computer. After receiving the images, the measurement processing algorithm embedded in the industrial computer begins to process the images. The processing results are returned to the control module. The control module controls the rejecting device to sort qualified and defective workpieces.

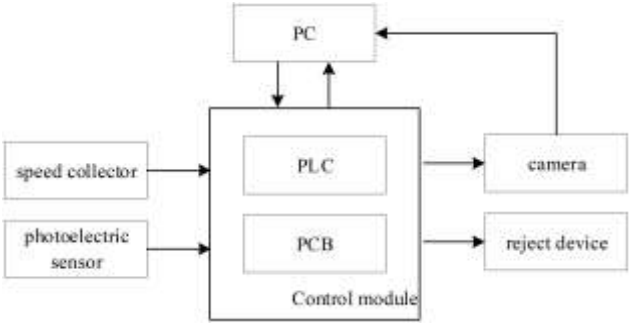


Fig. 3. The block diagram of the control module.

B. Image Processing

The overall flow of our proposed machine vision measurement algorithm for metal workpieces is shown in Fig. 4. The process of the measurement algorithm can be divided into the following five steps: 1) Image acquisition: The image of the target workpiece is captured by image acquisition equipment (light source, camera, lens, etc.). 2) Image normalization: Remove background information from the captured images to normalize the variation of background light. Determine the invalid images. Note that the complete workpiece contour has to be captured in a valid image. 3) Image preprocessing: Remove noise and other interferences using filter and smoother. Then image enhancement is used to highlight the details of the workpiece. Finally the segmentation method is used to separate the target from the background. 4) Edge detection: The size measurement needs to extract image information and features, mainly the contour and shape of the workpiece. 5) Size measurement: Calculate the physical size parameters of the workpieces using the workpiece contour.

III. AUTOMATIC MEASUREMENT ALGORITHM

In this paper, the measurement object is an irregularly shaped metal workpiece, which has an outer edge and an inner edge. Since image-based measurement algorithms rely on precise edge extraction, how to design a stable and reliable algorithm to extract target edges is a crucial part in our system. Image edge contours are grayscale values that change rapidly (quickly abruptly) [8] [9]. Since the workpiece is photographed in motion, the edge of the workpiece is blurred than the edge of the workpiece photographed in a stationary state. In addition, due to the operation of the mechanical platform, the position of the workpiece on the turntable may change slightly, resulting in out-of-focus and blurred edges. The above reasons cause the difficulty of edge extraction, and it is necessary to eliminate the influence caused by blurring.

The traditional Canny edge detection algorithm determines edge points through fixed high and low double thresholds, which is difficult to adapt the changes in illumination [10] [11]. In order to improve the robustness of our proposed measurement algorithm, a method using caliper tool and improved Canny algorithm is proposed. The flow chart of the improved measurement method using caliper tool is shown in Fig. 5.

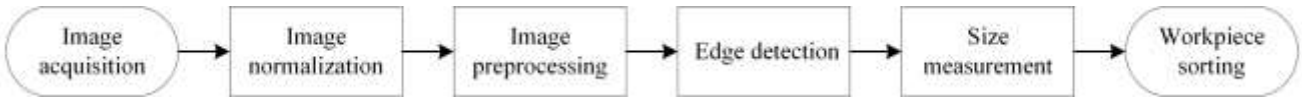


Fig. 4. Flow chart of the machine vision measurement algorithm.

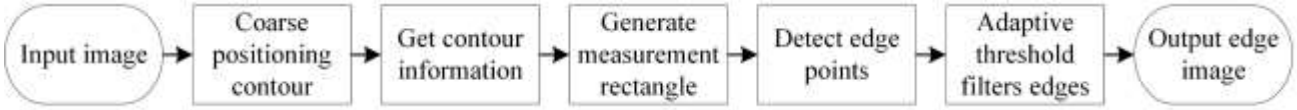


Fig. 5. Flow chart of the improved measurement method using caliper tool

The advantage of the caliper tool is that, by generating measurement rectangles of the same size and distance, the edge points with the largest gradient amplitude perpendicular to the rectangles can be detected, and the optimal edge points are obtained in accordance to the set of measurement rectangles. By combining all detected edge points, more accurate size measurements can be obtained, reducing traversal time and improving efficiency. Fig. 6 is a schematic diagram of the caliper tool method. The rectangle is a measuring caliper tool, and each rectangle generates an optimal edge point Mn in turn according to the detection steps.

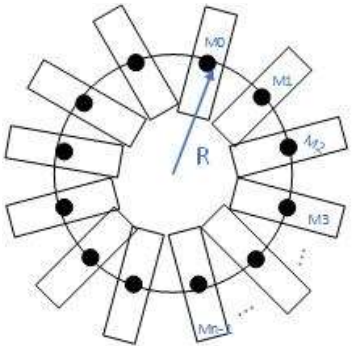


Fig. 6. Schematic diagram of the caliper tool method.

The parameters in the caliper tool, such as the size and number (interval) of the measurement rectangles are crucial for the measurement accuracy and efficiency [12]. Thus, they have to be selected carefully. We determine the optimal parameters after experimental research. The area size of the rectangle is chosen as 10×5 pixels. The interval between the two rectangles next to each other is selected as twenty pixels.

IV. Experiment Results

The configuration environment for the experiment: Windows10 operating system, CPU is Intel Core i5-10200H, clocked at 2.40 GHz, memory 8 GB, GPU is GRX 1650, pytorch1.2 deep learning framework. The Hikvision industrial camera (MV-CA050-10GM) are used to capture the images. Telecentric fixed focal lens are used. The field of view (FOV) is 12.8 mm×9.6 mm.

The experimental measurement and verification of two kinds of workpieces are presented here. In order to verify the stability and measurement accuracy of the algorithm, two kinds of workpieces are employed in our experiment. In the following part, two scenarios are considered to evaluate the repeatability and robustness, such as 1) measurement of a same workpiece repeatedly. 2) measurement of multiple workpieces.

A. *Scenario 1: Repeated measurement of the same workpiece*

This scenario is designed to test the stability of our proposed measurement algorithm. For each workpiece, ten images are randomly selected and the measurement algorithm is used to measure the size of the given workpieces repeatedly using the same industrial computer. The measurement parameters of workpiece 1 are the circle diameter (D) and roundness (R). The measurement parameters of workpiece 2 are the length of the workpiece (L) and the distance between the two circles' centers (C). The actual value of the size of the workpieces are used as the benchmark, which are shown in TABLE I.

TABLE I. ACTUAL VALUES OF THE SIZE OF THE TWO WORKPIECES.

	Workpiece 1		**Workpiece 2**	
Parameters	***D(mm)***	***R***	***L(mm)***	***C(mm)***
True Value	9.75	1	36	20

The statistics of the measured parameters are shown in TABLE II. The experiment results indicate that the measurement error using our proposed algorithm is less than 0.15 mm.

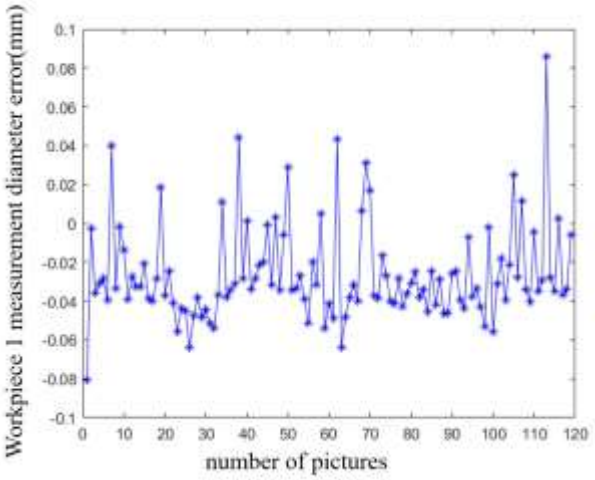
TABLE II. MEASUREMENT RESULTS AND ERRORS OF TWO WORKPIECES

	Workpiece 1		**Workpiece 2**	
Measurement parameters	***D(mm)***	***R***	***L(mm)***	***C(mm)***
Measurement results	9.752	0.978	35.922	19.88
Measurement error	+0.002	-0.022	-0.078	-0.12

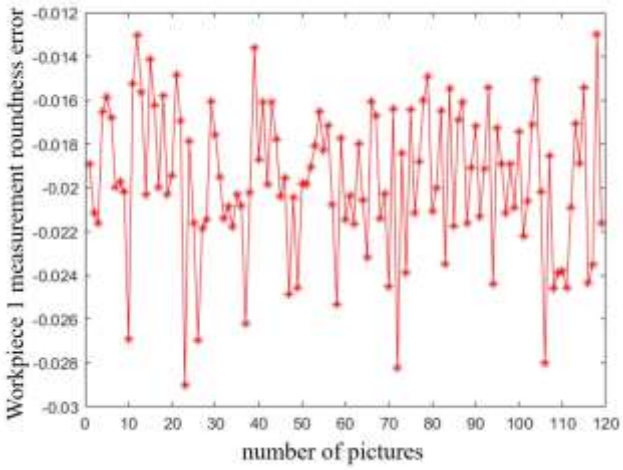
B. *Scenario 2: Test 120 images for two workpieces respectively*

In this scenario, for each type of workpiece, 2000 samples are randomly selected to collect images. Three images are taken for one sample. All the images are employed to evaluate our proposed measurement algorithm. Some experimental results are shown in Fig. 7.

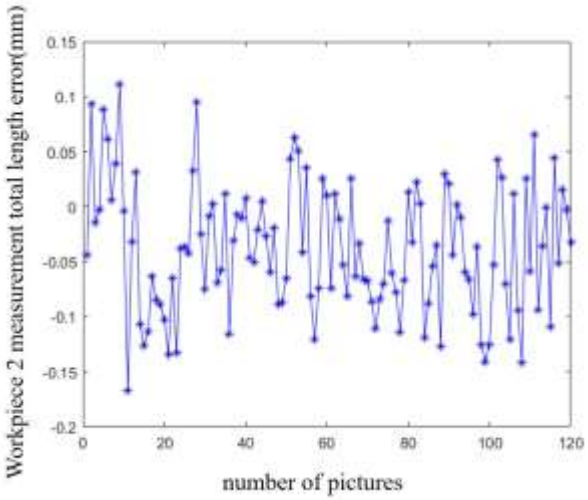
The maximum, minimum and average error of the two workpieces are shown in TABLE III.



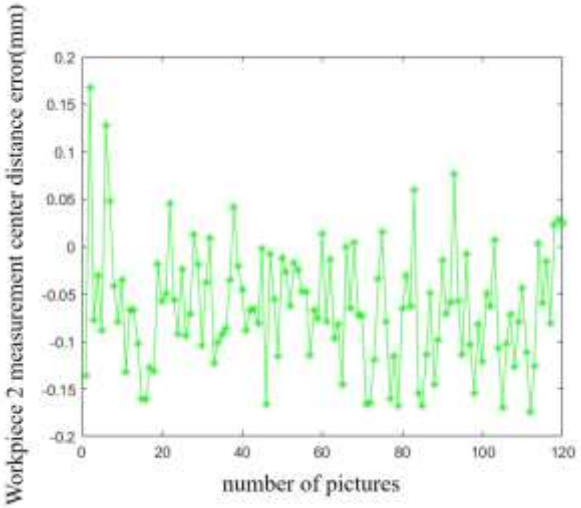
(a) Measurement error of D of workpiece 1.



(b) Measurement error of R of workpiece 1.



(c) Measurement error of L of workpiece 2



(d) Measurement error of C of workpiece 2

Fig. 7. Illustration of the measurement error of 120 samples for each of the two kinds workpieces.

TABLE III. MEASUREMENT RESULTS AND ERRORS OF TWO WORKPIECES.

	Workpiece 1		**Workpiece 2**	
Measurement parameters	*D*(mm)	*R*	*L*(mm)	*C*(mm)
Minimum error	-0.06	-0.029	-0.15	-0.17
Maximum error	+0.045	-0.013	+0.11	+0.17
Average error	+0.030	-0.020	+0.045	-0.012

The measurement results show that our proposed algorithm has higher measurement accuracy for circular diameters and regular-shaped objects. The average error of the diameter measurement of the workpiece 1 is within 0.03 mm, and the roundness measurement error is within 0.03. Since workpiece 2 is a complex polygon, there are more parameters to be measured than workpiece 1, and its size is larger than that of workpiece 1, so the measurement is more difficult. Therefore, the measurement error is also larger. The average error of straight line measurement is within 0.045 mm.

V. CONCLUSION

A metal workpiece dimension measurement algorithm is proposed. Based on the control module and our proposed algorithm, an intelligent sorting system for metal products is constructed. Using this system, subjective errors caused by manual measurements can be avoided. Non-contact, stable, high-precision automatic measurement is realized. The experimental results show that the accuracy of the system can reach 10^{-3} mm. The application of this system can improve the yield rate of products and has high application value to the production of enterprises.

REFERENCES

[1] Z. Y. Duan, N. Wang, J. S. Fu, W. H. Zhao, B. Q. Duan and J. G. Zhao, "High precision edge detection algorithm for mechanical parts," Measurement Science Review, vol.18, no.2, 2018, pp.65-71.

[2] Y. K. Dou, "Research on geometric shape detection of mechanical parts based on machine vision," Master thesis, Lanzhou University of Technology, 2018.

[3] T. H. Sun, C. C. Tseng and M. S. Chen, “Electric contacts inspection using machine vision,” Image and Vision Computing, vol. 28, no. 6, 2010, pp. 890-901.

[4] G. D. Leo, C. Liguori, A. Pietrosanto and P. Sommella, “A vision system for the online quality monitoring of industrial manufacturing,” Optics and Lasers in Engineering, vol. 89, feb, 2017, pp. 162-168.

[5] Y. Hong, “Research on measurement method of part plane geometric dimension based on computer vision,” Master thesis, Donghua University, 2017.

[6] M. Li, Y. T. Zhou, Z. W. Zhang and Y. J. Fan, ”Method for measuring geometric parameter measurement of black crystal panel based on machine vision,”Instrumentation Technology and Sensors, no.5, 2020, pp. 102-106.

[7] B. Liu, Z. T. Dong, C. H. Hu, P. H. Li and M. K. Gao, ”Measurement method of screen printing template size based on machine vision,” Journal of Metrology, vol. 42, no. 2, 2021, pp. 150-156.

[8] G Dougherty, ”Effect of sub-pixel misregistration on the determination of the point spread function of a CT imaging system,” Medical Engi- neering and Physics, vol. 22, no. 7, 2000, pp. 503-507.

[9] S. J. Tong, M. Jiang, and C. J. Jiao, ”Research on an improved edge detection method of workpiece,” Journal of Electronic Measurement and Instrumentation, vol. 35, no. 1, 2021, pp. 128-134.

[10] J. Kim, and S. Lee, ”Extracting Major Lines by Recruiting Zero-Threshold Canny Edge Links along Sobel Highlights,” IEEE Signal Processing Letters, vol. 22, no. 10, 2015, pp. 1689-1692.

[11] W. Xu, Q. Zhang, X. D. Wang, H. Gao, and H. R. Qin, ”Image edge detection method based on improved Canny operator,” Laser Magazine, vol. 43, no. 4, 2022, pp. 103-108.

[12] L. S. Wu, J. M. Xiang, and Y. Hu, ”Real-time Nuts Detection System Based on Machine Vision Caliper Tool Method,” Instrumentation Technology and Sensors, no. 2, 2020, pp. 50-55.

Optimal Stochastic Day-Ahead Power Management of Hybrid AC-DC Microgrids

Mahshid Javidsharifi
AAU Energy
Aalborg University
Aalborg, Denmark
mja@energy.aau.dk

Hamoun Pourroshanfekr Arabani
Division of Industrial Electrical Engineering and Automation
Lund University
Lund, Sweden
hamoun.pourroshanfekr_arabani@iea.lth.se

Tamas Kerekes
AAU Energy
Aalborg University
Aalborg, Denmark
tak@energy.aau.dk

Dezso Sera
Faculty of Science and Engineering
Queensland University of Technology
Brisbane, Australia
dezso.sera@qut.edu.au

Josep M. Guerrero
AAU Energy
Aalborg University
Aalborg, Denmark
joz@energy.aau.dk

***Abstract*— Due to the reappearance of DC loads in electrical systems and advanced improvement in energy storage systems (batteries) and environment-friendly properties of photovoltaics as a green energy supply, DC architecture is considered as a new solution for next-generation power distribution systems. Hybrid AC-DC microgrids (MG) can take advantage of DC and AC flows in a smart distribution system. The best strategy for the optimal operation of hybrid MGs is to minimize the converting energy between AC and DC sides such that DC loads are provided by photovoltaics, fuel cells, and the stored energy in batteries and AC loads are satisfied by AC-based sources including wind turbines (WTs) and diesel generators (DEs). Accordingly, this paper aims to scrutinize an optimal green power management strategy for hybrid AC-DC MGs from an economic viewpoint while considering photovoltaics as a prior source for the DC side and wind turbines for the AC side. Moreover, the uncertainties of renewable energy sources (RESs), DC and AC loads, and the correlation among them are investigated using the unscented transformation method.**

Keywords—Hybrid AC-DC Microgrids, Optimal Operation, Power Management, Uncertainties, Unscented Transformation

I. INTRODUCTION

The combination of distributed generators (DGs) to integrate renewable energy sources (RESs) into local distribution systems besides the advantages of DC power including decreasing the power loss in transmission lines and better controlling of power flow led to the reappearance of DC power [1, 2].

AC systems have some benefits due to the innate characteristics of AC appliances and the presence of transformers to transmit power over far areas to afford AC loads, however, gradual and permanent changes in the type of loads and DGs in AC distribution systems led to the combination of DC networks to the current AC networks [1, 2]. Even though most grids function in AC mode, the large penetration of distributed DC generations, energy storage units, and loads, along with other features, has necessitated the creation of DC distribution networks [3].

The main advantage of these networks is high efficiency due to lower power electronic interfaces, which leads to no flow of reactive power. In addition, it is not required to synchronize the DGs [3]. This configuration needs a lot of modifications to the current power grid and thus raises costs [1-3]. Although DC microgrids (MGs) have many advantages over AC MGs, this technology has not yet been fully adapted to seriously change existing systems. Since AC systems are more dominant, it is more likely to combine AC and DC MGs to solve existing problems efficiently [1-3].

Hybrid MGs which benefit from both AC and DC MGs ease the combination of DC technologies to the current AC systems. By using hybrid AC-DC MGs, DC power supplies are connected to DC loads while AC power sources supply AC loads and the bidirectional converter (BDC) shares power between these two sides [1]. A supervisory controller is needed for dividing the power among different sources. This led researchers to create power management systems [2, 3]. Accordingly, meeting the required power while maximizing the use of RESs, minimizing the use of fuel-based generators, increasing battery life, and limiting the use of the main power converter between AC and DC MGs are considered the main aspects [2]. The hybrid AC-DC MG configuration attracted much attention due to the simultaneous integration of the advantages of AC and DC structures. The main characteristic of this configuration is the integration of both AC and DC networks into the same distribution network, which helps the straight combination of AC and DC distributed loads, storage units, and generating units. This feature provides a convenient way to incorporate future renewable sources or electric vehicles with minimal modifications to the current distribution network and cost reduction [5].

Hybrid MGs consist of AC and DC networks and the BDC between these two networks, which helps the power flow between these two networks and the power grid. These arrangements have many benefits because AC and DC-powered appliances can easily be linked to the grid with fewer power electronic interfaces [2].

This research was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 812991. J. M. Guerrero was supported by VILLUM FONDEN under the VILLUM Investigator Grant (no. 25920).

The concept of hybrid MG as an efficient, economic, and environmentally friendly distribution network for the future has been investigated from several viewpoints [6-16]. In [42], a study on hybrid MG was done. This paper formulated a multi-objective optimization problem for AC-DC hybrid MG operation that minimized energy costs and pollutants.

In [6, 7], an algorithm was utilized for power transmission between AC and DC sides. It was shown that a lack of a proper control strategy in the system can lead to the shutdown of the entire system.

The dynamic evaluation of the AC-DC hybrid system was done in [8]. Authors in [9] implemented a droop-based controller to satisfy the constraints of the DC bus voltage. The studied system included a WT and several controllable loads without controllable generators, such as diesel generator (DE) units, fuel cells (FC), etc. In addition, reference [9] did not investigate the operation of off-grid MGs in different contingencies.

Authors in [13] studied the optimal utilization of hybrid MGs assuming a 24-hour time-dependent effects of the network. This article investigated two different structures (AC MG and hybrid MG) with the same production and local consumption. In [14, 15], issues related to power sharing in a hybrid MG were studied. These papers did not consider the day-ahead planning under different connection states (connected to or disconnected from the network) or the interface converter problem. In [16] the effects of connection inefficiencies in day-ahead scheduling for a hybrid MG were investigated.

The major goal of our current paper is to study the performance of a hybrid MG in the presence of uncertainties in load, and the output power of RESs. Main contributions of the paper can be mentioned as:

1. Optimal power management of hybrid MGs while taking into account the uncertainties of AC and DC loads and output power of RESs.
2. Solving the probabilistic optimal power management problem using a hybrid method.

II. Problem Formulation

A hybrid MG is illustrated in Fig.1. As is observed, the hybrid MG consists of an AC side that includes DE units, wind turbines (WT), and AC loads. The DC side includes FC, photovoltaic (PV) panels, battery storage devices (Batt), and DC loads. The DC and AC sides are connected via a bidirectional converter (BDC) which deals with the power-sharing between DC and AC sides.

A. Objective Functions

The objective is to demonstrate a daily schedule of units in the hybrid MG while considering the related constraints. It is considered that the studied AC-DC MG consists of one DE unit, one FC unit, PV and WT units, a battery, and a BDC. Moreover, the AC-DC MG can exchange power between AC and DC sides. The objective function is to minimize the total cost as follows:

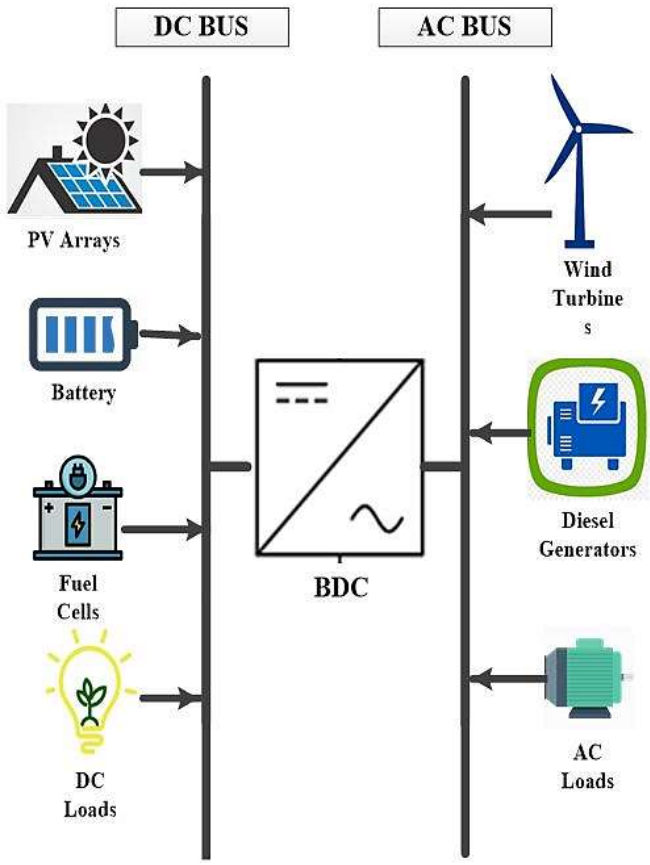


Fig. 1. Hybrid microgrid configuration.

$$min\left\{\sum_{t=1}^{T}\{DE\_cost_t + FC\_cost_t\}\right\} \quad (1)$$

$$DE\_cost_t = a_{DE}(P_{DE,t})^2 + b_{DE}P_{DE,t} + c_{DE} + SUC_{DE} \times V_{DE,t} \times (1 - V_{DE,t-1}) \quad (2)$$

$$FC\_cost_t = (C_{NG}P_{FC,t}/Q_{LHV}\eta_{FC}) + SUC_{FC} \times V_{FC,t} \times (1 - V_{FC,t-1}) \quad (3)$$

where t and T represent a time slot and the perspective of study, respectively. DE_cost_t and FC_cost_t are respectively the operational and start-up costs of DE and FC. a_{DE}, b_{DE} and c_{DE} are the coefficient of the cost function of DE. $P_{DE,t}$ is the output power of DE unit in time interval t. SUC_{DE} is the start-up cost of DE in time interval t. $V_{DE,t}$ is 0 or 1 to show the state of DE being on or off in time t. C_{NG} is the consumption cost of each meter cube of natural gas, $P_{FC,t}$ is the output power of FC in each time interval t, Q_{LHV} is the low heating value of natural gas in kW/m^3. SUC_{FC} is the start-up cost of FC in time interval t and $V_{FC,t}$ is 0 or 1 to show the on or off state of FC in time t.

B. Technical Constraints

1. The constraint of the rate of production changes

The following equations represent the change rate of output powers of DE and FC and the change of power of the BDC.

$$P_{DE,t+1} - P_{DE,t} \le R_{DE}^{up} \quad (4)$$

$$P_{DE,t} - P_{DE,t+1} \le R_{DE}^{down} \quad (5)$$

$$P_{FC,t+1} - P_{FC,t} \le R_{FC}^{up} \quad (6)$$

$$P_{FC,t} - P_{FC,t+1} \le R_{FC}^{down} \quad (7)$$

$$P_{BDC,t+1} - P_{BDC,t} \le R_{BDC}^{up} \quad (8)$$

$$P_{BDC,t} - P_{BDC,t+1} \le R_{BDC}^{down} \quad (9)$$

where $P_{BDC,t}$ represents the exchanged power of bidirectional converter in time interval t. R_{DE}^{up}, R_{DE}^{down}, R_{FC}^{up}, R_{FC}^{down}, R_{BDC}^{up} and R_{BDC}^{down} are respectively the upper and lower band of the changes of power of DE and FC units and bidirectional converter.

2. The constraint of the minimum on/off time of units

The minimum time that each unit should maintain its on/off status to be able to change its status can be demonstrated as the following (each unit that changes its status from off to on/ on to off, should continuously remain in the circuit for a minimum time for technical considerations. Moreover, whenever a unit turns off it should continuously remain off for a certain minimum time.)

$$V_{DE,t} - V_{DE,t-1} \le V_{DE,\tau}$$
$$\forall \tau \in [t+1, min(t + MUT_{DE}, T)] \; \forall t \in [2, T] \quad (10)$$

$$V_{DE,t-1} - V_{DE,t} \le 1 - V_{DE,\tau}$$
$$\forall \tau \in [t+1, min(t + MDT_{DE}, T)] \; \forall t \in [2, T] \quad (11)$$

$$V_{FC,t} - V_{FC,t-1} \le V_{FC,\tau}$$
$$\forall \tau \in [t+1, min(t + MUT_{FC}, T)] \; \forall t \in [2, T] \quad (12)$$

$$V_{FC,t-1} - V_{FC,t} \le 1 - V_{FC,\tau}$$
$$\forall \tau \in [t+1, min(t + MDT_{FC}, T)] \forall t \in [2, T] \quad (13)$$

where $MUT_{DE}, MDT_{DE}, MUT_{FC}$ and MDT_{FC} are respectively the minimum up and down time of DE and FC units.

3. Battery limits

The constraints of the battery charging/ discharging process include charging/discharging limits and up and down rates of the stored energy in the battery as follows [17]:

$$P_{Batt,Min}^{Ch} \le P_{Batt,t}^{Ch} \le P_{Batt,Max}^{Ch} \quad (14)$$

$$P_{Batt,Min}^{Dch} \le P_{Batt,t}^{Dch} \le P_{Batt,Max}^{Dch} \quad (15)$$

$$E_{Batt,Min} \le E_{Batt,t} \le E_{Batt,Max} \quad (16)$$

$$E_{Batt,t} = E_{Batt,t-1} + P_{Batt,t}^{Ch} \eta_{Batt}^{Ch} - \left(P_{Batt,t}^{Dch} / \eta_{Batt}^{Dch}\right) \quad (17)$$

where $P_{Batt,t}^{Dch}$ and $P_{Batt,t}^{Ch}$ represent the discharging and charging rates of the battery in *t*, $P_{Batt,Min}^{Ch}$ and $P_{Batt,Max}^{Ch}$ are the upper and lower bounds of battery charging rate and $P_{Batt,Min}^{Dch}$ and $P_{Batt,Max}^{Dch}$ are the upper and lower bounds of battery discharging rate in *t*. $E_{Batt,t}$ is the stored energy in the battery, $E_{Batt,Min}$ and $E_{Batt,Max}$ are the lower and upper bounds of the stored energy in the battery. η_{Batt}^{Dch} and η_{Batt}^{Ch} are respectively the battery discharging and charging efficiencies.

4. Power balance constraints in each AC and DC side

This constraint implies that the overall power of AC units should satisfy the demanded AC load while the overall the DC power satisfies the demanded DC load.

$$P_{DE,t} + P_{WT,t} + P_{BDC,t} = Load_{AC,t} \quad (18)$$

$$P_{FC,t} + P_{Batt,t} + P_{PV,t} + P_{BDC,t} = Load_{DC,t}$$
$$t = 1,2,\ldots,T \quad (19)$$

where $P_{WT,t}$ and $P_{PV,t}$ are the output powers of WT and PV units, $Load_{AC,t}$ and $Load_{DC,t}$ are the demanded electrical load of each DC and AC sides of the hybrid MG in time intervals *t*.

Moreover, the output power of DE and FC and the exchanged power of the bidirectional converter should fulfill the following:

$$P_{DE,Min} \le P_{DE,t} \le P_{DE,Max} \quad (20)$$

$$P_{FC,Min} \le P_{FC,t} \le P_{FC,Max} \quad (21)$$

$$-P_{BDC,Max} \le P_{BDC,t} \le P_{BDC,Max} \quad (22)$$

$P_{DE,Min}$, $P_{DE,Max}$, $P_{FC,Min}$, and $P_{FC,Max}$ are the lower and upper bounds of the output powers of DE and FC units. $P_{BDC,Max}$ represents the maximum exchangeable power between AC and DC sides of the hybrid MG.

III. PSO-UT Algorithm

To deal with the considered problem particle swarm optimization (PSO) algorithm is implemented. The efficiency of the PSO algorithm in solving optimization problems from stability, and accuracy viewpoints as well as its simple application and formulation are justified in the literature [18, 19].

In this paper, the uncertainties of demanded load and the inherent uncertainties of renewable energies are also considered. Consequently, the unscented transformation (UT) method as an efficient approach is applied in this paper to deal with the probabilistic nature of the considered optimal operation of hybrid MGs. UT is a suggested and widely used approach which is proved to be faster than while it is approximately as accurate as Monte-Carlo [20]. The detailed formulation of UT was studied in [19].

The probabilistic problem of optimal power management of AC-DC hybrid MGs is solved using the proposed PSO-UT approach. The uncertain variables including demanded loads, wind speed and solar irradiation are modeled by the UT. Afterward, the PSO algorithm is used for minimizing the cost while the constraints are satisfied.

IV. Simulation Results

The results of the optimal day-ahead power management of hybrid MGs in deterministic and probabilistic scenarios are presented and the planning of units is done such that the considered cost objective function is minimized. In the deterministic analysis, the solar irradiation, wind speed, and the demanded load in the DC and AC sides of the hybrid MG are considered without uncertainty and the problem is solved using

PSO with a population size of 50 and the number of maximum iterations equal to 200 in MATLAB. Afterward, to consider the uncertainties of the power management of hybrid MGs, UT is integrated with the PSO algorithm, and the probabilistic problem is solved.

Tables I to IV, respectively show the parameters of the DE, FC, battery, and bidirectional converter. The initial charge of batteries, the rated power of PV units, and the rated power of WTs are respectively 150 kWh, 200 kW and 150 kW.

A. Optimal Day-ahead Power Management of AC-DC Hybrid MGs (deterministic scenario)

It is considered that the demanded load in the AC and DC sides and the output power of PVs and WTs are deterministic. PSO is used to solve the problem and the scheduling of units in AC and DC sides in the studied horizon are shown in Figs. 2 and 3.

In Figs. 2 and 3, in each hour the power of units that are consumer or load is shown by negative values and the power of generative units is shown by positive values. The sum of the consumed and generated powers in each hour are equal. In Figs. 2 and 3, the negative values related to the battery are representative of battery charging which shows the battery is a consumer while positive values show that the battery is discharging, and acts as a power supplier.

B. Optimal Day-ahead Power Management of Hybrid MGs while Considering Uncertainties

The stochastic optimal power management of hybrid MGs is studied in this section. UT approach is used to deal with the uncertainties of DC and AC demanded loads and the output power of PV and WT. Assuming that the demanded load of the hybrid MG and the available output powers of WT and PV units are based on a normal distribution function and if the considered values of the deterministic scenario are the mean value (MV) of these variables, and the standard deviation (SD) equal to 5%, as well as a positive linear correlation among loads of the AC and DC sides of the MG, 8 scenarios according to equations of UT approach are originated. Afterward, the considered power management for each scenario is solved and the optimal planning of each unit is presented.

The optimal planning of the hybrid MG which is the result of the average of the eight considered scenarios based on the UT method is shown in Figs. 4 and 5. By comparing the results of Figs. 2 and 3 with Figs. 4 and 5, it is observed that the optimal power management of units in hybrid MG when considering uncertainties is different from those of the deterministic scenario.

The operational cost of the hybrid MG in different considered scenarios, the MV, and the SD of the total scenarios are tabulated in Table V.

According to Table V, in scenarios I and III where output powers of WT and PV units are increased the operation cost is decreased and vice versa. According to scenarios V-VIII, the effect of load changes has a considerable effect on the expected cost of hybrid MG.

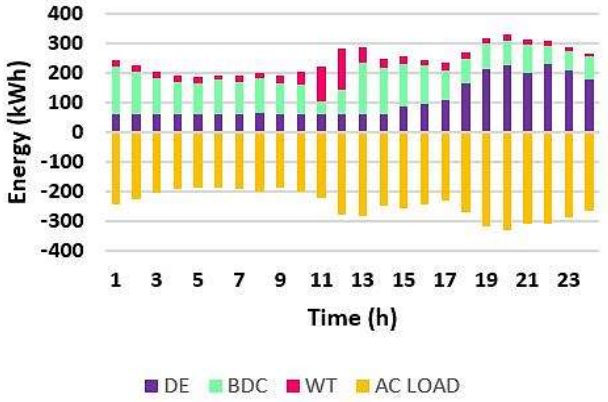


Fig. 2. Planning of units in the AC side of the hybrid microgrid without considering uncertainties.

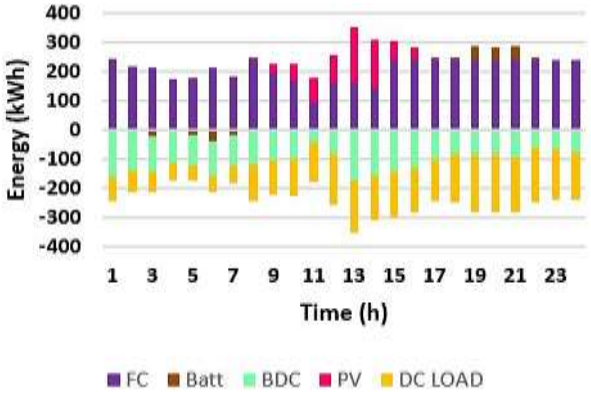


Fig. 3. Planning of units in the DC side of the hybrid microgrid without considering uncertainties.

TABLE I. PARAMETERS OF DE UNITS.

	P_{Min} (kW)	P_{Max} (kW)	a ($/kWh^2)	b ($/kWh)	c ($)	SUC ($)	R^{up}/R^{down} (kW)	MUT/MDT (h)
DE	60	360	0.007	0.03	2.9	0.68	120	2

TABLE II. PARAMETERS OF FC UNITS.

	P_{Min} (kW)	P_{Max} (kW)	C_{NG} ($/m^3)	Q_{LHV} ($kWh/m^3)	η (%)	SUC ($)	R^{up}/R^{down} (kW)	MUT/MDT (h)
FC	40	240	0.35	9.7	58	0.86	150	2

TABLE III. PARAMETERS OF BATTERY.

Unit	$P^{Ch/Dch}_{Batt,Max}$ (kW)	$P^{Ch/Dch}_{Batt,Min}$ (kW)	$E_{Batt,Min}$ (kWh)	$E_{Batt,Max}$ (kWh)	η_{Batt} (%)
Battery	45	0	60	240	90

TABLE IV. PARAMETERS OF BDC.

Unit	$P_{BDC,Max}(kW)$	$R^{up}/R^{down}(kW)$
BDC	250	150

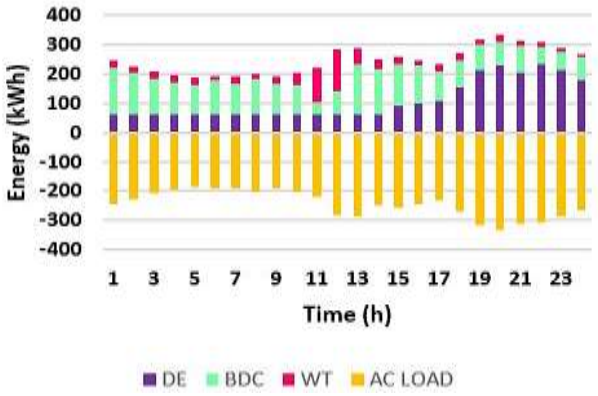


Fig. 4. Planning of units in the AC side for optimal power management of hybrid microgrid resulting from averaging of the eight generated scenarios by UT method.

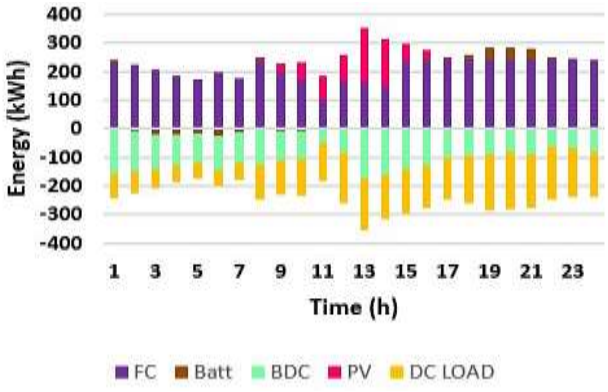


Fig. 5. Planning of units in the DC side for optimal power management of hybrid microgrid resulting from averaging of the eight generated scenarios by UT method.

TABLE V. OPERATIONAL COST IN DIFFERENT CONSIDERED SCENARIOS.

Scenario	Operational cost ($)
The base case (deterministic)	965
Scenario I: increasing the output power of PV unit	949
Scenario II: decreasing the output power of PV unit	978
Scenario III: increasing the output power of WT unit	941
Scenario IV: decreasing the output power of WT unit	990
Scenario V: increasing the AC demanded load	1308
Scenario VI: decreasing the AC demanded load	704
Scenario VII: increasing the DC demanded load	1152
Scenario VIII: decreasing the DC demanded load	800
The mean value and standard deviation	974 ±144.7

By comparing the base case (deterministic scenario) with the probabilistic scenario it is observed that the mean value of the expected operation cost increases with 15% tolerance while the tolerance of uncertain variables is considered equal to 10% of their mean values.

V. CONCLUSIONS

Recently, power management of hybrid MGs attracts attention due to the daily increase of DC loads including electric vehicles as well as the tendency to apply the maximum potential capacity of RESs and energy storage devices. The probabilistic problem of scheduling a hybrid MG is investigated in this paper using the PSO-UT algorithm. The problem is assessed from the economic point of view with RESs, AC, and DC load uncertainties, and the correlation among the random variables. It is concluded that the closer the MG parameters are to real conditions and the more accurate the modeling of uncertain variables, the more valid the solutions obtained from clarifying the MG power management problem. When the way of considering uncertain variables is far from the existing reality, the solution obtained as the optimal solution for MG power management and units' planning may not be the optimal solution.

REFERENCES

[1] S. Ali, Z. Zheng, M. Aillerie, J.-P. Sawicki, M.-C. Péra, and D. Hissel, "A Review of DC Microgrid Energy Management Systems Dedicated to Residential Applications." *Energies*, vol. 14, no. 14, p. 4308, 2021.

[2] S. K. Sahoo, A. K. Sinha, and N. Kishore, "Control techniques in AC, DC, and hybrid AC–DC microgrid: a review," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 6, pp. 738-759, 2017.

[3] S. Bahramirad, Review of the AC/DC microgrid operation and control: Illinois Institute of Technology, 2016.

[4] K. Gong, X. Wang, C. Jiang, M. Shahidehpour, X. Liu, and Z. Zhu, "Security-Constrained Optimal Sizing and Siting of BESS in Hybrid AC/DC Microgrid Considering Post-Contingency Corrective Rescheduling." *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2110-2122, 2021.

[5] E. Unamuno and J. A. Barrena, "Hybrid ac/dc microgrids—Part I: Review and classification of topologies," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1251-1259, 2015.

[6] X. Liu, P. Wang, and P. C. Loh, "A hybrid AC/DC microgrid and its coordination control," *IEEE Transactions on Smart Grid*, vol. 2, pp. 278-286, 2011.

[7] P. Wang, X. Liu, C. Jin, P. Loh, and F. Choo, "A hybrid AC/DC micro-grid architecture, operation and control," in *Power and Energy Society General Meeting, 2011 IEEE*, 2011, pp. 1-8.

[8] A. A. A. Radwan and Y. A.-R. I. Mohamed, "Assessment and mitigation of interaction dynamics in hybrid AC/DC distribution generation systems," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1382-1393, 2012.

[9] K. Kurohane, T. Senjyu, A. Uehara, A. Yona, T. Funabashi, and C.-H. Kim, "A hybrid smart AC/DC power system," *in Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference*, 2010, pp. 764-769.

[10] P. T. Baboli, M. Shahparasti, M. P. Moghaddam, M. R. Haghifam, and M. Mohamadian, "Energy management and operation modelling of hybrid AC–DC microgrid," *IET Generation, Transmission & Distribution*, vol. 8, pp. 1700-1711, 2014.

[11] P. T. Baboli, M. P. Moghaddam, M. R. Haghifam, M. Shafie-khah, and J. P. Catalão, "Serving flexible reliability in hybrid AC-DC microgrid using demand response and renewable energy resources," in *Power Systems Computation Conference (PSCC), 2014*, 2014, pp. 1-7.

[12] M. C. Bozchalui and R. Sharma, "Optimal operation of commercial building microgrids using multi-objective optimization to achieve emissions and efficiency targets," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1-8.

[13] P. T. Baboli, S. Bahramara, M. P. Moghaddam, and M.-R. Haghifam, "A mixed-integer linear model for optimal operation of hybrid AC-DC microgrid considering Renewable Energy Resources and PHEVs," in *PowerTech, 2015 IEEE Eindhoven*, 2015, pp. 1-5.

[14] P. C. Loh, D. Li, Y. K. Chai, and F. Blaabjerg, "Autonomous operation of hybrid microgrid with AC and DC subgrids," *IEEE transactions on power electronics*, vol. 28, pp. 2214-2223, 2013.

[15] N. Eghtedarpour and E. Farjah, "Power control and management in a hybrid AC/DC microgrid," *IEEE transactions on smart grid*, vol. 5, pp. 1494-1505, 2014.

[16] V.-H. Bui, A. Hussain, and H.-M. Kim, "A Strategy for Optimal Operation of Hybrid AC/DC Microgrid under Different Connection Failure Scenarios," *International Journal of Smart Home*, vol. 10, pp. 231-244, 2016.

[17] M. Javidsharifi, H. Pourroshanfekr Arabani, T. Kerekes, D. Sera, S. Spataru, and J. M. Guerrero, "Effect of Battery Degradation on the Probabilistic Optimal Operation of Renewable-Based Microgrids." *Electricity*, vol. 3, no. 1, pp. 53-74, 2022.

[18] M. Javidsharifi, H. Pourroshanfekr Arabani, T. Kerekes, D. Sera, and J. M. Guerrero, "Stochastic Optimal Strategy for Power Management in Interconnected Multi-Microgrid Systems." *Electronics*, vol. 11, no. 9, p. 1424, 2022.

[19] M. Javidsharifi, H. Pourroshanfekr Arabani, T. Kerekes, D. Sera, S. V. Spataru, and J. M. Guerrero, "Optimum Sizing of Photovoltaic-Battery Power Supply for Drone-Based Cellular Networks." *Drones*, vol. 5, no. 4, p. 138, 2021.

[20] M. Javidsharifi, T. Niknam, J. Aghaei, M. Shafie-khah, and J. P. S. Catalao, "Probabilistic Model for Microgrids Optimal Energy Management Considering AC Network Constraints." *IEEE Systems Journal*, vol. 14, no. 2, pp. 2703-2712, 2020.

A Review of Security Algorithms for Smart Home Internet of Things Devices

Maanda Magidimisa
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
MagidimisaM@gmail.com

Topside E. Mathonsi
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
MathosiTE@tut.ac.za

Vusumzi Malele
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
vusi.malele@dst.gov.za

Tonderai Muchenje
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
MuchenjeT@tut.ac.za

***Abstract*— Smart home with the Internet of Things (IoT) is now playing an important role in connecting most home devices to the Internet in real-time, and this directly improves the quality of life. This processing of data in real-time opens more opportunities for cyber-attacks to target the smart home IoT environment. Most cyber-attackers have taken advantage that most manufacturers are not focused on the security of those devices and other devices have limited processing power to handle security, because of their size or mobility. Distributed Denial of Service (DDoS) attack is one of the most serious threats to network security in smart IoT home devices and defines against DDOS attack is one of the most researched topics in smart IoT home devices. As a result, this research study proposes a DDoS algorithm to improve the detection accuracy of attacks in a smart IoT home environment. In the future, a computer network simulator will be used to compare the effectiveness of our proposed algorithm against that of state-of-the-art algorithms. Thereafter, simulation results will be presented.**

Keywords*— *Cyber-attacks, Distributed Denial of Service, Internet of Things, Limited Processing Power, Smart home.

I. INTRODUCTION

In recent years the adoption of smart Internet of Things (IoT) home devices has improved from close range connectivity to wide area network (WAN) and with this, the introduction of network-enabled devices emerged [1-3]. The IoT technology presents new convenience that enhances the availableness of smart home gadgets [4,5] and monitoring of smart IoT home infrastructure and systems from anywhere at anytime [6, 7] smart IoT home, has been developed to enhance small physical devices to exchange sensitive data wirelessly and enable remote management and collect data that can also be used for Automation. The most common IoT device in Smart Homes includes but are not limited to smart sensors, smart hubs, smart Fridge, Smart light, smart cameras, smart lock system, smart speakers, and smart TV [1-5].

In smart IoT, home network security is the primary concern when planning to adopt the home automation system since this opens up more cyber-attacks than physical attacks [8]. This means that hackers do not need to get into the home physically, they can virtually attack the system and have all control of home devices. Some common IoT attack types include Encryption Attacks, Denial of Service (DoS) attacks, Mirai-like, Botnets, Privilege Escalation, and Brute Force Password attacks.

[4, 9-11] they all have highlighted that DoS or DDoS is the most common attack in smart IoT home devices, they are the greatest threat to network security vulnerability and the defence against DDoS attacks is not easy, but very important. DDoS detection and mitigation are highly influenced by the location of the attack because it is easier to accurately detect close by the victim and easy for a Close attack to throttle the attacking source [12].

However, literature has shown that existing IoT security algorithms reduce the network speed or device performance, and when processing power is affected, the detection accuracy of network threats is also affected, mostly with DoS attacks [13, 14]. The main aim of this research is to develop a new enhanced DDoS algorithm that will improve the detection accuracy of a security vulnerability in a smart IoT home environment, without impacting device speed and performance.

To achieve the main objective of this research, we will adopt an experimental methodology and the methods that will be used are, Literature review for gathering our information or data, Mathematical models will be used to design and develop our DDoS algorithm, then the simulation will be used to test and to identify the effectiveness of our algorithm against already existing algorithms. The remainder of this research paper has been organize as follows: The research paper discusses the related works in Section 2. In addition, this research paper discusses the proposed DDoS algorithm in Section 3. In Section 4, this research paper provides the conclusion and forthcoming work.

II. RELATED WORK

With the advancement in machine learning and the adoption of smart IoT home devices, this section describes some related work for smart IoT home technology and how other researchers explore these security concerns. We will also look at how they have provided the solution, and what are the improvements and limitations of their solutions against security issues in smart home systems.

[1] has proposed a Network Intrusion Detection System (NIDS) which is placed at the router level in the smart home internal network. Their algorithm used a smart hub, the smart hub interprets the smart home devices network parameters like device ID, Media Access Control (MAC) address, IP address, among many more. The smart hub connects all network devices

in the smart home network and then routes their respective connections to the smart home network or connections between the devices within the smart home network from the external network. The obtained simulation results showed that all known intrusions in the database were detected, and all unauthorized connections were also detected at 100%. However, this proposed approach detects false-positive and requires additional security to protect the hub from being compromised.

An ML-based model to detect Mirai-like attacks was proposed for the IoT protocols this is because these protocols provide critical analysis of their performance against various security threats [2]. Their algorithm uses the following network topology model with Mininet with Open switch supported by OpenFlow, and Floodlight controller all installed in Ubuntu. The obtained simulation results exhibited that these ML algorithms with lightweight features would accurately classify DDoS attack traffic from normal IoT traffic in a smart home network. However, running Mirai code was not used to avoid trying any wide-range network damage and the algorithm was designed to defend against Mirai-like attacks.

[3] has proposed a rule-based approach towards generating Adversarial Machine Learning (AML) attack samples and explores how they can be used in machine learning algorithms for detecting DoS attacks in an IoT smart home network. Their algorithm used the following network parameters namely, IP Cameras, Smart plugs, Smart hub, motion sensors, Lamp, and computers. For the network authentic and suitable-sized IoT smart home dataset was used. The obtained simulation results showed that their rule-based approach to generating AML attacks was successful and reduced the performance of already existing algorithms when defending against DoS attacks.

[4] proposed the defined strategies of DDoS attacks based on the principle of DDoS and the defensive purpose. Their algorithm used the following network parameters namely, the Internet's entry router, and antivirus. This strategy was achieved by combining the victim, the intermediate network, and the source network, this is very effective for DDoS attack defines. However, the algorithm presented by [4] struggles when attackers are using counterfeit source IP addresses, which makes it difficult for this algorithm's attackers to be tracked.

[5] proposed a study to showcase that physical, network, and software (encryption) is viable for smart home IoT devices. Their compared security postures between well-known and less-known IoT vendors. This review has been covered in major vulnerability databases. This study has compared fewer known vendors such as Leeo and Feit Electric; and well-known vendors such as Google and Philips Hue in order to determine the best security solution for smart IoT networks. Their experiments exhibited that the well-known vendors/devices have stronger security postures, whereas less known manufacturers have weaker security for smart home IoT devices. In addition, their algorithm requires the security requirements of the smart IoT devices to be standardized along with the effectiveness categories and attack measurements.

The review of the literature has shown that other researchers have developed and implemented different DDoS algorithms for the Smart home IoT environment. However, due to the nature of IoT devices, size, and manufacturing standards, it has been very difficult and challenging to create a centralized algorithm to achieve data integrity, confidentiality, privacy, and availability smart home IoT environment without compromising device speed or power consumption. In the following Section, we will be presenting the idea of how this proposed solution will be designed in the future.

III. Proposed solution

This research paper proposes the design of an enhanced DDoS Algorithm for Smart Home Internet of Things Devices using a DDoS algorithm. The DDoS algorithm was chosen because this attack has been highlighted as one of the most serious threats to network security in smart IoT home devices and defines the DDOS attack as one of the most researched topics in smart IoT home devices. The aggressiveness of DDoS attacks has made this study consider the geographical area defending the system against this type of attack. In addition, the attack or connection to the smart IoT environment will need to be within the close range of the system's geographic area and any connection that will be out of the geographical area will be required to have unauthenticated with 2FA to be able to gain access to the environment. When those conditions are not met the sour IP will be added to the block IP list and a notification sent to the admin/owner.

IV. Conclusion and future work

The Smart Home IoT environment has been the centre and playground for cyber-attackers. The DDoS algorithm in smart IoT home devices plays an important role in data privacy and integrity in smart IoT home devices. Previously advanced DDoS algorithms failed to stop DDoS attacks, based on the nature of IoT devices, size, and manufacturing standards. In the future, the design of a DDoS algorithm with a high level of security that will achieve data integrity, confidentiality, privacy, and availability of data in smart home IoT devices will be presented. Thereafter, MATLAB/NS3 will be used as a tool to evaluate the proposed solution, and simulation results will be presented.

V. References

[1] M.M. Pillai, and A. Helberg, “Improving Security in Smart Home Networks through user-defined device interaction rules”. IEEE Africon, pp. 1-6, 2021.

[2] Y. Mtawa, H. Singh, A. Haque, and A. Refaey, “Smart Home Networks: Security Perspective and MLbased DDoS Detection”. IEEE Canadian Conference on Electrical and Computer Engineering, pp. 1-7, 2020.

[3] E. Anthi, L. Williams, A. Javed, and P. Burnap, “Hardening machine learning denial of service (DoS) defences against adversarial attacks in IoT smart home networks”. Published by Elsevier Ltd, pp. 1-12, 2021.

[4] L. Wenliang, and H. Wenzhi, “DDOS Defense Strategy in Software Definition Networks”. International Conference on Computer Network, Electronic and Automation (ICCNEA), pp. 1-4, 2019.

[5] B.D. Davis, J.C. Mason, and M. Anwar, “Vulnerability Studies and Security Postures of IoT Devices: A Smart Home Case Study”. IEEE Internet Of Things Journal, pp. 1-9, 2020.

[6] M. Zhang, J. Wang, and Y. Hu, “A New Approach to Security Analysis of Wireless Sensor Networks for Smart Home Systems”. International Conference on Intelligent Networking and Collaborative Systems, pp. 1-7, 2016.

[7] W.L. Costa, M.M. Silveira, T. Araujo, and R.L. Gomes, “Improving DDoS Detection in IoT Networks Through Analysis of Network Traffic Characteristics”. 2020 IEEE Latin-American Conference, pp. 1-6, 2020.

[8] S. Rehman, and V. Gruhn, “An Approach to Secure Smart Homes in Cyber - Physical Systems/Internet-of-Things”. Fifth International Conference on Software Defined Systems (SDS), pp. 1-4, 2018.

[9] Q. Chen, H. Chen, Y. Cai, Y. Zhang, and X. Huang, “Denial of Service Attack on IoT System”. 9th International Conference, pp. 1-4, 2018.

[10] R. Doshi, N. Apthorpe, and N. Feamster, “Machine Learning DDoS Detection for Consumer Internet of Things Devices”. IEEE Symposium on Security and Privacy Workshops, pp. 1-7, 2018.

[11] A. Verma, and V. Ranga, “Machine Learning Based Intrusion Detection Systems for IoT Applications”. Wireless Personal Communications, pp. 1-24, 2020.

[12] L. Xie, X. Xiao, Y. Shi, C. Zhang, and J. Jiang, “An Activatable DDoS Defense for Wireless Sensor Networks”, IEEE 9th International Conference on Information, Communication and Networks, pp. 1-5, 2021

[13] E. Kim, and C. Keum, “Trustworthy Gateway System Providing IoT Trust Domain of Smart Home”. Ninth International Conference on Ubiquitous and Future Networks, pp. 1-3, 2017.

[14] R. Yu, X. Zhang, and M. Zhang, “Smart Home Security Analysis System Based on The Internet of Things, IEEE 2nd International Conference on Big Data”, Artificial Intelligence and Internet of Things Engineering, pp. 1-4, 2021.

An Enhanced VANET'S Security Model for Mitigating Denial of Service attacks in Smart Cities

Ntshuxeko Makondo
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
ntshuxekomakondo@gmail.com

Topside E. Mathonsi
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
mathonsiTE@tut.ac.za

Tshimangadzo M Tshilongamulenzhe
Department of Information Technology
Tshwane University of Technology
Pretoria, South Africa
tshilongamulenzhetm@tut.ac.za

***Abstract*— Vehicle ad-hoc networks (VANETs) were created because of recent improvements in wireless communication. VANETs provide a platform for improving passenger comfort and safety. VANETs are a specific type of network that falls within the family of Mobile Ad hoc Networks (MANET) whereby safety is the primary focus since essential information on driver safety and support must be distributed between vehicle nodes. If network availability is raised, the security of the nodes can be improved. In the event that the network is subject to a Denial-of-Service (DoS) attack, the network's availability is reduced. The Cuckoo filter scheme and Advance malicious IP Detection (AMIPD) are integrated to develop an Enhanced Malicious IP Detection and Prevention (EMIDP) algorithm for identifying and mitigating DoS attacks. This approach identifies and prevents many malicious nodes in the network system from sending infected packets in order to block the communication in the network and prevent it from transmitting safety signals**.

Keywords—VANET, Dos Attack, MANET, Cuckoo Filter, AMIPD, EMIDP

1. INTRODUCTION

Incorporating software-based intelligence into vehicles has the potential to improve passengers' quality of life. vast arrays of mobile applications can be built on top of Vehicular Ad hoc Networks (VANETs). VANETs are self-organizing, dispersed networks that use moving vehicles, road-side units (RSUs), and base stations (BSS) as nodes to construct a mobile network, converting every vehicle or bs into a router [1].

VANETs are dispersed, and self-organized networks are made up of many cars, they have developed as a new strong technology to improve driving safety and traffic management. VANETs allow cars to communicate with one another and with adjacent roadside devices. VANETs cars can communicate with roadside devices or with each other to exchange information. To form a link, vehicles function as mobile nodes in a network; these nodes should interact with one another via a single hop or several hops. All communicating nodes are fitted with short-range radios. The transmission distance between the car nodes is less than 300m. RSUs are deployed at random based on the network's category in that specific location. RSUs allow authorities and vehicle nodes to interact. This network will most likely play a significant part in enabling a comfortable traffic system on highways, as well as in minimizing unnecessary traffic incidents [2]. Fig. 1 demonstrates VANETs architecture.

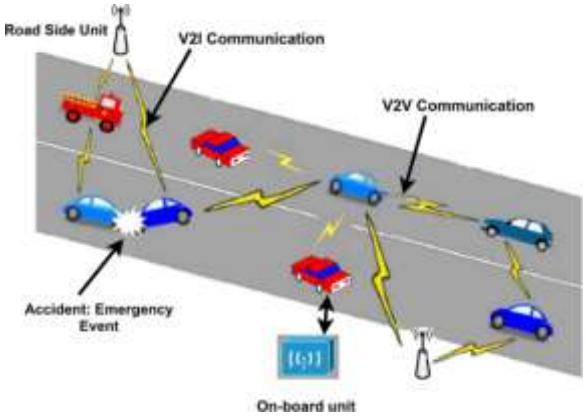


Fig. 1. VANETs architecture

Raghavan [3] explained that the nodes are linked together and communicate information via wireless connections. As a result, security is another critical criterion that must be carefully examined in the VANETs to avoid the spread of false messages that might interrupt traffic. Although data authentication and integrity can secure sent data, VANETs will not be able to fulfil their objectives if services to legitimate users are not accessible when they need them. In the VANETs, Denial of Service (DoS) attacks must be significant attacks that must always be managed. As a result, it's critical to comprehend DoS attacks in the context of VANETs. This research study investigates DoS attacks in VANETs to offer a better security model for detecting and preventing the attacker, the target, its capabilities, the kind of vulnerability utilised in the attack, and the victim's effect

VANETs faces downgraded performance due to a lack of adequate security measures in the network that might lead to malicious intrusions and service exploitation, which could be devastating to drivers in Smart cities. Due to the absence of fixed network infrastructure, vehicular networks rely on the vehicles themselves to provide network functionality. In contrast to wired networks, which are secured by many layers of defense, such as physical firewalls, wireless networks are not [4].

The aim of this research is to develop a security solution that will improve the performance of VANETs. After evaluating the gaps in the existing solutions, we will be able to implement a robust security defense solution against DoS attacks that is suitable for VANETs. The experiential methodology is going to be used for this research study to achieve the overall goal of the study, which is going to be divided into a literature view, modelling, and implementation of the proposed solution.

2. LITERATURE REVIEW

A DoS attack is an attempt to make a service unavailable or arrive late for users. Because it entirely interrupts the connection, this attack is extremely dangerous to the vehicles. The entire network is brought to a halt by this attack, which affects the network's dependability. This form of attack has a wide range of means, reasons, and goals. The attacker's purpose may be to exploit weaknesses in network hardware infrastructure or application services, resulting in varying effects on the victims depending on the attacker's capabilities. Knowing the attacker's goal and capabilities to carry out the attack is critical for preventing or detecting attacks [5].

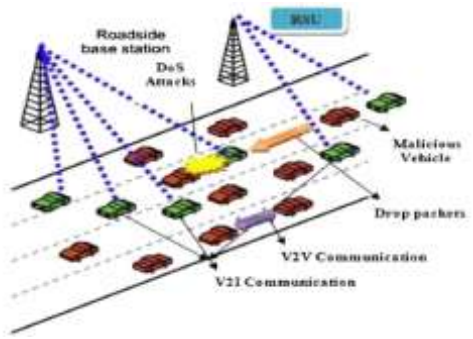


Fig. 2. DoS Attack in VANET

Based on capacity and motivation, attackers might be classified as follows.

a) Vandal: is careless they only want to demonstrate their attacking ability.
b) Hacker: a person who is motivated by curiosity and interest. You will get no personal benefit from the attack.
c) Deception hacker: motivated by personal or organizational financial and/or political gain.
d) Terrorist: a person who is driven primarily by political ideology, money, time, and staff are all well-equipped.

The DoS attack is the most common harmful attack on communication networks observed in VANETs these days, and it is one of the most common VANETs security risks. An attack like this disables all of VANET's services. Fig. 2 shows the many forms of DoS attacks and how they may be carried out in two ways [5].

3. RELATED WORK

Many attempts have been undertaken in recent years to reduce DoS attacks in vehicular networks to improve network system performance. This section addresses these previously offered remedies as well as the gaps that were found.

Rampaul et al. [6] presented the Attacked Packet Detection Algorithm (APDA), which improves VANET security while avoiding early delay costs. VANET security can be improved by using the technique before the overhead of verification time delay is reduced. This algorithm was evaluated and tested using Network Simulator-2 (NS-2). Its drawback is that the algorithm is ineffective when many invalid requests are sent simultaneously from several cars.

Durrani et al. [7] presented a novel approach for preventing DoS attacks in semiautonomous and self-driving cars. This security system's main objective is to single out infected vehicles in the network in real-time. The sophisticated Intrusion Detection System (IDS) may detect odd or malicious activity and prevent a hostile vehicle from gaining access to the system. The researchers put this method to the test and analyzed the outcomes using the MATLAB Simulation tool. They discovered that the detecting system only uses data from the network system in the form of a trace file.

Rajhi et al. [2] proposed a method to defend network systems against DoS attacks. When the network is attacked by malicious nodes, this algorithm will recognize the malicious nodes and reject all packets transmitted in the network by them. As a result, this algorithm will aid in the continuous availability of the network for the transmission of important life-related information. With the use of RSUs, this method will aid in the identification of rogue nodes by identifying irrelevant packets. Each node will interact with the RSUs,

allowing the RSUs to save each vehicle's information. When a node transmits hazardous signals, that car may be recognized and verified using information from the RSUs. Because the network contains numerous nodes, the simulation was done entirely in NS-2 and after evaluating the results, it was discovered that the network's throughput has been improved however, the network lifespan is reduced marginally, but the packet loss ratio has drastically increased.

The Bloom-filter-based DoS detection technique is a type of proactive and reactive DoS detection scheme. The proactive strategy is used to keep track of new IP addresses (nodes), whereas the reactive approach is used to figure out how many cars are linked (nodes). The network scalability is boosted in this sort of detection approach by generating a near zone by the closer vehicles, which can save bandwidth, collisions, and computational overhead. The proposed scheme offers an end-to-end solution for DoS attack mitigation. Based on the Bloom-filter, this is a unique approach for detecting DoS attacks. The Bloom-filter-based IP-CHOCK detection (BFICK) approach may be used to identify and fight IP spoofing addresses used in DoS attacks. It ensures that authentic vehicles in the VANETs have access to a service. In terms of computational cost and storage space, the proposed scheme is simple and effective. For high attack rates, this technique works well. This technique may be used to track down the source of the attacks. The simulation results show that the proposed method has lowered the average tracking detection time and average mistake detection probability while also improving the detection rate, but packet loss was a major issue [8].

Cooperative Message Authentication Protocol (CMAP) was used to identify malicious information broadcasted by malicious cars in the road transportation system. This car actively provides its position as well as other safety warnings to avoid a collision. As a result, it is necessary to authenticate each communication and vehicle. They have suggested three verifier selection algorithms in this protocol, which will be used to perform authentication [7]. After thoroughly testing and assessing the findings using the NS-2, researchers discovered that the system's primary drawback is that if no validator is present to validate the message, malicious messages may be received by the cars.

Enhanced Attacked Packet Detection Algorithm (EAPDA) that was created to identify DoS attacks on the VANETs. It makes use of time slots as well as Threshold values. The invader nodes are identified via a communication gap. Finally, the entire network is protected against the discovered danger. The simulation findings from NS-2 indicate that it improves network throughput while avoiding false alerts. When a DoS attack is identified, it is based on the average connection time of the nodes, as opposed to a previous method whose threshold was limited by area. As opposed to previous techniques, this technique never incorrectly detects any node as a malicious node. As a result, it is more responsive, verification takes less time, and throughput is higher. However, emergency vehicles, such as ambulances and fire extinguishers, are not given priority under this approach, thus they may be verified in a much longer time than expected [9].

Multivariate stream analysis (MVSA) that enables for RSU-based V2V connectivity. A DoS attack is recognized by scanning the network trace and estimating the average payload rate, the frequency at various intervals, and time to live per vehicle for each strike class. Using traces, the method calculates the weight of the flow. If it is valid or malicious, it will be flagged by the MVSA system [10]. With this method, network reliability and quality are assured however, excessive bandwidth usage remains a major concern, according to simulation results.

Raghavan [1] proposed a filter that is based on IP detection as a realistic method for detecting DoS attacks in VANET. The technique not just to improves the detection of the attack, but also the exchange of information about the attack to other nodes in the network. The system will outperform existing Dos attack detection methods in terms of true alarm rate, detection ratio, and latency. This present effort focuses solely on detecting the attack and transmitting it to the rest of the network's cars. The proposed technique produces extremely accurate detection and classification results at a minimal computing cost, according to simulation findings from NS-2. Their method is incapable of removing the malicious node from the network, therefore increasing the risk of network damage.

Enhanced Attacked Packet Detection Algorithm (EAPDA) was created to identify DoS attacks on the VANETs. It makes use of time slots as well as Threshold values. The invader nodes are identified via a communication gap. Finally, the entire network is protected against the discovered danger. The simulation findings from NS-2 indicate that it improves network throughput while avoiding false alerts. When a DoS attack is identified, it is based on the average connection time of the nodes, as opposed to a previous method whose threshold was limited by area. As opposed to previous techniques, this technique never incorrectly detects any node as a malicious node. As a result, it is more responsive, verification takes less time, and throughput is higher. However, emergency vehicles, such as ambulances and fire extinguishers, are not given priority under this approach, thus they may be verified in a much longer time than expected [9].

Jie et al. [8] created the Port Hopping security mechanism to reduce DoS attacks. Port hopping and a single linear space are

used to implement a simple and elegant security method for adjusting the vulnerable networks' port numbers for Vehicle to vehicle and Vehicle to infrastructure communications in various service slots. To overcome the complexity of DoS attacks in complex VANETs, a novel technique described as security strategy matrices were implemented to detect the port numbers checked by DoS attackers. The simulation results showed that the method does not address the hopping frequency issue.

The nodes responsible for network attacks are identified based on frequency and velocity. This method detects both irrelevant and legitimate packets. Instead of detecting a single node that is attacking the network, the algorithm also detects numerous nodes that are attacking the network. The network's lifespan is extended by detecting attacker nodes early. Other performance parameters indicate a significant change in their values, demonstrating that the suggested technique is an enhanced version of existing packet identification algorithms [9].

A well-known probabilistic data structure for handling large volumes of data is called the Cuckoo filter, it is believed that the Cuckoo filter is a more sophisticated version of the Bloom filter. The cuckoo filter will make up for the bloom filter's shortcomings. The filter's unique features include the capacity for data deletion, a low rate of false positives, and the application of cuckoo hashing to prevent collisions. The cuckoo filter stores the fingerprints of the objects, not the things to be added, in the cuckoo hash table. The bitstream that is obtained after hashing the object contains fingerprint representations. The cuckoo hash table, which is based on the two hash functions, maps each item that is put to one of two possible buckets. The buckets may store various numbers of fingerprints [3]. The strategy helps in both the attack's detection and the dissemination of information about it to other nodes in the network. The solution will perform better in terms of detection ratio, false positive rate, and end-to-end latency than the current solutions. According to simulation tests, the proposed technique uses fewer computer resources while delivering extremely accurate detection and classification results. However, this solution doesn't eliminate harmful IPs from the network [3].

The review of the literature has shown that many solutions have been proposed to mitigate DoS attacks in VANET to ensure that performance does not suffer any service disruptions. To overcome the identified problem caused by DoS attacks, this paper is proposing the EMIDP algorithm for identifying and mitigating DoS Attacks.

4. DESIGN OF EMIDP ALGORITHM

The goal of the proposed EMIDP Algorithm is to mitigate DoS attacks in VANET to ensure that the network performance is not disrupted. The architecture incorporates two critical strategies for mitigating DoS attacks in Vehicular Networks. Advanced Malicious IP detection is the initial stage, which entails keeping an eye out for unusual behaviour in packets. Cuckoo filter is used in the second step to make the choice. Both strategies are used in combination to identify and drop the malicious packet as quickly as possible. Also, inform other vehicles to prevent further disruption to the network.

I. Cuckoo Filter and Hashing

The cuckoo filter is a well-known probabilistic data structure used to handle massive amounts of data. The Cuckoo filter is thought to be a more advanced variation of the Bloom filter. The cuckoo filter will make up for the weaknesses of the bloom filter. Some of the filter's interesting characteristics include the ability to delete data, a low false positive rate, and the use of cuckoo hashing to avoid collisions. The cuckoo filter stores the fingerprints of the objects, not the things to be added, in the cuckoo hash table. The fingerprints are represented in the bitstream obtained after hashing the object. When an item is inserted, it is mapped to one of two potential buckets in the cuckoo hash table, which is based on the two hash functions. The buckets are designed to save a varying number of fingerprints [3].

Collisions are dealt with using cuckoo hashing in probabilistic data structures. In cuckoo hashing, values are assigned to one of two buckets and two hash algorithms are used for each key. Following hashing, the item is examined to see if the first bucket is empty, and if it is, the item is deposited. Place the item in the second bucket if the first bucket has it. If an item occupies the second bucket, the preceding item is removed to make room for the new item. This process is repeated for each new item that is added. While adding and removing items throughout this procedure, there is a small chance that we will enter an infinite loop. We must log the bucket entries to avoid the problem, but the only way out of the loop is to rebuild the hash table. It is also known as partial key cuckoo hashing and consists of the item insertion phases listed below [10].

a) *Add the new key* $\boldsymbol{K}$.
b) *Process the hash for key* $\boldsymbol{K}$ *using the first hash function, where* $\boldsymbol{h(k) = kh.}$
c) *If the first bucket is available, place the hashed value* $\boldsymbol{kh}$ *in it.*
d) *The key will now be processed using a second hash function on* $\boldsymbol{kh}$*, which contains the value o first function, where* $\boldsymbol{g(kh) = khg}$*, if the first bucket is already full.*
e) *To obtain the key for the second bucket, which is kg, use the XOR function on both the* $\boldsymbol{kh\ and\ khg}$ *hash values.*
f) *For the hashing function to be completed, look up for the second bucket*

Reversing the previous operations will produce the bucket and the fingerprint. The other bucket's position can be calculated and processed with relative ease, which enables the cuckoo filter to keep f-bit fingerprints and save storage space [11].

The first phase, during which traffic is monitored and IP addresses are registered, is entirely responsible for the process' conclusion. Every network activity, both inbound and outbound, is recorded and tracked. By evaluating the traffic patterns the node generates, the initial step was determining whether the node might potentially have an influence on the network. A decision is made based on this in the following phase. The filtered phase, which also includes the hash function, is the Cuckoo clock's last stage. Since the filter maintains a database of all the IP addresses for the vehicles, if a malicious IP address is found in the data collected during the first step of the detection strategy, an alarm is sent to every other vehicle. The malicious IP is finally deleted from the network. Fig.3 illustrates the flowchart for the proposed EMIDP algorithm.

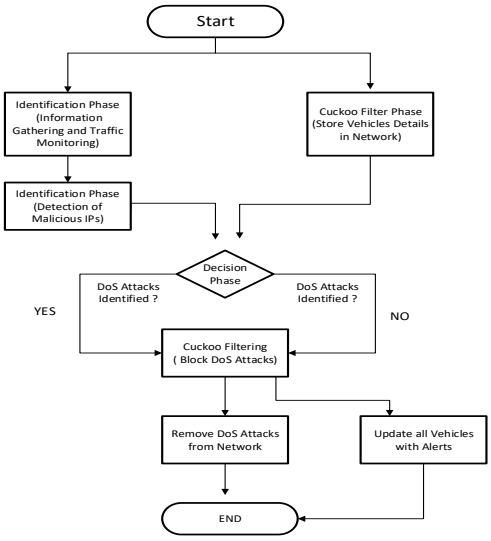


Fig. 3. Flowchart for proposed EMIDP Algorithm

5. EXPERIMENTAL EVALUATION

TCL scripts were used in this study to generate the proposed simulation topology in NS-2. The purpose of this was to assess the impact of integrating the Cuckoo filter system with AMIPD to create the proposed algorithm.

This section of the research study delves into the examination of the findings produced from simulations done to compare the proposed EMIDP Algorithm to the AMIPD algorithm and Cuckoo filter. To give credible results, this research study ran the simulations many times. Our simulations examined the following performance metrics:

I. Average Detection Ratio

The detection ratio measures the proportion of malicious vehicles to all other vehicles in the network. The number of vehicles in the network has a direct impact on the detection ratio of the system. The number of packets transferred increases with the number of vehicles. Results are obtained by changing the number of attackers present in the network. The average detection ratio for the proposed technique was 92%, while AMPIPD and the Cuckoo filter had average detection ratios of 86% and 82%, respectively as depicted in Fig. 4.

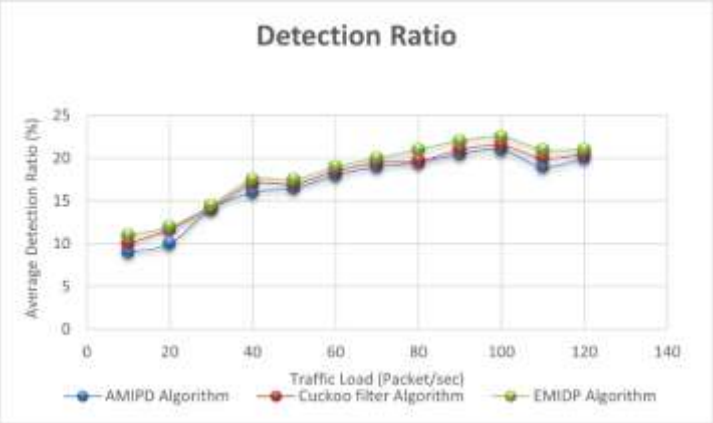


Fig. 4. Average Detection Ratio

According to the simulation data and graph generated, there are fewer malicious vehicles overall, which raises the detection ratio of such vehicles. The detection abilities increase as the number of malicious files increases, but not as much as the AMIPD algorithm and Cuckoo filter. Even if there are a lot of nodes, the EMIDP will eventually stabilize the detection ratio. This is due to the possibility of communication between nodes that are just close together.

ii. Average False Positive Ratio

False positives are seen to be crucial statistic for evaluating the effectiveness of any system. The ratio of negative events labeled as positive to all instances of negative behavior is generally used to define the false positive rate. The False Positive Ratio contrasts the overall number of authorized vehicles across the network with the number of identified approved vehicles. One of the most important criteria is this one because a legitimate vehicle could be mistaken for a harmful one. As shown in Fig. 5, the proposed algorithm had an average False Positive Ratio of 0.2% while AMPIPD and the Cuckoo filter had average detection ratios of 0.8% and 1.3%, respectively.

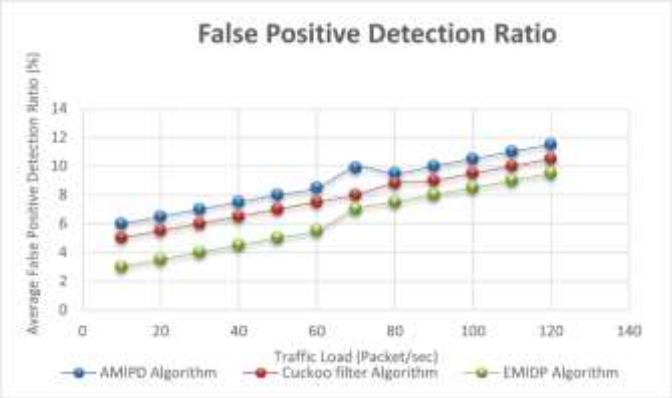


Fig. 5. False Positive Ratio

The results of the simulation show that there are not many malicious vehicles and a low false-positive ratio. Like the detection ratio, the false-positive ratio rises as more malicious vehicles are present. The communication range will eventually cause this tendency to stabilize as well. The proposed solution, however, outperforms the AMIPD algorithm and Cuckoo filter in terms of performance.

iii. Average End-to-End Delay

The end-to-end delay is sometimes referred to as the time it takes for a packet to be transmitted from source to destination through a network. It is sometimes referred to as the packet's network travel time. The end delay is considered important because it depends on how many malicious vehicles are present in the network. If there are several malicious vehicles, the delay will be longer. According to Fig. 6, the proposed algorithm had an average end-to-end delay of 0.5%, while AMPIPD and the Cuckoo filter had average detection ratios of 1.5% and 1.9%, respectively.

Fig. 6. End-to-End Delay

iv. Average Packet Delivery Ratio

Another crucial metric for every network system is the packet delivery ratio. The packet delivery ratio is calculated by dividing the total number of packets sent by the total number of packets received. In general, the wireless network will experience substantial packet losses owing to congestion and obstructions in the path of transmission. Furthermore, there are attackers on the network, which affects packet delivery. The proposed algorithm showed an average packet delivery ratio of 88% while AMPIPD and Cuckoo filter showed an average detection ratio of 80% and 75% respectively as depicted in Fig. 7.

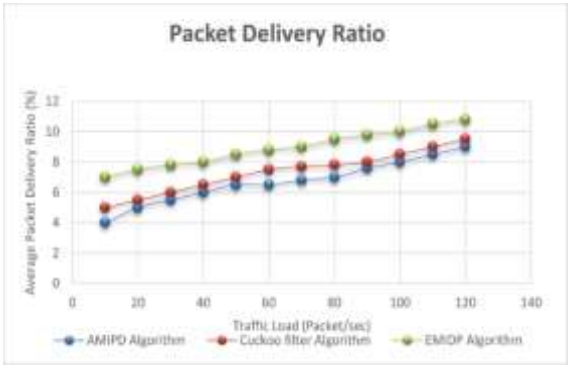


Fig. 7. Packet-delivery ratio

v. Average Packet-loss Ratio

The ratio between the number of packets received and the total number of packets transmitted is known as the packet loss ratio. Since the connection is wireless, as was explained in the PDR section, the presence of attackers causes significant packet losses in the network. According to the simulation results, the use of the wireless media and the presence of attackers have a significant influence in Packet-delivery ratio and Packet-loss Ratio The proposed algorithm showed an average Packet-loss Ratio of 0.4% while AMPIPD and Cuckoo filter showed an average detection ratio of 0.8% and 1.2%, respectively as depicted in Fig. 8.

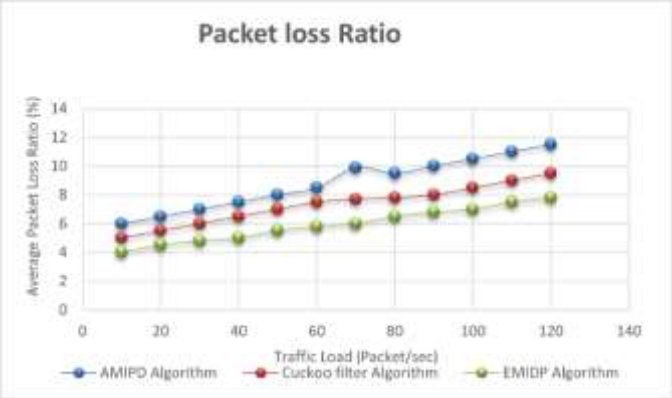


Fig. 8. Packet-loss Ratio

6. CONCLUSION AND FUTURE WORK

In this study, an EMIDP algorithm was designed and implemented to mitigate packet malicious nodes generated by intruders, which leads to packet loss, packet delay, and High False-positive Ratio in a VANETs environment. To mitigate DoS attacks in a VANET context, the proposed EMIDP algorithm integrates the Cuckoo filter with AMIPD. The paper solely discussed the detection and prevention of DoS attacks, as well as broadcasting them to other vehicles in the network. Experiment findings reveal that the proposed method produces extremely accurate detection and classification results at a low computational cost. In the future, we can also strive to detect and eliminate Distributed Denial of Service (DDoS) attacks from the network by disconnecting them, which would limit the damage to the network. Another consideration is to categorize it as random spoofing, subnet spoofing, or fixed spoofing by examining a hash table for the source IP characteristics.

Acknowledgment

The authors would like to thank the Tshwane University of Technology for its assistance.

REFERENCES

[1] U. Srinivasa Raghavan, "Detection of Denial of Service (DoS) Attacks in VANET using Filters," Dublin, National College of Ireland, 2020.

[2] M. Rajhi, H. Madkhali, and I. Daghriri, "Comparison and Analysis Performance in Topology-Based Routing Protocols in Vehicular Ad-hoc Network (VANET)," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 2021: IEEE, pp. 1139-1146.

[3] Raghavan, "Detection of Denial of Service (DoS) Attacks in VANET using Filters," Dublin, National College of Ireland, 2020.

[4] A. S. Mustafa, M. M. Hamdi, H. F. Mahdi, and M. S. Abood, "VANET: towards security issues review," in *2020 IEEE 5th international symposium on telecommunication technologies (ISTT)*, 2020: IEEE, pp. 151-156.

[5] N. A. Alsulaim, R. A. Alolaqi, and R. Y. Alhumaidan, "Proposed solutions to detect and prevent DoS attacks on VANETs system," in *2020 3rd international conference on computer applications & information security (ICCAIS)*, 2020: IEEE, pp. 1-6.

[6] C. Rudraraju, "Simulation of Detecting and Preventing DDoS in Vehicular Ad-hoc Networks (VANETS)," Dublin, National College of Ireland, 2020.

[7] A. Durrani, S. Latif, R. Latif, and H. Abbas, "Detection of Denial of Service (DoS) Attack in Vehicular Ad Hoc Networks: A Systematic Literature Review," *Adhoc & Sensor Wireless Networks*, vol. 42, 2018.

[8] Y. Jie, M. Li, C. Guo, and L. Chen, "Dynamic defense strategy against DoS attacks over vehicular ad hoc networks based on port hopping," *IEEE Access*, vol. 6, pp. 51374-51383, 2018.

[9] S. Kumar and K. S. Mann, "Prevention of DoS attacks by detection of multiple malicious nodes in VANETs," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 2019: IEEE, pp. 89-94.

[10] Ö. Kasim, "An efficient and robust deep learning based network anomaly detection against distributed denial of service attacks," *Computer Networks*, vol. 180, p. 107390, 2020.

[11] V. V. Mahale, N. P. Pareek, and V. U. Uttarwar, "Alleviation of DDoS attack using advance technique," in *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2017: IEEE, pp. 172-176.

Inspection of the classifying performance of the deepfake voices by the latest text-to-speech model

Yuta Yanagi
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
yanagi.yuta@ohsuga.lab.uec.ac.jp

Ryohei Orihara
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
orihara@acm.org

Yasuyuki Tahara
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
tahara@uec.ac.jp

Yuichi Sei
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
seiuny@uec.ac.jp

Tanel Alumae
Department of Software Science
Tallinn University of Technology
Tallinn, Estonia
tanel.alumae@taltech.ee

Akihiko Ohsuga
Department of Informatics
The University of Electro-Communications
Tokyo, Japan
ohsuga@uec.ac.jp

***Abstract*—On the one hand, the development of speech synthesis technology has made it possible to produce more natural-sounding voices. On the other hand, there are threats of deepfake voices that impersonate real people and make fake claims. Until now, ASVspoof has considered biometric measures. However, they added a new measure for social media for the first time in 2021. In this study, we tested whether the latest deepfake voice detection model could detect deepfake voices caused by the most recently proposed text-to-speech. Experimental results showed that the model misclassified almost all deepfake voices as bona fide voices. In the future, the effect of speech categorization, which also considers speech content, will be investigated regarding factors specific to deepfake voices.**

***Index Terms*—datasets, neural networks, voice processing, natural language processing, speech synthesis**

I. Introduction

Speech synthesis technology has developed significantly over the last few years [1]–[3]. On the one hand, this impact has made it much easier for anyone to obtain more authentic voices. On the other hand, there are fears of deepfake voice(a.k.a. fake voice, spoofing) attacks by users with malicious intent to deceive others. So far, fake news has been a means of deceiving others, and research conducted with the aim of early detection [4]. The deepfakes can shape a multimodal version of fake news and, like fake news, need to be detected before they are widely shared. Indeed, Figure 1 shows an example of the deepfake. The social media users widely shared a video of Ukrainian President Zelensky calling for surrender, but this was fake, using the President's previous videos and audio [5]. Hence, there has been widespread research into countering deepfake voices [6]–[9].

ASVspoof is the most popular initiative to make anti-spoofing methods [7], [10]–[12]. This initiative has continued every other year since 2015, with many participants each time proposing models to classify the deepfake voice(they call them spoofing) and bona fide provided. However, they initially intended to use deepfake voices against biometric systems, adding to the tasks of targeting users on social media in 2021 [12]. Specifically, in 2015, they held the first time, generated two types of speech, text-to-speech (TTS) and voice conversion, and played back over a telephone line [13]. In 2017, they provided a dataset that envisaged an attempt to break through biometrics by playing back a pre-recorded real voice [10]. In 2019, the first scenario was named Logical Attack (LA), and the second was named Physical Attack (PA), with new methods and environments added [11]. In 2021, they added a new DeepFake (DF) task to the LA, which involved the classification of audio played in an online environment [12]. They also observed that generalization performance was lacking in common with the methods that participated in the DF task. When trained on a training set consisting of methods up to 2019, they could not correctly classify a validation set consisting of methods from 2020 onwards [12]. Therefore, this paper experiments explicitly on how well the classification models trained on the ASVspoof 2021 DF dataset classify deepfake voices from the recently proposed TTS method. As a result, this paper reports that the classification model failed to detect the deepfake voices at almost all. In response, this paper discusses how we should detect deepfake voices on social media in the future.

Henceforth, this paper will describe the contents in the following order. Section II introduces the related research,

Fig. 1: The corresponding of Ukrainian President Zelensky with the deepfake video [14].

which focuses on the automatic detection of deepfake voices and the developments of TTS. After that, we introduce the models which generate and classify deepfake voices in Section III. We explain the process of the experiment, like generating a deepfake voices dataset and setting up the models in Section IV. We report the result of the experiment in Section V. Finally, we discuss the result from two perspectives in Section II and we conclude in Section VII.

II. Related researches

A. Automatic detection of deepfake voices

There are many attempts for the countermeasures of speech synthesis [6], [15], [16], voice conversion [17]–[20], and replay attacks [21] before 2013. However, a research team pointed out in 2013 the necessity of making standard datasets, protocols, and metrics [22], [23]. After then, the ASVspoof challenge proposed in 2015 [7], [13]. This project is the first initiative for making a shared dataset, evaluation protocols, and metrics. The first one has the Logical Access(LA) task, which focuses on spoofing the telephony environment. The next one, ASVspoof 2017, focuses on Physical Access(PA). This task contemplates a replay attack using audio recorded from the target [10]. They updated these two tasks in ASVspoof 2019 [11]. In these tasks, they added new voice generation models and environments for replaying. In the latest round in 2021, the DeepFake (DF) task: classification of voices in the online environment, was added. In particular, they noted that in 2021, generalization performance is insufficient [12]. This trend is because the models trained by the methods up to 2019 could not correctly classify the deepfake voices from the new methods later in 2020. Furthermore, the extended shared task from ASVspoof is proposed as the Audio Deepfake Detection by the end of 2021 [24].

B. Developments of text-to-speech

Text-to-speech(TTS) methods are used in ASVspoof 2015 as the speech synthesis [13], [25]. In the ASVspoof 2019, many neural-network-based TTS are added [26]. So far, the most common method to generate speech from preprocessed sentences has been to extract acoustic [2], [27] and linguistic features [28] by one model and then generate output waveforms by another model [28], [29]. The two-stage pipeline structure needs sequential training or fine-tuning [2], [30]. The wave generation model needs to wait for the training of the feature extract one. There are some researches to make end-to-end models to reduce the training cost and capture hidden representations for outputs [1], [31], [32]. The performance of end-to-end models is also improving [33].

III. Methodology

We explain the method of generating and classifying deepfake voices. According to the literature, both are existing methods whose results are said to be good.

A. Generating deepfake voices

In generating deepfake voices, we used the Variational Inference with adversarial learning for the end-to-end Text-to-Speech(VITS) model [33]. The end-to-end text-to-speech model consists of the Conditional Variational AutoEncoder (CVAE), alignment estimation by variational inference, and adversarial training. The VITS can be expressed as a CVAE to maximize the variational lower bound, also called the evidence lower bound (ELBO), of the intractable marginal log-likelihood of data: $\log p_\theta(x|c)$ [33]:

$$\log p_\theta(x|c) \geq \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z) - \log \frac{q_\phi(z|x)}{p_\theta(z|c)}\right] \quad (1)$$

where $p_\theta(x|c)$ denotes a prior distribution of the latent variables z when condition c is given, $p_\theta(x|z)$ shows the likelihood function of a data point x, and $q_\phi(z|x)$ is an approximate posterior distribution [33]. The training loss is the negative value of ELBO, which is calculated by the sum of reconstruction loss $-p_\theta(x|z)$ and KL divergence $q_\phi(z|x) - p_\theta(z|c)$, where $z \sim q_\phi(z|x)$ [33]. Please see the reference for detailed calculation methods for each loss.

B. Classifying the voices

In classifying deepfake voices, we used RawNet2 [34] and RawBoost [35].

1) RawNet2: The RawNet2 is an improved method from the RawNet [36]. Both models are end-to-end classifying ones by integrating the extracting utterance-level features and the feature enhancement phase [36]. Table I shows the structure of RawNet2. The first layer is a sinc-convolution layer from SincNet [37], [38]. The SincNet has a Convolutional Neural Network(CNN) structure, and it filters the raw waveform by a bank of band-pass filters. The model sets the filters in the shape of sinc functions. The second is the residual block. This consist of batch normalisation(BN), LeakyReLU [39], convolutional layer, max-pooling, and Feature Map Scaling(FMS). The FMS is proposed in [34], it behaves like an attention layer with a sigmoid activation function [34]. The output layer implements for binary classification of deepfake voice or bona fide.

The ASVspoof applied the model in the ASVspoof 2019 LA task [40]. The combined method with a baseline performed the second-best result in the ASVspoof 2019 LA task [40].

TABLE I: The RawNet2 architecture which is applied for ASVspoof 2019 [40] and 2021 [12].

Layer	Input	Output shape
Fixed Sinc filters	Conv(129,1,128) Maxpooling(3) BN & LeakyReLu	(21290, 128)
Res block	BN & LeakyReLU Conv(3,1,128) BN & LeakyReLU Conv(3,1,128) Maxpooling(3) FMS	(2365, 128)
Res block	BN & LeakyReLU Conv(3,1,128) BN & LeakyReLU Conv(3,1,128) Maxpooling(3) FMS	(29, 512)
GRU	GRU(1024)	(1024)
FC	1024	(1024)
Output	1024	2

Moreover, it also ranked 24th out of 41 participants in the ASVspoof 2021 LA task [12].

2) RawBoost: RawBoost is a data boosting and augmentation method that operates at the raw waveform level [35]. It implements noise addition as a data augmentation in three forms: (1) linear and non-linear convolutive noise, which reproduces the noise from encoding, compression, and transmission, (2) impulsive signal-dependent additive noise like clipping, non-optimal device (microphones and amplifiers) operation, Etc., (3) stationary signal-independent additive noise by applying a single finite impulse response filter [35].

IV. Experiment

A. Making a dataset of deepfake voices

1) News dataset: We had to make our deepfake voices by the latest text-to-speech model from fake news. We used the news from the MuMiN dataset [41]. This dataset contains multilingual news which shares fake or real news. Table II shows the statistics of the dataset. We got 465 news articles in English. Some of the articles obtained were confirmed as fake news, while others appeared to be factual. This dataset may be insufficient amount to fine-tune the models. Therefore we are still considering adding more fake/factual news articles from other datasets. We are also examining adding not only the news themselves but also the tweets that propagate them. We limit longer articles to 480 words to ensure that the audio is not too long.

TABLE II: Statistics of our dataset.

statistics	value
Number of news articles	465
Upper limit of words	480
Average num. of words	342
Ave. duration of voice [s]	118.7

2) Generating voices by text-to-speech: They employed speech production models(text-to-speech and voice conversion) proposed before 2019 in the ASVspoof 2021 DF dataset [12]. According to the report, they did not use the model since 2020 in training datasets because they wanted to check the generalization performance. Indeed, it reports that the models of participants remained challenged in generalization performance. However, the report did thinly analyze the classification results for each voice generation model. Therefore, we checked the performance of the models with the latest text-to-speech model. We selected the VITS [33] model for deepfake voice generation. We used the pre-trained model for the generation by the VCTK corpus [42].

B. Classifications models

We selected two models from ASVspoof 2021. One is the RawNet2 [40], another is the RawBoost [35]. RawNet2 is one of the baseline models in the ASVspoof 2021 DF

RawNet2 through data augmentation [35]. They described that they could implement the model for other detection tasks.

C. Experiment process

First, we checked the reproducibilities of the two models by the ASVspoof 2021 DF dataset. We used Equal Error Rate(EER) as a reference index. After that, we got outputs of the 234 deepfake voices by RawNet2 and RawBoost. We refrained from using the EER because all voices are fake.

V. Result

Table III, IV and Figure 3 shows the result of the experiment. Table III shows the EER values of the models. In the RawNet2, they are close to the ones in the report article of the ASVspoof 2021 [12]. In the RawBoost, the performance of the ASVspoof DF task was worse than RawNet2. This trend may be because they did not initially develop the model for the task. We used the threshold value in the EER to evaluate our proposed dataset.

Table IV shows the output values of the models. The common point of both models is that the output value shows the suspiciousness of the voice. According to the results of the ASVspoof dataset, we interpret values higher than the threshold value in the EER in the ASVspoof dataset as the deepfake voices. Compared to the ASVspoof dataset from Figure 3, our provided voices have a much lower detection rate of deepfake voices. This result is a severe problem, as the provided one does not contain a single authentic voice. We can confirm a similar trend from the mean and median values in the table. For example, the average and median values of the provided ones are also smaller than the values in ASVspoof 2021 DF task. Table V shows the number of deepfake voices whose outputs are higher than the threshold. The dataset of the ASVspoof sets the threshold. If the output value exceeds the threshold, we can regard the voice as a deepfake. The RawNet2 detects five voices as the deepfake. The RawBoost detects 32 voices as the deepfake, but it is unreliable because the EER of RawBoost is small. This result means that both

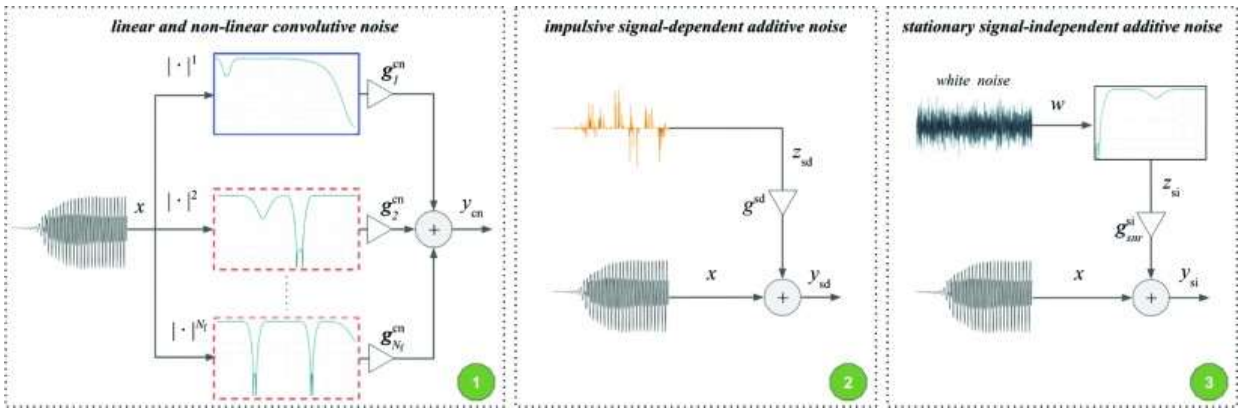


Fig. 2: The three ways of the data argumentation in the RawBoost [35].

models detected less than 10% of the total of our provided dataset as the deepfake. We confirmed that there are no specific fatal artifacts from listening to the audio, which classified the obvious deepfake. In summary, there are differences in the range of values. However, both have little doubt about our deepfake voices.

TABLE III: The EER and the threshold values of the classification models for the ASVspoof 2021 DF dataset.

Model Name	RawNet2	RawBoost
Equal Error Rate(EER)[%]	25.5	81.0
Threshold in EER	-5.74	2.77×10^{-6}

VI. Discussion and future perspective

A. Generalization performance

According to Table IV, the lack of generalization performance is a critical problem. RawNet2, in the middle of the pack in ASVspoof, does not detect the latest text-to-speech deepfake voices. So is RawBoost. This result means the model trained by the dataset with text-to-speech and voice conversion models before 2019 cannot compete with the other models from 2020 onward. The report of ASVspoof 2021 has already raised the generalization performance issues [12]. However, as we know, this is the first time we have obtained similar results for a specific post-2020 methodology. We cannot check the performance of top-rated models in ASVspoof 2021. The best model may detect the latest deepfake voices. Nonetheless, it is unsure if it can handle future developments when developers propose the brand-new voice-generating structure.

B. Measures to take

Deceiving users is the purpose of deepfake voices on social media. Therefore, if they seek to assert their claims, the content of the deepfake voices should also be unique. We think what they say is more important to detect them than the wave features. Hence, in the future, we aim to detect deepfake voices by considering speech content.

VII. Conclusion

The progress of speech processing gives us the benefit of getting natural voices. It also means we must be careful if it is bona fide or deepfake voices. Some attempts to spot them by using wave features of the speech. However, the trained models by old voice-generating models are challenging to identify the deepfake voices that the latest models provide. If this trend continues in the future, detection within the existing framework will keep fraught with dangers. Therefore alternative detection scenarios need to be considered. We intend to propose new detection methods that focus on what they are trying to tell us.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP21H03496, JP22K12157.

References

[1] Y. Wang, R. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. Le, Y. Agiomyrgiannakis, R. Clark, and R. A. Saurous, "Tacotron: Towards End-to-End Speech Synthesis," in *Proc. Interspeech 2017*, 2017, pp. 4006–4010.

[2] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerrv-Ryan, R. A. Saurous, Y. Agiomvrgiannakis, and Y. Wu, "Natural tts synthesis by conditioning wavenet on mel spectrogram predictions," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4779–4783.

[3] T. Wang, R. Fu, J. Yi, J. Tao, Z. Wen, C. Qiang, and S. Wang, "Prosody and voice factorization for few-shot speaker adaptation in the challenge m2voc 2021," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8603–8607.

[4] Y. Yanagi, R. Orihara, Y. Sei, Y. Tahara, and A. Ohsuga, "Fake news detection with generated comments for news articles," in *2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES)*, 2020, pp. 85–90.

[5] B. Otte, "No, president zelenskyy did not tell ukrainian soldiers to surrender," Apr 2022. [Online]. Available: https://www.poynter.org/fact-checking/2022/no-president-zelenskyy-did-not-tell-ukrainian-soldiers-to-surrender/

[6] P. L. De Leon, M. Pucher, J. Yamagishi, I. Hernaez, and I. Saratxaga, "Evaluation of speaker verification security and detection of hmm-based synthetic speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 8, pp. 2280–2290, 2012.

TABLE IV: The statistics of output values a.k.a. suspiciousness from two voice datasets: DF scenario of ASVspoof 2021 and fake news from MuMiN with VITS.

Dataset name	Index	RawNet2	RawBoost
ASVspoof 2021 DF	min.	-9.11	1.5×10^{-7}
	max.	0.00	1.00
	ave.	-6.31	0.06
	med.	-7.57	1.02×10^{-6}
deepfake voices from fake news in MuMiN	min.	-8.67	1.5×10^{-7}
	max.	0.00	1.00
	ave.	-7.88	2.3×10^{-3}
	med.	-8.05	1.05×10^{-6}

TABLE V: The number of deepfake voices whose output value is greater than the threshold value. The threshold values are those of the situation when calculating the EER.

Index	RawNet2	RawBoost
Greater than the threshold	5	32
Percentage of total[%]	1.08	6.88

REFERENCES

[7] Z. Wu, J. Yamagishi, T. Kinnunen, C. Hanilc¸i, M. Sahidullah, A. Sizov, N. Evans, M. Todisco, and H. Delgado, "Asvspoof: The automatic speaker verification spoofing and countermeasures challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 588–604, 2017.

[8] T. Chen, A. Kumar, P. Nagarsheth, G. Sivaraman, and E. Khoury, "Generalization of Audio Deepfake Detection," in *Proc. The Speaker and Language Recognition Workshop (Odyssey 2020)*, 2020, pp. 132–137.

[9] H. Ma, J. Yi, J. Tao, Y. Bai, Z. Tian, and C. Wang, "Continual Learning for Fake Audio Detection," in *Proc. Interspeech 2021*, 2021, pp. 886–890.

[10] T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, and K. A. Lee, "The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection," in *Proc. Interspeech 2017*, 2017, pp. 2–6.

[11] M. Todisco, X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. H. Kinnunen, and K. A. Lee, "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection," in *Proc. Interspeech 2019*, 2019, pp. 1008–1012.

[12] J. Yamagishi, X. Wang, M. Todisco, M. Sahidullah, J. Patino, A. Nautsch, X. Liu, K. A. Lee, T. Kinnunen, N. Evans, and H. Delgado, "ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection," in *Proc. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*, 2021, pp. 47–54.

[13] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilc¸i, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge," in *Proc. Interspeech 2015*, 2015, pp. 2037–2041.

[14] D. Evon, "Bad deepfake of zelenskyy shared on ukraine news site in reported hack," Mar 2022. [Online]. Available: https://www.snopes.com/news/2022/03/16/zelenskyy-deepfake-shared/

[15] P. L. De Leon, I. Hernaez, I. Saratxaga, M. Pucher, and J. Yamagishi, "Detection of synthetic speech for the problem of imposture," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4844–4847.

[16] Z. Wu, X. Xiao, E. S. Chng, and H. Li, "Synthetic speech detection using temporal modulation feature," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 7234–7238.

[17] Z. Wu, E. S. Chng, and H. Li, "Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition," in *Proc. Interspeech 2012*, 2012, pp. 1700–1703.

[18] Z. Wu, T. Kinnunen, E. S. Chng, H. Li, and E. Ambikairajah, "A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012, pp. 1–5.

[19] F. Alegre, R. Vipperla, A. Amehraye, and N. Evans, "A new speaker verification spoofing countermeasure based on local binary patterns," in *Proc. Interspeech 2013*, 2013, pp. 940–944.

[20] F. Alegre, A. Amehraye, and N. Evans, "Spoofing countermeasures to protect automatic speaker verification from voice conversion," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 3068–3072.

[21] J. Villalba and E. Lleida, "Preventing replay attacks on speaker verification systems," in *2011 Carnahan Conference on Security Technology*, 2011, pp. 1–8.

[22] N. Evans, T. Kinnunen, and J. Yamagishi, "Spoofing and countermeasures for automatic speaker verification," in *Proc. Interspeech 2013*, 2013, pp. 925–929.

[23] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification," *Speech Commun.*, vol. 66, no. C, p. 130–153, feb 2015. [Online]. Available: https://doi.org/10.1016/j.specom.2014.10.005

[24] J. Yi, R. Fu, J. Tao, S. Nie, H. Ma, C. Wang, T. Wang, Z. Tian, Y. Bai, C. Fan, S. Liang, S. Wang, S. Zhang, X. Yan, L. Xu, Z. Wen, and H. Li, "Add 2022: the first audio deep synthesis detection challenge," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 9216–9220.

[25] J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, and J. Isogai, "Analysis of speaker adaptation algorithms for hmm-based speech synthesis and a constrained smaplr adaptation algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 66–83, 2009.

[26] X. Wang, J. Yamagishi, M. Todisco, H. Delgado, A. Nautsch, N. Evans, M. Sahidullah, V. Vestman, T. Kinnunen, K. A. Lee, L. Juvela, P. Alku, Y.-H. Peng, H.-T. Hwang, Y. Tsao, H.-M. Wang, S. L. Maguer, M. Becker, F. Henderson, R. Clark, Y. Zhang, Q. Wang, Y. Jia, K. Onuma, K. Mushika, T. Kaneda, Y. Jiang, L.-J. Liu, Y.-C. Wu, W.-C. Huang, T. Toda, K. Tanaka, H. Kameoka, I. Steiner, D. Matrouf, J.-F. Bonastre, A. Govender, S. Ronanki, J.-X. Zhang, and Z.-H. Ling, "Asvspoof 2019: A large-scale public database of synthesized, converted and replayed speech," *Computer Speech & Language*, vol. 64, p. 101114, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230820300474

[27] H. Ze, A. Senior, and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 7962–7966.

[28] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," in *Proc. 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9)*, 2016, p. 125.

[29] N. Kalchbrenner, E. Elsen, K. Simonyan, S. Noury, N. Casagrande, E. Lockhart, F. Stimberg, A. van den Oord, S. Dieleman, and K. Kavukcuoglu, "Efficient neural audio synthesis," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 2410–2419. [Online]. Available: https://proceedings.mlr.press/v80/kalchbrenner18a.html

[30] R. J. Weiss, R. Skerry-Ryan, E. Battenberg, S. Mariooryad, and D. Kingma, "Wave-tacotron: Spectrogram-free end-to-end text-to-speech synthesis," in *ICASSP*, 2021. [Online]. Available: https://arxiv.org/abs/2011.03568

[31] Y. Ren, C. Hu, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T.-Y. Liu, "Fastspeech 2: Fast and high-quality end-to-end text to speech," in

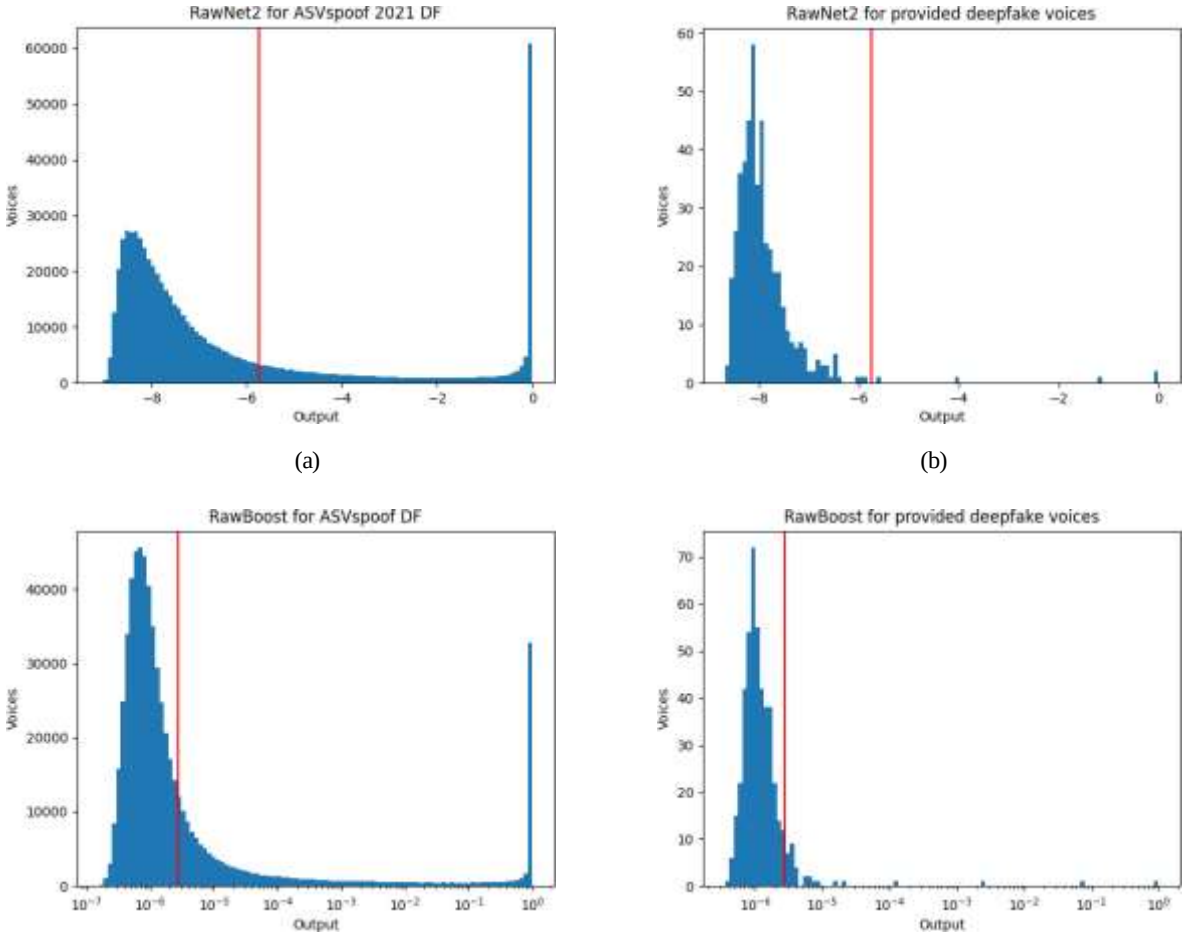


Fig. 3: The histogram of output values by (a) RawNet2 for ASVspoof 2021 DF, (b) RawNet2 for our provided deepfake voices, (c) RawBoost for the ASVspoof, and (d) RawBoost for the provided. The red lines mean the threshold value in the EER situation of the ASVspoof dataset.

International Conference on Learning Representations, 2021. [Online]. Available: https://openreview.net/forum?id=piLPYqxtWuA

[32] J. Donahue, S. Dieleman, M. Binkowski, E. Elsen, and K. Simonyan, "End-to-end adversarial text-to-speech," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=rsf1z-JSj87

[33] J. Kim, J. Kong, and J. Son, "Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 5530–5540. [Online]. Available: https://proceedings.mlr.press/v139/kim21f.html

[34] J. weon Jung, S. bin Kim, H. jin Shim, J. ho Kim, and H.-J. Yu, "Improved RawNet with Feature Map Scaling for Text-Independent Speaker Verification Using Raw Waveforms," in *Proc. Interspeech 2020*, 2020, pp. 1496–1500.

[35] H. Tak, M. Kamble, J. Patino, M. Todisco, and N. Evans, "Rawboost: A raw data boosting and augmentation method applied to automatic speaker verification anti-spoofing," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 6382–6386.

[36] J. weon Jung, H.-S. Heo, J. ho Kim, H. jin Shim, and H.-J. Yu, "RawNet: Advanced End-to-End Deep Neural Network Using Raw Waveforms for Text-Independent Speaker Verification," in *Proc. Interspeech 2019*, 2019, pp. 1268–1272.

[37] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with sincnet," in *2018 IEEE Spoken Language Technology Workshop (SLT)*, 2018, pp. 1021–1028.

[38] ——, "Learning Speaker Representations with Mutual Information," in *Proc. Interspeech 2019*, 2019, pp. 1153–1157.

[39] A. L. Maas, A. Y. Hannun, A. Y. Ng *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1. Atlanta, Georgia, USA, 2013, p. 3.

[40] H. Tak, J. Patino, M. Todisco, A. Nautsch, N. Evans, and A. Larcher, "End-to-end anti-spoofing with rawnet2," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 6369–6373.

[41] D. S. Nielsen and R. McConville, "Mumin: A large-scale multilingual multimodal fact-checked misinformation social network dataset," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. ACM, 2022.

[42] J. Yamagishi, C. Veaux, and K. MacDonald, "Cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit (version 0.92)," 2019.

A Multibody Dynamics Approach to Predict the Gear Shift Force

Salah A Sabri
Eaton India Innovation Centre,
Eaton Technology,
Pune, India
salahasabri@eaton.com

Sachin Ahirrao
Eaton India Innovation Centre,
Eaton Technology,
Pune, India
sachinahirrao@eaton.com

Bruno Mussulini
Eaton Transmission,
Eaton Technology,
SP, Brazil
brunocmussulini@eaton.com

Rafael Garcia
Eaton Transmission,
Eaton Technology,
SP, Brazil
rafaelhgarcia@eaton.com

Carlos Sena
Eaton Transmission,
Eaton Technology,
SP, Brazil
carloshsena@eaton.com

Guilherme Biagio
Eaton Transmission,
Eaton Technology,
SP, Brazil
guilhermegbiagio@eaton.com

***Abstract*— Gear shift quality is the assessment of a driver's perception of the gear knob response when shifting up or down a gear. Due to customer's expectations and awareness of drive comfort, this subject has received increased attention over the past decades in the automotive industry. Manufacturers are searching for methodologies to best attend and surpass the customer's expectations. One of the most important driving parameters of the gear shift quality comes from the variation of gear shift forces over the entire shift travel event. This majorly consists of the highest force, the lowest force, and the force characteristics during transition between highest and lowest forces. A transmission gear shift system comprises of different subsystems such as control, detent, rails with interlocks and synchronizer systems. These subsystems contribute towards the gear shift force variation.**

A novel methodology to predict the dynamic forces during gear shifting is presented in this paper. Multibody dynamics (MBD) model of the gear shift system with transmission is developed. Detailed modeling comprising of all the subsystems is carried out in the virtual environment. This captures contribution of individual subsystems on total shift forces and subsequently gear shift quality. Simulation predicts the gear shift force of different gear pairs for the given displacement of shift lever as an input which is then validated with test data. This methodology shall be used to improve shift quality by enhancing the performance at subsystem level. This will also assist in reducing design lead time and minimize testing along with prototype building cost.

Keywords*— *Gear Shift System, Shift force, MBD, Showroom Shifting

I. INTRODUCTION

Shift system in the transmission is required to change gears. This gear shifting is performed by applying a force on the gear knob. This shift force is essential for two purposes, first of which is to prevent the auto engagement and disengagement of gears due to vibration, jerk, or sudden change in the vehicle condition. Secondly, it gives the driver a feel of gear shifting. However, it is of equal importance to keep the effort applied for gear shifting low for a smooth and convenient shift. Consequently, a compromise is made between the two conditions by optimizing the shift subsystem parameters. Kent [1] gives an overview of Gear shift quality and parameters affecting it. Dynamic model of shift system and other transmission parts using MATLAB/SIMULINK was developed to assess the effect of each subsystem on gear shift force. Davis et al [2] developed a methodology to predict the gear shift force by building a dynamic model of the shift system and transmission components in Dymola and its parameters were established by correlating with test data. The shift parameters were optimized using Ricardo OGAS software. More et al [3] discussed various dimensions of gear shift quality and parameters affecting it. Experimental method of data gathering, and data analysis were also discussed, and gear shift quality was defined in measurable number. Singh et al. [4] proposed methodology to predict shifting and selecting forces by building rigid body shift tower model using multi-body dynamic tool. However little attention has been paid to investigate dynamic behavior of shift system along with complete transmission to simulate the shift forces.

The present work focusses on developing a methodology to predict gear shift forces for showroom shifting condition using multibody dynamics analysis approach. The virtual model of the shift system and the complete transmission is built in ADAMS MBD environment. The simulations are performed for shifting different gear pairs in idle transmission condition.

II. PROBLEM DESCRIPTION

The forces experienced during a gear shift serves two purposes. It prevents the transmission from self-shifting, and it also provides confirmative predictability of gear shifting to the

driver. However, if the shift forces are too high, the driver might find it difficult and harsh. This shift force requirement varies with transmission types, applications, demographic conditions etc. The problem discussed in this paper deals with developing a methodology to predict shift force when the engine is running under no load idle condition [5]. The shift force is an indicator of shift quality and can be tested on idle vehicle. In the present work, total shift force is predicted, and the contribution from various subsystems are also evaluated. MSC ADAMS software is used to develop detailed model of transmission.

A. *Selection*

Selection is the process of changing gear pair without engaging the gears. In the given example there are 3 selection gear pairs i.e., 3-4, 5-6 and 1-2 gear pairs with an additional reverse gear mode. Selection operation is performed by pushing the knob on the selector arm as shown in "Fig. 1" and "Fig. 5".

B. *Shifting*

Shifting is the process of engaging the gear within the selected pairs. In the given example shifting to 3rd or 4th gear can be done directly from the neutral position. For other gear, the particular gear pair need to be selected and shifting is performed as shown in "Fig. 1" and "Fig. 5".

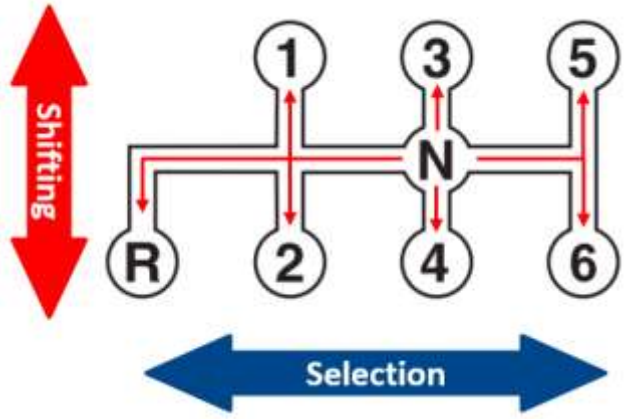


Fig. 1. Selection and shifting of gears. There are 3 selection pairs and for each selection pair, two gear shifting can be performed.

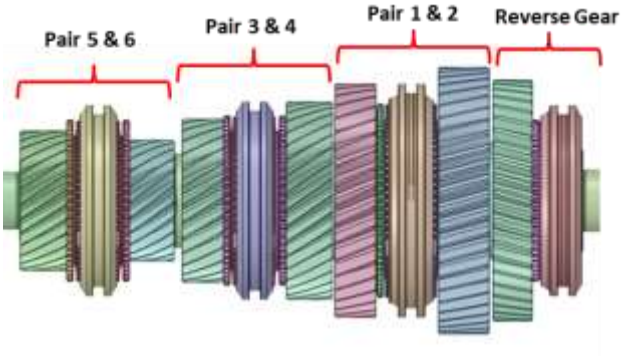


Fig. 2. Gear pairs of gearbox for selection and shifting. Each gear pair have 1 synchronizer assembly responsible for shifting.

III. METHODOLOGY DESCRIPTION

A. *Modelling*

The virtual model is the replica of an actual vehicle. The rigid body model of transmission and its component should exhibit the same dynamic behavior as the physical one. Rigid body model of complete transmission including shift system, synchronizer assembly and interlocking mechanism is built in ADAMS. In rigid body dynamics, the governing dynamic equations are developed using augmented Newton-Euler formulation with Lagrange multipliers. We can write the system equation of motion of the rigid body [6]

$$\begin{bmatrix} M & C_q^T \\ C_q & 0 \end{bmatrix} \begin{bmatrix} \ddot{q} \\ \lambda \end{bmatrix} = \begin{bmatrix} Q_e + Q_v \\ Q_c \end{bmatrix} \quad (1)$$

Here, M represents Mass matrix of rigid bodies, $\ddot{q}$ represents acceleration vector, λ is the vector of Lagrange multipliers, Q_e is the vector of externally applied forces and Q_v is quadratic velocity vector, C_q^T is constraint Jacobian Matrix. The equation is a system of algebraic equations that can be solved for the acceleration vector and vector of Lagrange multipliers. This equation (1) is solved to obtain the accelerations and the Lagrange multipliers. The position and velocity of the individual internal components are obtained by numerical integration.

To simplify the analysis only relevant parts are modelled while the parts having same motion are combined. Springs are modelled, as active force element and its inertia effects aren't captured in the analysis. Translation joint is added with a spring to prevent its transverse movement. Springs are modelled in detent system which is charged with preload and stiffness.

To perform a shift, the shift lever is rotated which pushes the shift selector block to translate the main rail as shown in "Fig. 4". Selection is achieved by rotating the select lever which further rotates the shift selector block about main rail axis and main rail also rotates as shown in "Fig. 4".

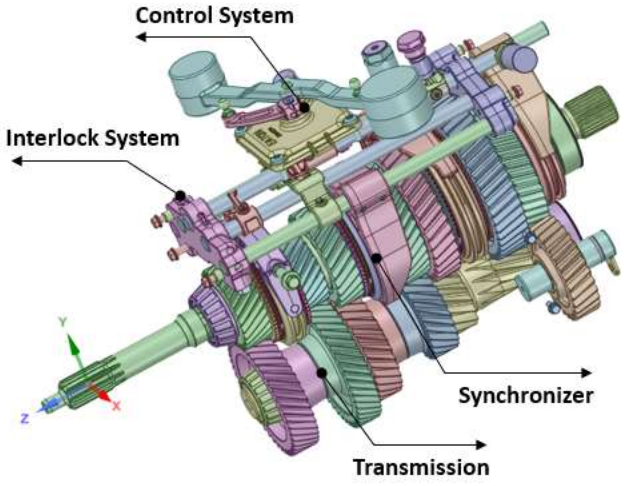


Fig. 3. Complete transmission along with subsystems are modelled in multibody dynamics (MBD) virtual environment

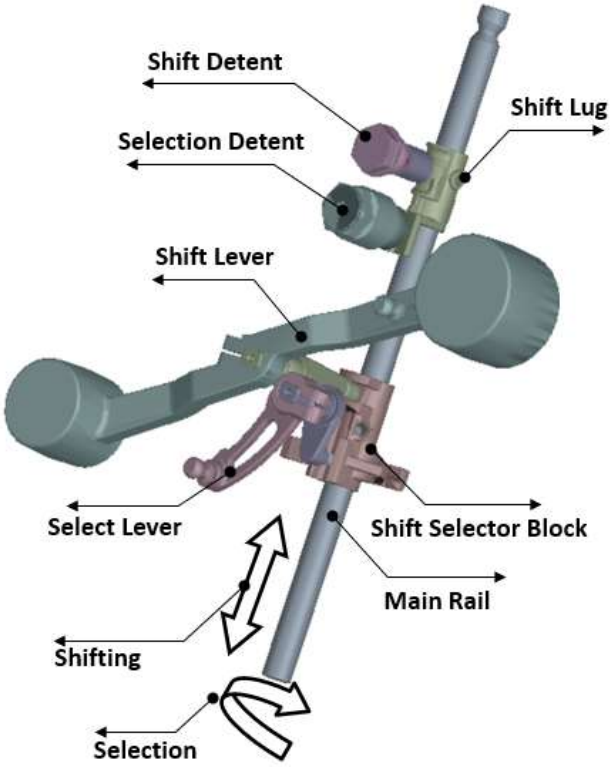


Fig. 4. Shift system of a transmission comprises of detent system, rails, control system along with its components.

The detent system is charged with spring load. When the main rail is translated for the shifting or rotated for the selection, there is a resistance force between detent CAM and ball, hence the driver experience the shifting effort.

If two rails move together due to improper selection, the interlocking mechanism will lock the travel of rail and simultaneous shifting of two gears won't take place. Simplified rigid body model of transmission consists of input shaft, counter shaft, main shaft, gears and synchronizer assembly. Analysis is performed for all gear set but result is included for 3-4th gear shifting cycle. Synchronizer strut pre-energizes the synchro ring in the axial movement and the resistance offered contributes to the total force while shifting the gear through the shift lever

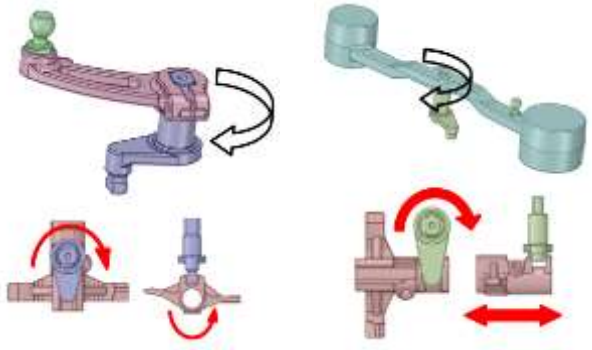


Fig. 5. Selector (left) and shifter (right) assembly are part of control system used for selection and shifting respectively.

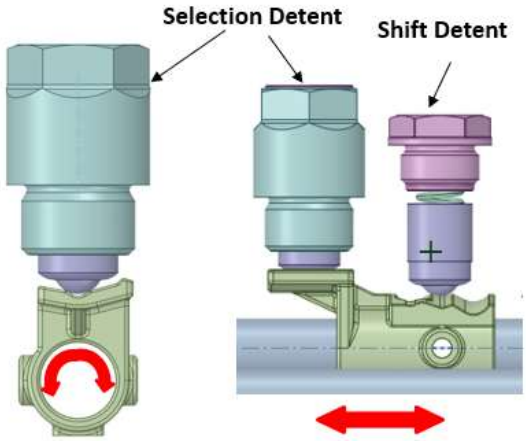


Fig. 6. Detent system with shift and selection detent offers resistance to selection and shifting through charged spring and CAM profile

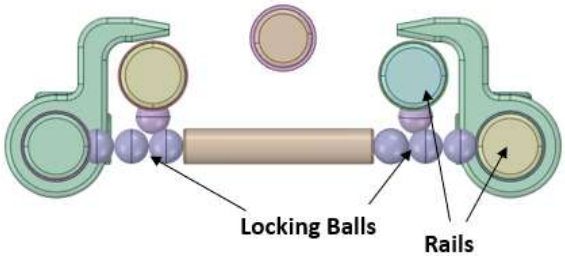


Fig. 7. Interlock system prevents the simultaneous movement of multiple rails.

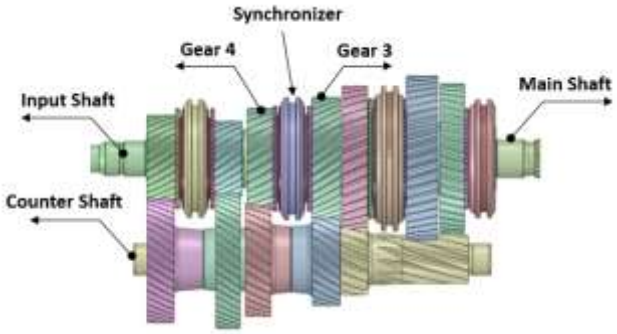


Fig. 8. Transmission comprises of input shaft, counter shaft, main shaft and different gears for speed and torque control.

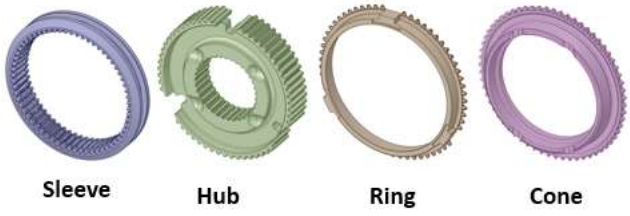


Fig. 9. Synchronizer parts are responsible for smooth shifting of gear.

B. Test Condition and Simulation

Testing is conducted on the vehicle in IDLE condition which is also called showroom shift condition. The transmission along with the shift system is mounted on a test bench. There is no external excitation to the transmission either from the dynamometer or from the engine. The shift force is measured at the shift pin location and the selection force is measured at pin location on selection lever as shown in "Fig. 5". Virtual environment for simulation is prepared to replicate the physical testing condition in ADAMS environment.

C. Results

The shift force is plotted against the shift lever angle for complete shift cycle of 3-4 gears. The shift angle is proportional to the axial displacement of main rail or shift detent. Initially the shift rail moves from neutral position towards 3rd gear and gets engaged. The shift lever is moved back to disengage from 3rd gear to come to neutral position. This completes one cycle of engagement and disengagement for 3rd gear and the same is repeated for 4th gear. The force experienced in the complete cycle of engagement and disengagement can be understood with the help of 6 regions in the curve as depicted in "Fig. 11"

- **Peak Load during Engagement (1):** Initially when the force is applied, it must overcome the preloads of the springs within the system. The detent ball ramps up over the CAM slope. Normal as well as frictional force opposes the detent ball movement. This corresponds to the region on the CAM slope where first linear slope is about to end as depicted in the "Fig. 10".
- **Decreasing Load during Engagement (2)**: Load decreases with decreasing slope. At maximum spring compression condition, only frictional force opposes the motion as depicted in the "Fig. 10".
- **Auto Pull during Engagement (3)**: The direction of the resultant contact force by CAM on the detent ball in the axial direction is same as the direction of motion of detent ball. In this region external force is not needed to move the detent and auto pull phenomenon is observed. This give feeling of comfort at hand of driver as depicted in the "Fig. 10".
- **Peak Load during Disengagement (4):** When disengagement starts, the detent ball has to ramp up over the slope. Normal as well as frictional force opposes the motion of detent ball. Hence the load is maximum as depicted in the "Fig. 10".
- **Decreasing load during Disengagement (5)**: Load decreases with decreasing slope as depicted in the "Fig. 10".
- **Auto Pull during Disengagement (6)**: The similar behaviour is exhibited as region 3 during disengagement as depicted in the "Fig. 10".

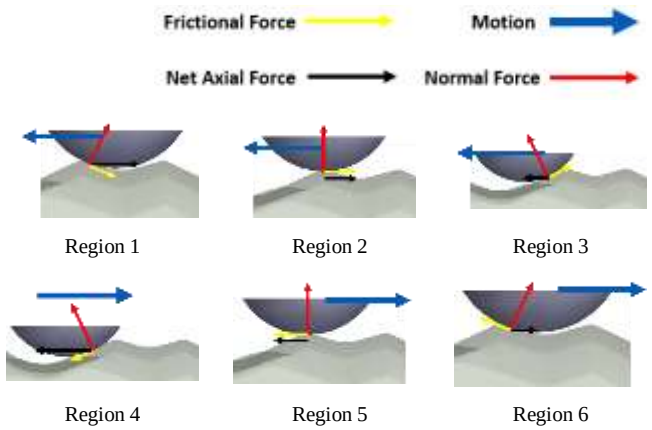


Fig. 10. Contact Force in different region during engagement and disengagement. Regions 1 and 4 offers highest resistance to the motion, regions 2 and 5 depict decreasing load whereas regions 3 and 6 are auto pull region.

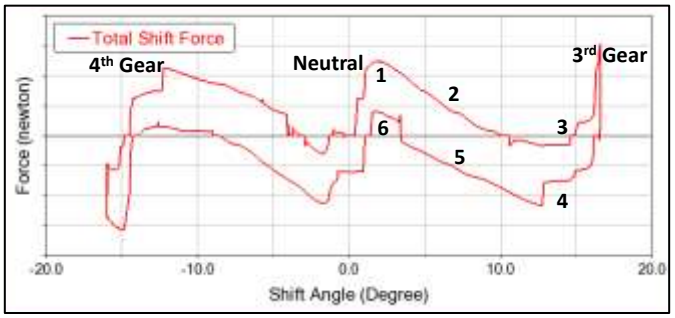


Fig. 11. Total shift force on shift lever for engagement and disengagement for complete 3rd and 4th gear cycle.

D. Contribution from Subsystems

The total force required for shifting comes from different subsystems. These subsystems are shift detent system, synchronizer system and interlock system. There is considerable contribution of friction as well. The force in these subsystems depends on various parameters such as slopes, spring specifications, contact friction etc. as shown in "Fig. 12" and "Fig. 13".

- **Detent System:** Detent system is the major contributor of the force to the total shift force. The difference in the force during engagement and disengagement is hysteresis loss
- **Synchronizer System**: Along with detent system, synchronizer system is also one of the major contributors of force
- **Interlock System:** It is used primarily for interlocking but it also offers resistance and contributes to total force
- **Forces due to Friction:** These forces come mainly from rail, support and other contact interfaces, hence contribute to total shift force

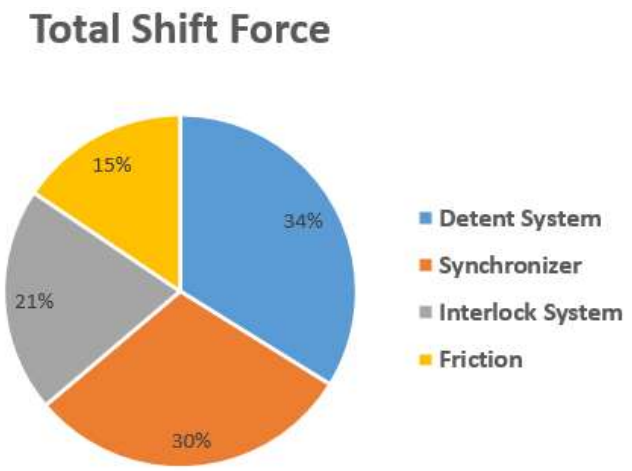


Fig. 12. Contribution of different subsystems towards the total shift force at peak load condition during engagement. Detent system and synchronizer are major contributor to total force, whereas share of interlocking system and friction are also significant.

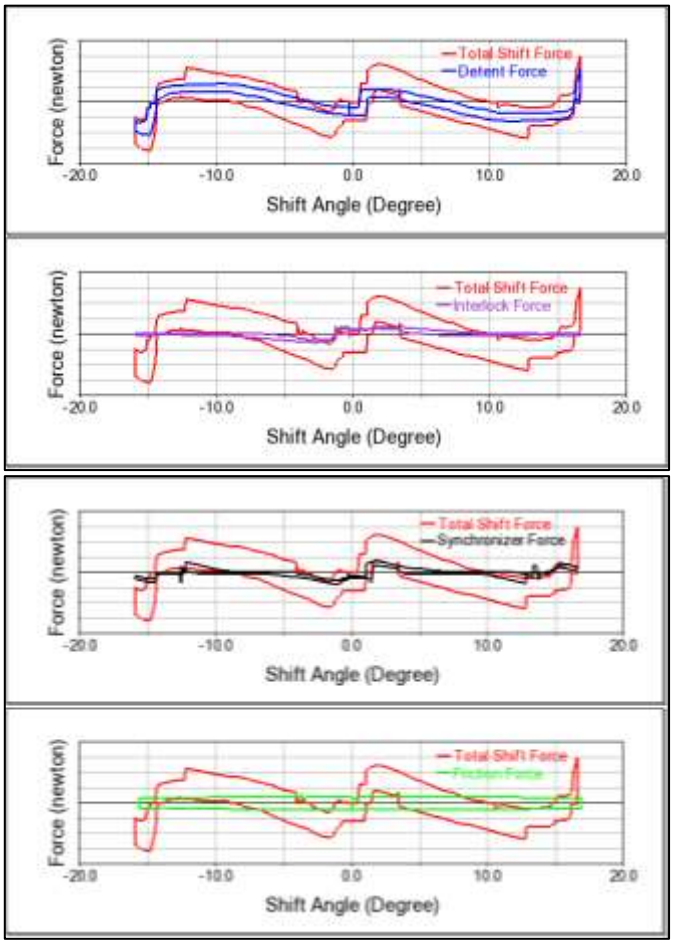


Fig. 13. Contribution of different subsystem towards the total shift force. Graph shows the contribution of different subsystems in individual plots for complete cycle of engagement and disengagement for 3-4 gear pair.

IV. Correlation with Test Result

To validate the virtual dynamic model, the results are compared with experimental test data. As seen in "Fig. 14", the simulation agrees well with the experimental result for complete engagement and disengagement cycle for the 3rd gear shift. All the regions such as maximum load during engagement and disengagement, auto pull load during engagement and disengagement are in agreement with test result. Correlation during engagement and disengagement can be summarized below.

- Overall, there is a good correlation achieved between the test and simulation (> 90 %) for entire engagement and disengagement cycle for 3rd gear.
- This correlation level is achieved as all the subsystems contributing for shift force are considered
- Area for improving further correlation between test and simulation are clearance consideration in the model which is present in actual assembly. Considering parts flexibility may improve the correlation which is not captured in the rigid body analysis.

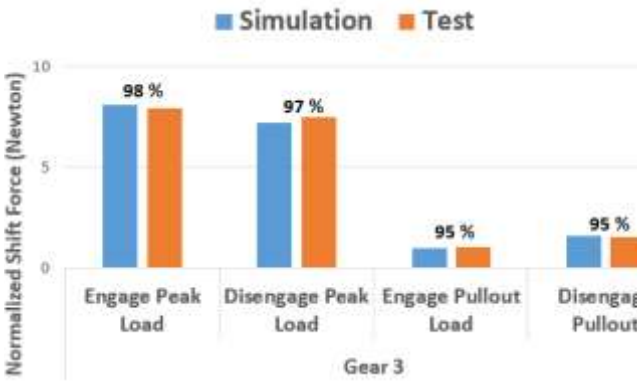


Fig. 14. Correlation of simulation with test result at different regions. Good correlation (>90%) for complete engagement and disengagement cycle is achieved.

V. Conclusion and Future Work

MBD analysis is performed for a complete shift system and the results are in agreement with the experimental data in terms of load value as well as trends for different phenomenon during complete cycle. This CAE virtual model has been used in multiple program and results are validated with test results.

This study can be further used for sensitivity analysis of the parameters such as CAM profile geometry, spring parameters, detent ball diameter and their effect on shift forces for the complete cycle. Parameter study can lead to optimize the design.

Acknowledgment

We are thankful to Eaton India Innovation Centre, Pune and Eaton Transmission, Valinhos, Brazil for providing this opportunity and infrastructure to perform this task. We are thankful to leadership at EIIC Pune and Eaton Transmission Valinhos for providing guidance and continuous support.

References

[1] Kent, David Kelly Christopher. *Gear shift quality improvement in manual transmissions using dynamic modelling*. No. 2000-05-0126. SAE Technical Paper, 2000.

[2] Davis, Geoff, Rolland Donin, Mark Findlay, Peter Harman, Mark Ingram, and David Kelly. "Optimization of gear shift quality: By Means of Simulation." *ATZ worldwide* 106, no. 7-8 (2004): 23-26.

[3] More, Prajwal, Santosh Deshmane, and Onkar Gurav. "Gear Shift Quality Parameters Optimization for Critical to Quality Dimensions." *SAE International Journal of Engines* 11, no. 3 (2018): 265-276.

[4] Singh, Bhupinder, Rajesh Vats, Rohan Garg, and Manoj Tanwar. *Dynamic Simulation of Shift Tower*. No. 2013-01-2790. SAE Technical Paper, 2013.

[5] Goel, Somya, and Vikraman Vellandi. Investigations on the Effect of Synchronizer Strut Detent Groove Profile on Static and Dynamic Gear Shift Quality of a Manual Transmission. No. 2020-28-0319. 2020.

[6] A. A. Shabana, Dynamics of Multibody System, Third ed., Cambridge University Press., 2005.

Accessibility Evaluation of Saudi Health-Related Websites

Redhwan Nour
Department of Computer Science, College of Computer Science and Engineering
Taibah University
Medina 42353, Saudi Arabia
rnour@taibahu.edu.sa

***Abstract*—Health websites are empowering people to find health information and access online services that allow handling of health affairs. However, health websites need to be accessible to support a greater variety of people and provide equal access to individuals regardless of their abilities and disabilities. This study set out to analyze the current accessibility of Saudi health-related websites. Using two automatic evaluation tools, 20 websites were evaluated for compliance with WCAG 2.1 guidelines. None of the tested websites passed the evaluation, and the majority of errors were related to basic web accessibility requirements. These findings highlight that Saudi health-related websites are currently not very satisfactory in their accessibility, and the reported violations could cause barriers that hinder people with disabilities in accessing and benefiting from online health information and services.**

Keywords—web accessibility; accessibility evaluation; health websites; WCAG; disability

I. INTRODUCTION

Reliance on the web for finding health information and accessing health data and services is increasing, as is demand to satisfy the needs of people regardless of their abilities. Some search the web for health information to look for a diagnosis or information on a chronic disease; others access health websites for services like making an appointment, buying medicines, or renewing a health insurance plan. Additionally, some are looking for health-related information not just out of sheer curiosity, but because they do not get enough information from their doctor, or do not trust the information given by their doctor [1]. Although the web can provide valuable information, using it may pose some difficulties for people with disabilities if the accessed websites do not sufficiently consider their needs [2].

The web is an E-Health technology that has opened new opportunities for patients to benefit from health services and for providers to enhance and improve upon a wide range of health and care services [3], [4]. Regardless of their capabilities, people with disabilities need to not be neglected and have the right to fully utilize health-related websites. These websites should be accessible such that disabled individuals can discern, comprehend, navigate, engage with, and add their own contributions to the sites [5].

As one element of the Saudi Vision 2030, the Ministry of Health (MOH) has initiated an E-Health strategy to expand and improve health services for all related sectors [6]. One of the strategy goals is to provide inclusive services that support a range of disabled individuals. The disabled represent 7.1% of the Kingdom's population of 32 million [7]. Within that subpopulation, 52.2% are males and 47.8% are females, and six types of disabilities are represented: physical, visual, hearing, Down syndrome, autism, and attention deficit hyperactivity disorder (ADHD).

In Saudi Arabia, use of the web is increasingly widespread; according to the Saudi Communications and Information Technology Commission (CITC), over 98% of the population (98.5% male, 97.7% female) was using the internet by the end of 2021[8]. In addition, they reported that over 54% spent upwards of seven hours online daily, and searching for health information accounted for 34.5% of the total online activity [8]. Therefore, it is essential to construct accessible websites so as to enable those with disabilities to have the same opportunity to make use of the services and information provided through health websites.

Web accessibility concerns designing a website such that it can be used by or otherwise benefits a disabled person. The World Wide Web Consortium (W3C) has released a set of Web Content Accessibility Guidelines (WCAG) that constitute an international standard aimed at increasing the accessibility of webpages for those who have disabilities. Two versions of these guidelines are in use at present: WCAG 2.0 dating from 2008 and its backwards-compatible successor WCAG 2.1 released in 2018 [9]. Multiple layers of guidance make up the WCAG 2.1; the highest level comprises four principles, followed by 13 broad guidelines under which are a total of 78 testable Success Criteria (SC). Sixty-one of those criteria were inherited from WCAG 2.0, and 17 were added to specifically address access through mobile devices, users having poor vision, and users having cognitive or learning disabilities [10].

At its foundation, web accessibility is comprised of four principles, encapsulated by the acronym "POUR". The first principle, *Perceivable*, means a user must be able to perceive the site interface and information regardless of sense capabilities and access method. The second, *Operable*, refers to supporting interaction through multiple modalities. The third,

Understandable, mandates that interface operations and the information presented must be comprehensible. Finally, *Robust* requires that site content be accessible to diverse technologies [10].

When testing SC in WCAG 2.1, test outcomes are rated in terms of three levels. A score of "A" represents minimal conformance, adhering to only the most basic of accessibility requirements. The score "AA" represents intermediate compliance, in which significant barriers have been accounted for. The highest score, "AAA," represents maximal conformance and is the most desirable score. Notably, these levels are cumulative; that is, a site that achieves AA conformance also meets the criteria for Level A [11].

The accessibility of a website can be evaluated in several ways, including through an initial check, by people, and with automated tools. Notably, tools are used to overcome some challenges faced by other methods [12], [13]. Definitionally, such tools are "*Software programs or online services that help determine if web content meets accessibility standards*" [14] and can provide automated checks of accessibility violations in accordance with the WCAG 2.1 standard and generate reports in a number of formats such as PDF, CSV, HTML, and XML. Additionally, different tools have different features and focuses, for example guidelines, language, type of tool, supported formats, etc. The W3C has created a list [15] of web accessibility evaluation tools, which has more than 100 different tools to choose from.

This study set out to evaluate 20 Saudi health-related websites in terms of their degree of accessibility. The websites were categorized into four groups: government websites, public hospitals, private hospitals, and health insurance companies. The evaluation was conducted by using automated tools to report accessibility violations for each website. The aim of this endeavor is to highlight the current status of web accessibility regarding health websites and hence to communicate areas of improvement to web developers, health practitioners, and service providers in addition to those decision makers who focus on E-Health initiatives and programs in the kingdom.

The rest of the paper is organized as follows: Section II provides an overview related literature. Section III focuses on the evaluation approach used here. Section IV presents the outcomes of the evaluation by each tool. Section IV discusses the obtained results. Finally, Section VI highlights the conclusion of the study.

II. Related Literature

There is a scarcity of research focusing on the accessibility of health-related websites in Saudi Arabia. Alhadreti [16] used one evaluation tool to audit accessibility of the respective top ten Saudi hospital websites from the public and private sectors. The outcomes indicated that these websites have critical accessibility issues: just 20% of the sites examined were fully compliant with WCAG 2.0. Also, no significant difference was found between public and private hospitals. On average, homepages had 569.7 violations, and major errors were found in the areas of keyboard access, information structure, headings, labels and instructions, and non-text content.

Alajarmeh [17] used different evaluation tools to evaluate public health websites representing over two dozen countries and four continents. The findings highlighted that the vast majority are afflicted by many accessibility problems; the highest average of combined errors was obtained for websites in Europe, followed by websites located in South America, Asia, and North America respectively. The most frequent issues related to the perception and operable principles and to errors corresponding to basic accessibility requirements (WCAG 2.0, Level A). The barriers found to be most common were an absence of alternative descriptions for interface items (links and images) and incompatibility with assistive technologies. One of the websites evaluated in the study was that of the Saudi MOH, which ranked in 9th place with 68 total errors after combining results from all tools. Most errors for the Saudi MOH website related to the operable principle, while the robust principle featured the least issues.

Kuzma [18] used a single online tool to investigate problems with website accessibility for 160 hospitals representing 16 countries and spanning four continents. The results indicated no compliance with basic accessibility requirements; collectively, the tested websites returned a total of 15,663 violations (10,832 for Level A and 4,830 for Level AAA). Moreover, only two of the 160 websites were fully Level A compliant. The most frequent errors concerned non-text content, parsing, and info and relationships. In this study, the Africa/Middle East category, which included Saudi Arabia, returned the fewest errors among continents; however, at 30% of the total, Saudi Arabia had more accessibility violations than any other country within that category.

Another study used multiple tools [19] to evaluate over 50 official US websites for COVID-19 vaccine registration (50 state websites plus four territory-level vaccine websites). The official websites were found to be encumbered with accessibility violations in regard to both WCAG 2.0 and 2.1. The most common accessibility issues concerned text alternatives, content distinguishability, contrast, input assistance, audio components, text resizing, and navigability. In addition, several other studies have evaluated health websites from different countries with various evaluation tools [20]–[23]. All of these studies concluded that health-related websites have very critical accessibility violations and that they are not meeting the needs of disabled individuals.

III. Method

A. Selected Websites

In the interest of analyzing the current accessibility status of health-related websites, four different categories of websites were chosen and five websites selected within each category. The first category consisted of websites selected from the Saudi Health Sector Transformation Program (HSTP) [24]. These belong to leading parties in various initiatives related to the Saudi Vision 2030 and are considered government institutes. The second and the third category focused on top public and private hospitals respectively. Hospitals for these categories were selected via the "Ranking Web of World Hospitals" published by Cybermetrics Lab, an online ranking of healthcare institutions worldwide [25]. The final category targeted health insurance companies and referenced the Atlas

TABLE I. SELECTED HEALTH-RELATED WEBSITES

#	Name
	Health Government Institutes
1	Ministry of Health (MOH)
2	Saudi Food and Drug Authority (SFDA)
3	Saudi Red Crescent Authority (SRCA)
4	Saudi Health Council (SHC)
5	King Faisal Specialist Hospital and Research Centre (KFSHRC)
	Public Hospitals
6	National Guard Health Affairs (NGHA)
7	Royal Commission Hospital in Jubail (RCHJ)
8	Security Forces Hospital (SFH)
9	King Fahad Medical City (KFMC)
10	King Khaled Eye Specialist Hospital (KKESH)
	Private Hospitals
11	Magrabi Hospitals and Centers (MHC)
12	Al Moosa General Hospital (AGH)
13	Saudi German Hospitals Group Jeddah (SGHJ)
14	Adama Hospital (AH)
15	Dallah Hospital (DH)
	Health Insurance Companies
16	Bupa
17	Tawuniya
18	Alrajhi Takaful
19	Walaa
20	Medgulf

Magazine website, which ranked Saudi insurance companies' performance in 2021 [26]. Table I lists all 20 selected websites.

B. Evaluation Tools

To enhance the reliability of the results, overcome the limitations of using a single tool, and benefit from more features and functionality [27], this study employed two online automatic evaluation tools, WAVE [28] and Siteimprove [29]. Both have been used in previous literature [17], [19], [30], [31] and are listed on the W3C website as appropriate for checking against WCAG 2.1 [15]. WAVE is an online service that provides three color-coded icons to identify issues. The red icon indicates accessibility and contrast errors, the yellow icon points to alerts that need human intervention, and the green icon relates to features and areas of improvement. Siteimprove is an online browser extension that highlights the locations of errors and reports accessibility violations at all levels with suggestions for fixes and improvements.

C. Evaluation Procedure

The 20 landing pages evaluated in this study were analyzed using two automatic analysis tools, WAVE and Siteimprove. Both evaluated all three conformance levels (i.e., checks corresponded to WCAG 2.1 Levels A, AA, and AAA). Evaluations were carried out in Google Chrome on a computer running 64-bit Microsoft Windows 10 in June 2022 and analyzed the Arabic version of each homepage.

IV. RESULTS

A. WAVE

The evaluation results are listed in Table II; in total, there were eleven different accessibility errors and none of the 20 websites passed the evaluation with no errors at all. Three errors ("Very low contrast," "Empty link," and "Missing alternative text") accounted for around 82% of the 1,708 total violations and affected the majority of websites. The most frequent error was "Very low contrast," which constituted 42% of all violations and was the only error affecting all 20 websites. This indicates all websites as having low contrast between text and background colors and needing to fix the contrast ratio. The second most prevalent error was "Empty link," which comprised 29% of all errors. This error affected 18 websites that had links with no text. In third place, "Missing alternative text," constituted around 11% of all violations; this error also ranked third in sites affected, with 17 in total. This error highlights the absence of descriptive text for non-text items. Overall, the first six of the eleven errors affected more than 50% of the evaluated websites. Referring to the SC, it can be seen that most websites violated the basic accessibility requirements; only two of the unmet SC, "2.4.6" and "1.4.4," related to Level AA; all of the others were associated with Level A. The SC "1.1.1 - non-text content" was very likely to be unsatisfied, appearing in more than 50% of errors.

TABLE II. RESULTS OF THE ERROR ANALYSIS USING WAVE

Error Type	Number of Websites	Success Criteria	Number of Occurrences	Error %
Very low contrast	20	1.4.3	717	42
Empty link	18	2.4.4	499	29.21
Missing alternative text	17	1.1.1	191	11.18
Linked image missing alternative text	13	1.1.1, 2.4.4	130	7.61
Missing form label	17	1.1.1, 1.3.1, 2.4.6, 3.3.2	72	4.22
Empty button	12	1.1.1, 2.4.4	42	2.46
Broken ARIA menu	4	2.1.1, 4.1.2	23	1.34
Empty heading	5	1.3.1, 2.4.1, 2.4.6	16	0.94
Spacer image missing alternative text	3	1.1.1	10	0.58
Language missing or invalid	5	3.1.1	5	0.29
Empty form label	3	1.1.1, 1.3.1, 2.4.6	3	0.17
		Total	**1708**	**100**

TABLE III. RESULTS OF THE ERROR ANALYSIS USING SITEIMPROVE

Error Type	Number of Websites	Success Criteria	Occurrences	Error %
Link without a text alternative	20	2.4.4, 2.4.9	530	27.33
Font size is fixed	6	1.4.8	254	13.1
Element IDs are not unique	12	4.1.1	165	8.51
Links are not clearly identifiable	6	1.4.1	144	7.43
Text is clipped when resized	7	1.4.4	137	7.07
Color contrast is not sufficient	16	1.4.3, 1.4.6	123	6.34
Line height is fixed	11	1.4.8	114	5.88
Image without a text alternative	16	1.1.1	110	5.67
Line height is below minimum value	12	1.4.8	101	5.21
Visible label and accessible name do not match	6	2.5.3	83	4.28
Hidden element has focusable content	3	1.3.1, 4.1.2	43	2.22
Role not inside the required context	3	1.3.1	34	1.75
Button without a text alternative	8	4.1.2	34	1.75
Container element is empty	7	1.3.1	15	0.77
Uneven spacing in text	3	1.4.8	12	0.62
Form field is not labelled	7	4.1.2	10	0.52
Page language has not been identified	4	3.1.1	9	0.46
Page zoom is restricted	7	1.4.4	7	0.36
Empty headings	4	1.3.1, 2.4.6	5	0.26
Inline frame without a text alternative	5	4.1.2	5	0.26
Invalid role	1	1.3.1	2	0.1
Content language not recognized	1	3.1.2	1	0.05
Scrollable element is not keyboard accessible	1	2.1.1	1	0.05
		Total	**1939**	**100**

TABLE IV. SUMMARY OF ERRORS BY LEVEL AND PRINCIPLE

Error Type	Total Error	Error Pct	Min	Max	Average	Median
A	2181	50%	1	530	87.24	34
AA	1081	24%	1	717	120	44
AAA	1135	26%	1	530	162	118.5
P	2282	50%	2	717	103.72	72
O	1404	30%	1	530	127.63	57
U	90	2%	1	72	18	7
R	810	18%	5	530	115.71	38.5

B. Siteimprove

Similar to WAVE, no website passed the Siteimprove evaluation with no errors at all (Table III). However, Siteimprove reported 23 different errors, four of which ("Link without a text alternative," "Font size is fixed," "Element IDs are not unique," and "Links are not clearly identifiable") constituted around 56% of the 1,939 total violations. The most frequent error was "Link without a text alternative," constituting nearly 27% of all violations. Considering that all 20 websites were affected by this error, more work is needed to identify the purpose of links without relying on additional context. The second most prevalent error was "Font size is fixed," which made up 13% of all violations despite only affecting six websites. This suggests a need to allow browsers to adjust the default font size so as to make reading easier. The third most common error was "Element IDs are not unique," which encompassed around 8.5% of all violations and affected over 50% of the evaluated websites; this highlights the importance of parsing content. Similar to the font size error, the fourth most common error "Links are not clearly identifiable" affected only six websites while constituting around 7% of all violations. Overall, as with WAVE, this evaluation also indicates that more than 50% of the websites were affected by only six different errors. Regarding SC, the majority of evaluated websites violated basic accessibility requirements; only seven unmet SC pertained to Levels AA and AAA, while the remaining 17 SC related to Level A. The SC "1.3.1 - Info and Relationships" was the most common unmet criterion, indicating the importance of providing content in multiple formats to aid user perception.

V. DISCUSSION

This study assessed the state of web accessibility in Saudi health-related websites by means of automatic evaluation tools. The findings indicate there is a significant problem in adherence to accessibility standards. The accessibility situation of Saudi health-related websites is not currently sufficient as in total, 3,647 violations were identified by the two tools and not one of the evaluated websites could pass the accessibility test. These findings are similar to some reported in [16], [17].

Among these websites, not one passed the lowest bar of accessibility conformance, i.e. an absence of Level A errors. In fact, as presented in Table IV and Fig. 1, Level A had the most errors at 2,181, with an average of 87.24; this level comprised 50% of all violations. Level AA had 1,081 errors with 120 on average, and encompassed 24% of all violations. Moreover, this level included the single most frequent error at 717

TABLE V. SUMMARY OF ERRORS PER GROUP

Group	Total	Avg	Median	Max	Min	Site with Least Errors
Public Hospitals	1182	236.4	210	514	105	NGHA
Government	1168	233.6	150	533	91	KFSHRC
Private Hospitals	659	131.8	100	234	22	Adama
Health Insurance	638	127.6	107	240	33	Alrajhi Takaful

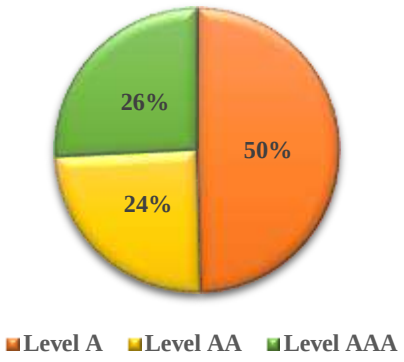


Fig. 1. Error distribution by conformance level.

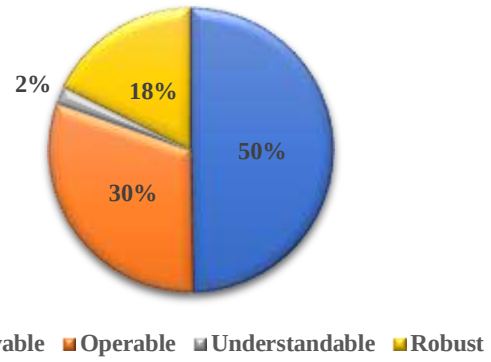


Fig. 2. Error distribution by principle.

violations, while the most frequent errors in other levels tied at 530 violations. Overall, 74% of violations in the evaluated websites represent failures to meet minimal (A) and acceptable (AA) levels of compliance.

These are not promising findings and indicate the neglect of various disabled users who could benefit from these health websites. With regard to web accessibility principles, it is evident in Table IV and Fig. 2 that the majority of errors related to the perception category, constituting 50% of all violations with 2,282 in total and 103.72 on average. This suggests that the developers of these health websites must consider presenting web content in different ways so that users can use one or more of their senses to perceive it. The second most affected category was operable, representing 30% of all violations with 1,404 in total and an average of 127.63. This indicates difficulties in navigating the websites and using their components and features. Robust was the third most impacted category, encompassing 18% of all violations, while understandable was the least with only 2%.

As presented in Table V, the group having the most combined errors on average was public hospitals, then health government institutes, private hospitals, and health insurance companies in that order. The average total violations is 182.35, with the least being 22 for the Adama website and the most 105 for NGHA. It is clear that in terms of avoiding accessibility violations, the private sector (private hospitals and health insurance companies) performs better than the government sector; combined, government and public hospital sites accounted for 64% of all violations. This finding is consistent with [32] as the private sector is more active in providing E-Health services in general. Moreover, the overall results underscore the absence of standard awareness. This support [33] findings where education programs for health informatics in Saudi Arabia lack accreditation and standardization, which is reflected in noncompliance with guidelines such as WCAG.

VI. Conclusion

The study highlights the current status of accessibility compliance for 20 Saudi health-related websites. The results showed that due to issues with accessibility, many Saudi health websites are not in compliance with current guidelines; in fact, none of the evaluated websites fully conformed with the lowest conformance level of WCAG 2.1. Moreover, these websites in their present form do not have satisfactory accessibility status, as the evaluation tools reported numerous violations. In total, 3,647 errors were identified, with an average of 182.35; moreover, 74% of those errors related to Levels A and AA, which cover the accessibility requirements that many organizations strive to meet. In particular, more work is evidently needed from public and government institutes compared to private ones. The reported accessibility barriers may hinder disabled individuals who attempt to access online health services and information.

References

[1] K. M. AlGhamdi and N. A. Moussa, "Internet use by the public to search for health-related information," *Int J Med Inform*, vol. 81, no. 6, pp. 363–373, Jun. 2012, doi: 10.1016/j.ijmedinf.2011.12.004.

[2] G. Berget and A. MacFarlane, "What is known about the impact of impairments on information seeking and searching?," *Journal of the Association for Information Science and Technology*, vol. 71, no. 5, pp. 596–611, 2020, doi: 10.1002/asi.24256.

[3] F. Barbabella, M. G. Melchiorre, S. Quattrini, R. Papa, and G. Lamura, *How can eHealth improve care for people with multimorbidity in Europe?* Copenhagen (Denmark): European Observatory on Health Systems and Policies, 2017. Accessed: May 05, 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK464571/

[4] G. Eysenbach, "What is e-health?," *J Med Internet Res*, vol. 3, no. 2, p. e20, Jun. 2001, doi: 10.2196/jmir.3.2.e20.

[5] World Wide Web consortium (W3C), "Introduction to Web Accessibility," *Web Accessibility Initiative (WAI)*. https://www.w3.org/WAI/fundamentals/accessibility-intro/ (accessed May 05, 2022).

[6] Ministry of Health, "National E-health Strategy MOH Initiatives 2030." https://www.moh.gov.sa/en/Ministry/nehs/Pages/vision2030.aspx (accessed May 10, 2022).

[7] General Authority for Statistics, "Disability Survey 2017," Saudi Arabia, Dec. 2017. [Online]. Available: https://www.stats.gov.sa/en/904

[8] "Saudi Internet 2021," Communications & Information Technology Commission, Saudi Arabia, 2021. Accessed: May 14, 2022. [Online]. Available: https://www.citc.gov.sa/ar/indicators/PublishingImages/Pages/saudi_internet/internt-saudi-2021.pdf

[9] World Wide Web consortium (W3C), "WCAG 2 Overview," *Web Accessibility Initiative (WAI)*. https://www.w3.org/WAI/standards-guidelines/wcag/ (accessed May 14, 2022).

[10] World Wide Web consortium (W3C), "Web Content Accessibility Guidelines (WCAG) 2.1." https://www.w3.org/TR/WCAG21/ (accessed May 14, 2022).

[11] "WebAIM: Web Content Accessibility Guidelines," Sep. 21, 2020. https://webaim.org/standards/wcag/ (accessed May 14, 2022).

[12] J. Grantham, E. Grantham, and D. Powers, “Website accessibility: an Australian view,” in *Proceedings of the Thirteenth Australasian User Interface Conference - Volume 126*, AUS, Jan. 2012, pp. 21–28.
[13] H. Petrie, F. Hamilton, N. King, and P. Pavan, “Remote usability evaluations with disabled people,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2006, pp. 1133–1141. doi: 10.1145/1124772.1124942.
[14] World Wide Web consortium (W3C), “Evaluating Web Accessibility Overview,” *Web Accessibility Initiative (WAI)*. https://www.w3.org/WAI/test-evaluate/ (accessed May 14, 2022).
[15] World Wide Web consortium (W3C), “Web Accessibility Evaluation Tools List.” https://www.w3.org/WAI/ER/tools/ (accessed May 14, 2022).
[16] O. Alhadreti, “An accessibility evaluation of the websites of top-ranked hospitals in Saudi Arabia,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 1, Art. no. 1, Aug. 2021, doi: 10.14569/IJACSA.2021.0120180.
[17] N. Alajarmeh, “Evaluating the accessibility of public health websites: An exploratory cross-country study,” *Univers Access Inf Soc*, pp. 1–19, Jan. 2021, doi: 10.1007/s10209-020-00788-7.
[18] J. Kuzma, J. Law, V. Bell, and N. Williams, “A study of global hospital websites for accessibility compliance,” *European Journal of Business and Social Sciences*, vol. 6, no. 6, pp. 15–28, 2017.
[19] S. Alismail and W. Chipidza, “Accessibility evaluation of COVID-19 vaccine registration websites across the United States,” *Journal of the American Medical Informatics Association*, vol. 28, no. 9, pp. 1990–1995, Sep. 2021, doi: 10.1093/jamia/ocab105.
[20] J. L. Brobst, “United States federal health care websites: A multimethod evaluation of website accessibility for individuals with disabilities.,” The Florida State University, 2012. Accessed: May 05, 2022. [Online]. Available: https://www.proquest.com/docview/1034593107
[21] Y. J. Yi, “Web accessibility of healthcare web sites of Korean government and public agencies: a user test for persons with visual impairment,” *Univ Access Inf Soc*, vol. 19, no. 1, pp. 41–56, Mar. 2020, doi: 10.1007/s10209-018-0625-5.
[22] J. Martins, R. Gonçalves, F. Branco, J. Pereira, C. Peixoto, and T. Rocha, “How ill Is online health care? An overview on the Iberia Peninsula health care institutions websites accessibility levels,” in *New Advances in Information Systems and Technologies*, Cham, 2016, pp. 391–400. doi: 10.1007/978-3-319-31307-8_41.
[23] A. Kaur, D. Dani, and G. Agrawal, “Evaluating the accessibility, usability and security of hospitals websites: An exploratory study,” in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Jan. 2017, pp. 674–680. doi: 10.1109/CONFLUENCE.2017.7943237.
[24] “Health Sector Transformation Delivery Plan,” Health Sector Transformation Program 2021, Saudi Arabia. Accessed: May 05, 2022. [Online]. Available: https://www.vision2030.gov.sa/media/0wop2tds/hstp_eng.pdf
[25] “Saudi Arabia,” *Ranking Web of Hospitals*. https://hospitals.webometrics.info/en/aw/saudi%20arabia%20 (accessed May 05, 2022).
[26] “Top 10 insurance companies in Saudi Arabia,” *Atlas Magazine*, Apr. 11, 2022. https://www.atlas-mag.net/en/article/insurance-companies-in-saudi-arabia-ranking-2018 (accessed May 05, 2022).
[27] M. Vigo, J. Brown, and V. Conway, “Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests,” in *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, New York, NY, USA, May 2013, pp. 1–10. doi: 10.1145/2461121.2461124.
[28] “WAVE Web Accessibility Evaluation Tool.” https://wave.webaim.org/ (accessed May 05, 2022).
[29] “Accessibility hub: Resources for an inclusive web experience,” *Siteimprove*. http://prod.siteimprove.com/accessibility-hub/ (accessed May 05, 2022).
[30] S. Panda and R. Chakravarty, “Evaluating the web accessibility of IIT libraries: a study of Web Content Accessibility Guidelines,” *Performance Measurement and Metrics*, vol. 21, no. 3, pp. 121–145, Jan. 2020, doi: 10.1108/PMM-02-2020-0011.
[31] A. Alsaeedi, “Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: proposed frameworks,” *Information*, vol. 11, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/info11010040.
[32] A. Noor, “The Utilization of E-Health in the Kingdom of Saudi Arabia,” *International Research Journal of Engineering and Technology*, vol. 6, no. 9, pp. 1229–1239, Sep. 2019.
[33] A. Noor, “Discovering Gaps in Saudi Education for Digital Health Transformation,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 10, Art. no. 10, 31 2019, doi: 10.14569/IJACSA.2019.0101015.

Tele-Health Security Framework for Medical Image Storage

Mohammed Ayad Saad
Dept of Electrical, Electronics & Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Malaysia
Department of Medical Instrumentations Technique Engineering, Alkitab University, Kirkuk, Iraq
Mohmmad9alani@gmail.com
0000-0001-9527-264

Ahmed Hashim Rashid
Dept of Electrical, Electronics & Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Malaysia
P109034@siswa.ukm.edu.my
0000-0002-9944-1505

Rosmina Jaafar
Dept of Electrical, Electronics & Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Malaysia
rosmina@ukm.edu.my
0000-0001-8019-0446

Kalaivani Chellappan
Dept of Electrical, Electronics & Systems Engineering, Universiti Kebangsaan Malaysia, Bangi, Malaysia
kckalai@ukm.edu.my
0000-0002-2618-216X

Abstract: Since the emergence of COVID-19, awareness on telehealth and related discussions have escalated because of the increasing demand for online applications and services to minimize in-person meetings and treatments. After decades of using telehealth services in only a few situations, nowadays such services have gained prevalence. Besides the increasing need for these services, the highly developed telecommunication networks and the use of the Internet -of things have enhanced the use of such services. One of the technologies used in telehealth services is medical imaging, where images store important information for later usage. However, medical image storage has several problems associated with it, including latency, image size, and security. The first two issues can be overcome using a cloud storage environment, but security challenges are on a rise with the introduction of cloud storage and telehealth services. In this review paper, we investigate the requirement of medical image management that should be taken into consideration in designing a security framework for telehealth services. A framework is devised for the storage security management techniques using two different parameters: device security and access security. Furthermore, we discuss some related works that propose advanced techniques to ensure medical image security.

Keywords— ***Medical Image, Storage Security, Telehealth, Device Security, Access Security***

I. INTRODUCTION

Development and civilization have changed lifestyle and human needs. These include the need for treatment, continuous health monitoring, accessing medical files, and remote data access, which can reduce treatment costs. Telehealth concept has witnessed improvements, especially with the aid of advanced telecommunication systems and information technologies. Telehealth requires the application of advanced telecommunication systems such as the fifth generation (5G) technology, which provides terahertz of data transfer to allow access to healthcare services remotely and manage healthcare. By using telehealth services through mobile applications, medical websites, and data links to the medical files and images, it offers medical services to rural and isolated communities, making basic health services more readily available anytime and anywhere to save time and improve the communication and coordination between healthcare personnel and patients [1], [2].

A recent notable example of the application of telehealth services is in reducing the healthcare professionals' exposure to infected individuals during the COVID-19 pandemic. Implementation of telehealth services during a pandemic can help reduce patient surge in healthcare facilities, minimize the loss of healthcare staff lives due to infection, and reduce operation cost by avoiding extensive use of personal protective equipment. Telehealth services allow evaluating, giving treatment prescriptions, and monitoring the patient without in-person services, and they are considered a very powerful way to offer medical services. The concept of telehealth is not new; it was introduced in the 1960s, when wireless communication systems were used to monitor the physiological parameters of astronauts. In the years following that, the development of telehealth applications took more consideration from researchers, governments, and industries for technological advancements in telecommunication networks implemented to enable fast transfer of medical information, data, consultations, and recommendations remotely from one place to another .Telehealth applications require typical data delivery techniques including mobile applications connected to a network linking patients with hospitals or medical centers, high-connection video applications, point-to-point connection to hospitals and clinics such as using microwave links and/or fiber optics communication, web-based e-health service pages, and home monitoring links [3].

Notwithstanding their advantages, telehealth services require massive investments, and the distribution of the

telehealth hospital-based networks is considered very costly. ,Broadband services and networks facilitate the construction of telehealth infrastructure and allow increased use of telehealth services such as capturing medical images, transferring images, and storing images in a secured way. Medical image security is therefore very important, especially for big data requirements [4].

This paper overviews the extant literature on medical image security. Based on the information gathered from the literature review, a framework for a secured medical imaging system for telehealth is proposed. The paper concludes with a discussion on several open research challenges in addition to a discussion on some possible solutions for adequately applying security techniques.

II. Literature Review

In any network, data security is very important to be considered to overcome all challenges that come from the third party. Medical images and biomedical data are also very important to be stored securely because of their sensitivity to small changes. Two types of technologies can be used to secure medical images; the first one depends on hiding some of the medical data to secure transmission of it, and the second one involves the use of encryption techniques of the whole image to completely hide the image's content so that no one can clearly see the hidden content. Cellular neural network crumb coding transform (CNN-CCT) introduces a technique to encrypt the medical images using the concept of hyper-chaotic cellular neural network (CNN) technology. This technique ensures greater efficiency in the hardware storage capacity with a high-security process because of the use of cipher block chain technology in security process [5]–[7].

Simulation results obtained after using this technique show that the medical storage system is highly robust against various types of external attacks with competitive advances in storage capacity and efficiency. Researchers have analyzed the sensitivity of medical data using machine learning algorithms to facilitate the use of encrypted images in the cloud storage. They have used the concept of data-at-rest encryption to decrease the security load on the cloud and to ensure security. Next, they have presented an improved watermarking technique capable of protecting patient data by embedding multiple watermarks in the medical cover image using discrete wavelet transform with singular value decomposition domain. The process of watermarking depends on encryption and compression. The researchers have used a combination of hyper-chaotic algorithm and Lempel–Ziv–Welch (LZW) to ensure robustness against Gaussian noise, speckle noise, and histogram equalization attacks [8], [9].

As mentioned in the introduction, encryption and authentication techniques have been used in medical image security. These two techniques employ cryptographic methods to prevent the stored medical data from being attacked by third parties. Their security algorithm is based on the idea that the big medical data gathers, and their redundancy requirement cannot be met from the traditional security techniques. Because of this, researchers have proposed an enhanced version of spatial domain methods such as the discrete cosine transform (DCT) method for encryption, which does not affect the histogram of the medical image, in addition to the use of Fresnel transform and Arnold transform. Researchers have also proposed securing electronic health records (EHR) containing X-rays, computerized tomography scans, and magnetic resonance imaging (MRI) reports using a decoy strategy. Internal attacks and external penetration pose a threat to EHRs. A previous study was focused on reducing the time consumption and cost of image storage and use. The researchers used the concept of cloud computing, which decreased the cost of storing and processing and displayed improved competence and quality. They also proposed a hex code-oriented encryption method to encrypt the stored images [5], [10].

In two other studies [16, 17], the authors proposed a method that can trigger two levels of encryption for achieving a higher grade of security. The first level encrypts the data by assigning hex codes to each character to generate a unique key for each data, and this key is converted into hex code and is added with the encrypted file, which, in turn, provides strong encryption and high efficiency. The authors found it to be a very powerful technique to use, but noted that it suffers from complexity when there are big data. Thus, it should be used under the condition that each data entry should have a unique generated and converted key.

It is important to discuss basic medical image storage requirements and critically address the most common storage problems that the new technologies promise to overcome. The most common problem is the image size, where the concept of higher-quality images taking up less space in recent advanced storage techniques is not true for medical images. The big file size of medical imaging causes not only the requirement of a huge amount of storage spaces but also a latency when one wants to call the image to use it because image of large sizes tax one's digital bandwidth, making it more difficult to transfer images into storage and pull them back. The solution here is to avoid wasting the storage space by using cloud storage in the medical field because it is approved to immediately purchase additional storage space at a reasonable price, and the organizations meet their providers' latency requirement [11], [12]. The latency of communication is the second critical problem faced in medical image storage. Latency of more than 3 seconds in telehealth systems cannot be accepted as it does not allow the synchronization of the operations. The third problem is the security of the storage system, which can offer to satisfy privacy and secure connection of patient information and data. This is an important aspect to be considered in image storage system management [11] .

The three problems discussed above lead to formation of medical image storage requirements that should be satisfied to ensure user security and privacy in telehealth services. The requirements are defined while designing a framework for telehealth services that has the objective of maximizing the perception of security and privacy of the system described as follows:

- The security process should not limit the purpose of telehealth services, which are the accessibility and affordability to users, regardless of their place or their environments

• Ensure the accessibility of the system of with user data and information in a low-latency period anytime and anywhere.
• Several parameters should be ensured in a telehealth system, which are confidentiality, authentication, non-repudiation, and integrity [13].
• Personalization is the core of a telehealth system to enhance the security level.

To ensure security, patients and healthcare providers must trust telehealth systems to keep personal information private and safe for telemedicine to be more trustworthy. To ensure this, two considerations in the security process should be satisfied, which are individual device security and access to the system security. Fig. 1 shows the framework of medical image storage security techniques. Not all devices in a telehealth system are secured. One can be sure about the organization device to be secure, but what about the user's devices? It is impossible for providers to assure that 'users' home networks are secure. As a result, on all provider-owned telehealth equipment, technological programs such as firewalls and intrusion detection systems must be installed to ensure device security. Another security concern is data access security, which implies that the data in the storage should be secured using various ways such as secure logins for both the patient and the provider to the data and multifactor authentication [14].

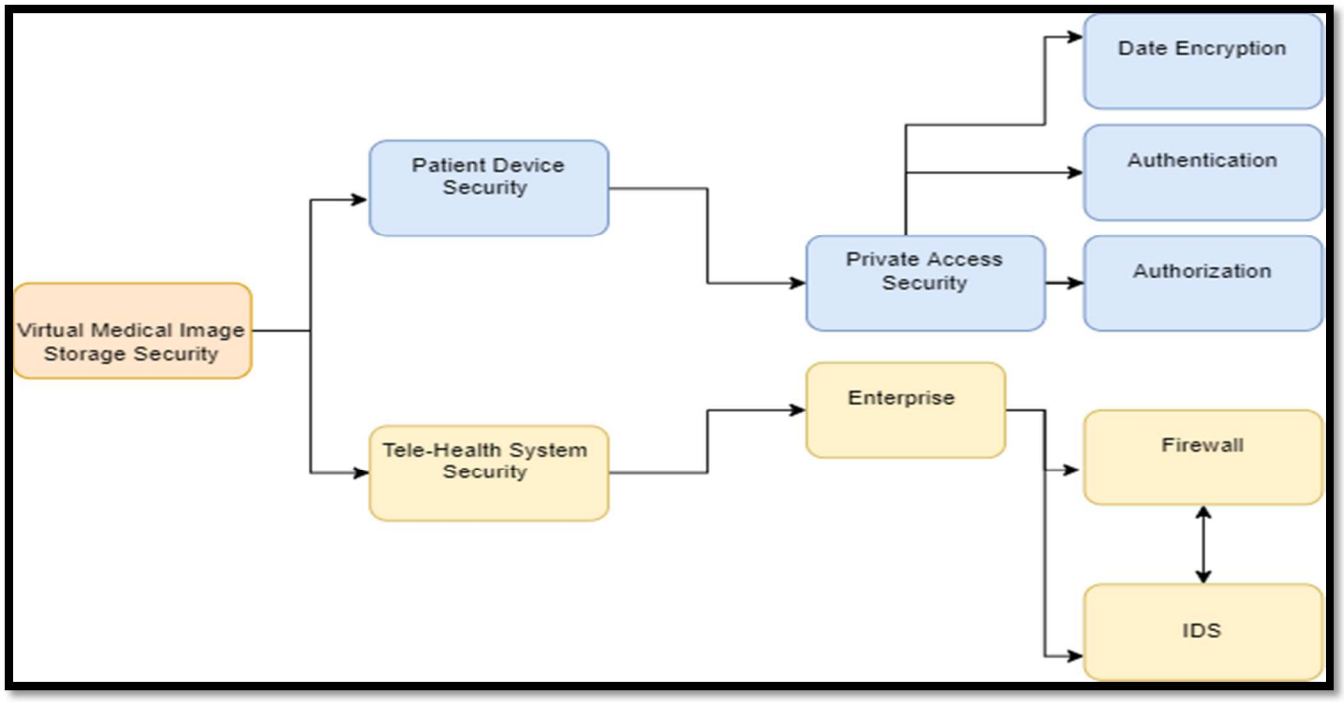


Fig. 1. Medical Image Storage Security Framework

Data, especially medical images, should be kept safe when stored. These security techniques make data unsafe in case of spamming or hacking. Some of these techniques and the most proper ones that are used in access security are described as follows.

A. Data Encryption:

Medical images can be kept safe by employing suitable encryption keys. Data will be meaningless if a system does not have reliable security in ensuring the validity of the service provided. Encryption ensures that data is meaningless in the transmission media, encryption of data in transit, on the other hand, ensures that data is rendered meaningless in the transmission medium. End-to-end encryption assures data security on both ends of a communication channel, including the user side and the provider side [15].

B. Authentication:

To gain access to medical images, one can use one of the following authentication methods: (1) knowledge-based authentication, which necessitates the 'user's knowledge of a hidden piece of information or the entry of a password; (2) biometric authentication, which uses something unique to a 'user's attributes such as their fingerprint; and (3) multifactor authentication, which combines the two preceding approaches [16].

C. Multiple Device Access

One of the techniques used to ensure access security is the distribution of image storage. This means that images can be distributed in different types of storage platforms such as cloud storage and device storage [11].

This technique provides the ability to restore data immediately when required with less latency because it can be gathered from the available storage type. Another advantage of this technique is that the image itself is divided into several parts, and thus, it is difficult for attackers to gather them. To our knowledge, we are the first to offer the overall taxonomy,

requirements, and discussion of medical image storage security techniques used in telehealth services. We can conclude our contributions as the following:

- We present and investigate recent advances in using security techniques in medical imaging storage such as encryption, storage distribution, watermarking, and blockchain technology.
- We derive a framework for general security management in telehealth services based on security processes and access techniques [17], [18].
- We highlight different security frameworks introduced in the field of medical imaging storage.

We discuss the available security frameworks to reduce latency and to improve security. Elliptic curve cryptography (ECC) is a public-key cryptography scheme proposed by Lawrence C. Washington in May 2003. It is regarded as a productive method for picture security with a small key size, but it is extremely difficult in terms of latency and complexity owing to the small number of processes involved in key generation, encryption, and decryption. [21], [22]. A chaotic system allows the generation of different random numbers for each image using an adaptive grasshopper optimization algorithm. This optimization technique enhances the chaotic algorithms, the same concept of generating random data as keys of images discussed. The method used is based on diffusion of medical image pixel using two rounds of high-speed scrambling. This combination is used to randomly shuffle neighboring pixels using bitwise XOR coding and modulo arithmetic. This proposed algorithm can better adapt to impulse noise and data loss interference [20], [23].

The concept of distributing images was discussed in another work [18]. The authors proposed a segmentation technique to improve cloud storage security. They split a medical image into several parts using the K-means clustering method. Each split was stored in a distinct cloud within a multi-cloud architecture at the same time. This proposed scenario and technique allowed a robust security scheme in cloud storage. The proposed algorithm consisted comparable pixels clustered in a certain place while breaking up the hidden image. Another work presented a two-stage encryption process based on DWT-DCT. The method of the proposed algorithm depends on two stages; the first one is the encryption of the medical image on the frequency domain to extract the image features. The second stage involves adding the eigenvector to the encrypted image to be stored in the database. To restore the original signal, the researchers used normalized correlation coefficient to test the similarity between the original and the encrypted images [5], [24]. Table I summarizes the development of the Internet-of-Thing (IoT) networks allowing an increase in the use of telehealth applications.

TABLE I. SUMMARY OF THE LITERATURE REVIEW

Ref. / Year	**Security type**	**Security algorithm**	**Storage media**
S. J. Sheela [25]	Encryption	CNN-CCT	Cellular neural network
I. Blanquer [1]	Encryption	Data-at-rest encryption	Cloud storage
A. Anand [20]	Encryption	Hyper-chaotic (LZW)	Cloud storage
M. Roy [26]	Encryption/Authentication	Fresnel transform, Arnold transform, and DCT	Cloud storage
U. S. Bhargavi [24]	Encryption	Decoy technique	Cloud storage
P. Preethi [27]	Encryption	Hex code-oriented encryption	Two-level encryption in cloud storage
B. A. Alqaralleh [28]	Encryption/Authentication	ECC	Cloud storage

K. Shankar [29]	Encryption/Authentication	Adaptive grasshopper optimization algorithm	Cloud storage
Z. Hua [16]	Encryption/Authentication	Two stages: High-speed scrambling and pixel adaptive diffusion	Cloud storage
M. Marwan [15]	Distributing	K-means clustering algorithm	Cloud storage
S. Wang [30]	Encryption	NCC	Two-level encryption in cloud storage

III. Medical Image Storage Security Framework

This section discusses two main proposed security frameworks. The frameworks are explained based on access security to the image storing security in terms of three main aspects of encryption, authentication, and distribution/segmentation. Both frameworks depend on using cloud storage to reduce the latency of calling images and to allow storing in big size. The first framework is shown in Fig. 2. In this framework, the authors consider using an IoT network. The security framework depends on the encryption of images and allows authentication to access these data by using six stages of the security process, which are:

- Stage 1: Medical data is transferred to be verified using device authentication (DA) between device-to-device, device-to-collection gateway, and transmit data securely through secure channel (SC) and data en/decryption (DED) during data transmission.
- Stage 2: The collected data of the encrypted image should be transmitted using contextual-based access control (CBAC) and role-based access control (RBAC) after completing DA on the collection gateway. The use of DED and SC comes from the need of more security services.
- Stages 3 and 4: The medical staff and anyone else allowed to access the stored data can serve this data by passing the user authentication (UA) test to ensure security of the stored data using DED, SC, CBAC, and RBAC.
- Stage 5: interface through telemedicine, remote monitoring, remote control and new services Remote control is a requirement in forming an important and basic factor for monitoring and controlling things remotely and is a transitional element in the control of objects and their contents
- Stage 6: The relation between DA and UA is transverse which means that DA is used in the transmitter and the UA is used in the receiver.

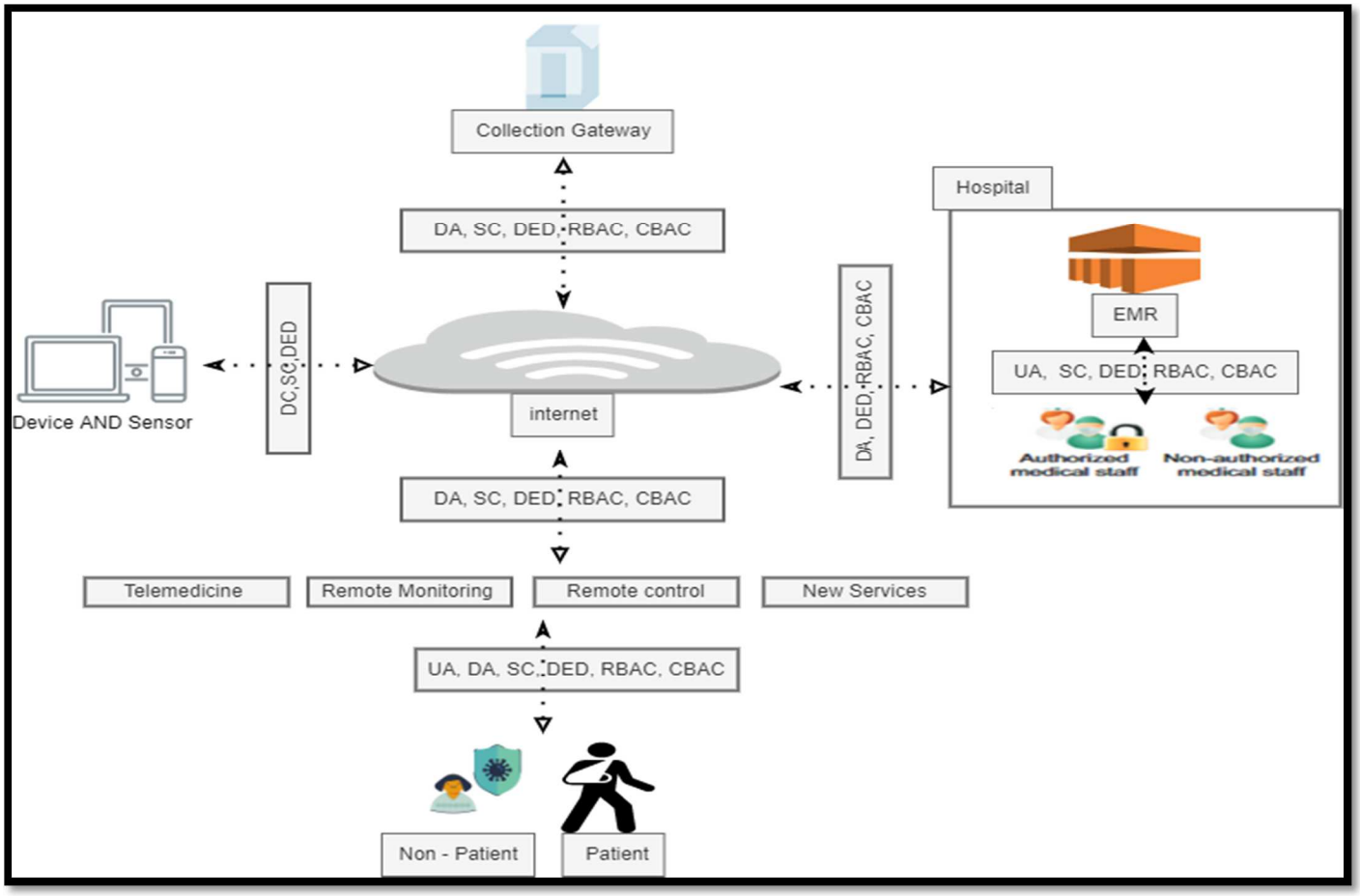


Fig. 2. Collection of Internet Structures with the Proposed Security Framework

As discussed above in the literature review section, [25, 26] proposed a segmentation technique to provide highly secure medical image storage depending on the concept of distributing images along with the cloud and segment each image to several regions. They proposed using parallel uploading and downloading of the stored data using the concept of multi-cloud environment.

This technique allows increasing the security level of the network and reducing the risks that a network can face. Fig. 3 presents another proposed approach, which has adopted a multi-region segmentation security framework. The proposed framework aims to enhance the security level and system performance using two modules, namely, Cloud Sec, for the first and main cloud, and the Cloud Slave, which is the second cloud. To enable secure data transmission, Secure Sockets Layer (SSL) protocol is used to send data to the Cloud Sec. At this time, Cloud Sec stores the data locally to ensure data privacy and security besides segmenting the data into several subsets of data to enhance confidentiality. This technique enhances the security level of the network, but it also increases the image storing process and computation, which increase the latency of the system [31], [32]

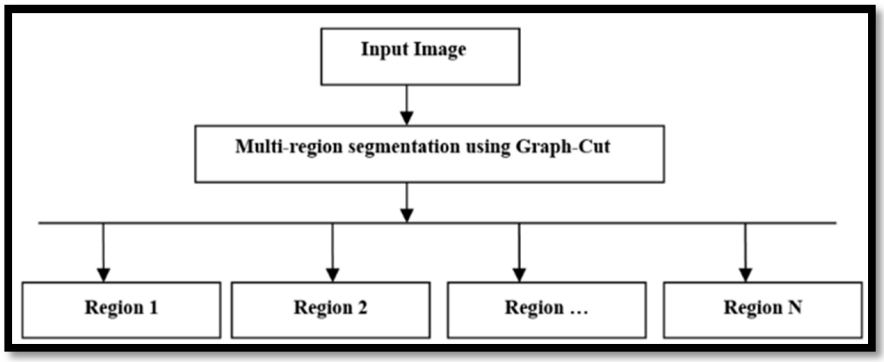


Fig. 3. Multi-region Segmentation Framework

IV. Open Research Areas

In the field of medical image storage, there are various open research problems. These difficulties and potential solutions are discussed in this section. We provide fresh challenges in our survey, which are summarized as follows:

A. Blockchain technology

This technology has gained widespread attention in cybersecurity. It can enable safe and secure healthcare data management. It can also be used to satisfy encryption, allow authentication, and make data distribution along with different cloud storage feasibility. It can also allow real-time healthcare monitoring with an up-to-date secure healthcare service [33].

B. High-speed data transfer

The increasing demand for telehealth services have increased the need for storing huge images and data, which has, in turn, necessitated high-speed connectivity. 5G can play a role in giving the terahertz connection between users and healthcare providers. The use of 5G systems can enhance the data rate transfer, but some new challenges can appear such as security, privacy, and the routing protocols to deal with a huge amount of data and dense cell deployment, which increase the security requirement

C. Big data Storage

The expected growth of IoT devices has led to having big data that needs to be stored. This storage of big data causes high latency when calling this data again to be used. This also leads to a complex process to satisfy security and increased latency at the end. Because of this, the proposed algorithms deals with big data to reduce the security process complexity by using machine learning, artificial intelligence, and blockchain technology [28].

V. Lessons Learned: Summary And Insights

In this paper, we have investigated the benefits of telehealth services in modern lifestyle and the need for enhancing the medical image storage security to reduce latency and enhance the security challenges. Furthermore, we devised a security architecture framework for medical image storage, with a discussion of some recent related works and presented some of the security frameworks used for enhancing medical image storage security. We learned the following lessons from this study:

- Data-driven algorithms can be used for cloud storage environment to enhance the security in telehealth services by incorporating blockchain, machine learning algorithms, and data analytics.
- The advanced networks like 5G can enhance the connectivity between users and medical providers and allow real-time or near-real-time treatment and monitoring services.
- The main security problems in medical data storage are latency and data size, where the former can be solved using cloud storage and terahertz networks like 5G, while the latter can be overcome by using distributed storage.
- Developing a low-complexity security process is important to enhance the connectivity of the telehealth services

VI. Conclusions And Future Recommendations

In this paper, we discussed the importance of enhancing the security of medical image storage, especially in the context of telehealth services. The benefits of cloud storage as a promising environment to reduce latency in connectivity have been discussed in addition to discussion on some security related works. The security framework shown in Fig. 1 has also been discussed in relation to access and device security. Encryption, authentication, and authorization are the three techniques to ensure security for medical image storage. A few open research topics have also been mentioned, especially the integration between security and terahertz networks such as 5G, the use of blockchain to allow security, and the need of using machine learning algorithms to deal with medical big data. Our primary research proposals for the future are to use artificial intelligence and machine learning in medical security, as they have been deemed a vital part of wireless communication and beyond networks to enable various smart applications in telehealth services. Machine learning is predicted to be utilized to facilitate the intelligent management of massive data. However, training of machine learning algorithms is required to ensure data security. The complexity of medical security is expected to be reduced in the future by using machine learning and artificial intelligence.

References

[1] I. Blanquer *et al.*, "Federated and secure cloud services for building medical image classifiers on an intercontinental infrastructure," *Futur. Gener. Comput. Syst.*, vol. 110, pp. 119–134, 2020, doi: 10.1016/j.future.2020.04.012.

[2] C. E. V. B. Hazenberg, W. B. aan de Stegge, S. G. Van Baal, F. L. Moll, and S. A. Bus, "Telehealth and telemedicine applications for the diabetic foot: A systematic review," *Diabetes. Metab. Res. Rev.*, vol. 36, no. 3, pp. 1–11, 2020, doi: 10.1002/dmrr.3247.

[3] M. T. Gatta and S. T. A. Al-Latief, "Medical image security using modified chaos-based cryptography approach," *J. Phys. Conf. Ser.*, vol. 1003, no. 1, 2018, doi: 10.1088/1742-6596/1003/1/012036.

[4] Y. C. Jonk *et al.*, "Telehealth Use in a Rural State: A mixed-methods study using Maine's All-Payer Claims Database," *J. Rural Heal.*, no. December, 2020, doi: 10.1111/jrh.12527.

[5] M. A. Saad *et al.*, "Total energy consumption analysis in wireless mobile ad hoc network with varying mobile nodes," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 3, pp. 1397–1405, 2020, doi: 10.11591/ijeecs.v20.i3.pp1397-1405.

[6] A. Kumar, K. Chellappan, A. Nasution, and R. Kanawade, "Non-invasive blood oxygenation monitoring from different sites of human body using diffuse reflectance spectroscopy: a feasibility study of diabetic foot monitoring," no. March, p. 16, 2021, doi: 10.1117/12.2585732.

[7] M. S. Fathillah, K. Chellappan, R. Jaafar, R. Remli, and W. A. W. Zaidi, "Time-frequency analysis in ictal and interictal seizure epilepsy patients using electroencephalogram," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 11, pp. 3433–3443, 2018.

[8] D. Hayn *et al.*, "Telehealth Services for home-based rehabilitation of cardiac patients," *Comput. Cardiol. (2010).*, vol. 2020-Septe, pp. 5–8, 2020, doi:

10.22489/CinC.2020.150.

[9] M. A. Saad, M. H. Ali, S. Alani, A. H. Ali, and Y. A. Hussein, "Performance evaluation improvement of energy consumption in adhoc wireless network," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 3, pp. 4128–4137, 2020.

[10] A. Mohammed *et al.*, "Weighted round Robin scheduling algorithms in Mobile ad hoc network," pp. 1–5, 2021, doi: 10.1109/hora52670.2021.9461358.

[11] R. S. Khabipov, "Development of a cloud database for storage and processing of medical images," *Biomed. Eng. (NY).*, vol. 54, no. 2, pp. 135–139, 2020, doi: 10.1007/s10527-020-09990-6.

[12] M. A. Saad, S. T. Mustafa, M. H. Ali, M. M. Hashim, M. Bin Ismail, and A. H. Ali, "Spectrum sensing and energy detection in cognitive networks," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 1, pp. 465–472, 2019, doi: 10.11591/ijeecs.v17.i1.pp464-471.

[13] J. D. Lee, T. S. Yoon, S. H. Chung, and H. S. Cha, "Service-oriented security framework for remote medical services in the internet of things environment," *Healthc. Inform. Res.*, vol. 21, no. 4, pp. 271–282, 2015, doi: 10.4258/hir.2015.21.4.271.

[14] S. Bassan, "Data privacy considerations for telehealth consumers amid COVID-19," *J. Law Biosci.*, vol. 7, no. 1, pp. 1–12, 2021, doi: 10.1093/jlb/lsaa075.

[15] M. Marwan, A. Kartit, and H. Ouahmane, "Secure cloud-based medical image storage using secret share scheme," *Int. Conf. Multimed. Comput. Syst. -Proceedings*, vol. 0, pp. 366–371, 2017, doi: 10.1109/ICMCS.2016.7905649.

[16] Z. Hua, S. Yi, and Y. Zhou, "Medical image encryption using high-speed scrambling and pixel adaptive diffusion," *Signal Processing*, vol. 144, pp. 134–144, 2018, doi: 10.1016/j.sigpro.2017.10.004.

[17] A. Kumar, K. Chellappan, A. Nasution, and R. V. Kanawade, "Diffuse reflectance spectroscopy based blood oxygenation monitoring: a prospective study for early diagnosis of diabetic foot," no. March, p. 27, 2021, doi: 10.1117/12.2583022.

[18] M. R. M. L. M. Lazim *et al.*, "Is heart rate a confounding factor for photoplethysmography markers? A systematic review," *Int. J. Environ. Res. Public Health*, vol. 17, no. 7, 2020, doi: 10.3390/ijerph17072591.

[19] K. M. Hosny, A. M. Khalid, and E. R. Mohamed, "Efficient compression of volumetric medical images using Legendre moments and differential evolution," *Soft Comput.*, vol. 24, no. 1, pp. 409–427, 2020, doi: 10.1007/s00500-019-03922-7.

[20] A. Anand and A. K. Singh, "An improved DWT-SVD domain watermarking for medical information security," *Comput. Commun.*, vol. 152, no. January, pp. 72–80, 2020, doi: 10.1016/j.comcom.2020.01.038.

[21] M. Z. Suboh, R. Jaafar, N. A. Nayan, and N. H. Harun, "ECG-based detection and prediction models of sudden cardiac death: Current performances and new perspectives on signal processing techniques," *Int. J. online Biomed. Eng.*, vol. 15, no. 15, pp. 110–126, 2019, doi: 10.3991/ijoe.v15i15.11688.

[22] C. I. Irma Syarlina, R. Suzaimah, W. Muslihah, and H. Nor Asiakin, *Technology Adoption Models: Users' Online Social Media Behavior Towards Visual Information*, no. March. 2021.

[23] A. Z. Sameen, R. Jaafar, E. Zahedi, and G. K. Beng, "A novel waveform mirroring technique for systolic blood pressure estimation from anacrotic photoplethysmogram," *J. Eng. Sci. Technol.*, vol. 13, no. 10, pp. 3252–3262, 2018.

[24] U. S. Bhargavi, S. Gundibail, K. N. Manjunath, and A. Renuka, "Security of medical big data images using decoy technique," *2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019*, pp. 310–314, 2019, doi: 10.1109/ICACTM.2019.8776696.

[25] S. J. Sheela, K. V. Suresh, D. Tandur, and A. Sanjay, "Cellular neural network-based medical image encryption," *SN Comput. Sci.*, vol. 1, no. 6, 2020, doi: 10.1007/s42979-020-00371-0.

[26] M. Roy, K. Mali, S. Chatterjee, S. Chakraborty, R. Debnath, and S. Sen, "A study on the applications of the biomedical image encryption methods for secured computer aided diagnostics," *Proc. - 2019 Amity Int. Conf. Artif. Intell. AICAI 2019*, pp. 881–886, 2019, doi: 10.1109/AICAI.2019.8701382.

[27] P. Preethi and R. Asokan, "A high secure medical image storing and sharing in cloud environment using hex code cryptography method-secure genius," *J. Med. IMAGING Heal. INFORMATICS*, vol. 9, no. 7, pp. 1337–1345, Sep. 2019, doi: 10.1166/jmihi.2019.2757.

[28] B. A. Y. Alqaralleh, T. Vaiyapuri, V. S. Parvathy, D. Gupta, A. Khanna, and K. Shankar, "Blockchain-assisted secure image transmission and diagnosis model on Internet of Medical Things Environment," *Pers. Ubiquitous Comput.*, 2021, doi: 10.1007/s00779-021-01543-2.

[29] K. Shankar, M. Elhoseny, E. D. Chelvi, S. K. Lakshmanaprabu, and W. Wu, "An efficient optimal key based chaos function for medical image security," *IEEE Access*, vol. 6, pp. 77145–77154, 2018, doi: 10.1109/ACCESS.2018.2874026.

[30] S. Wang, "An encrypted medical image ?etrieval algorithm based on DWT-DCT frequency domain 1Chunyan," pp. 135–141, 2017.

[31] N. K. Al-qazzaz *et al.*, "Role of EEG as biomarker in the early detection and classification of dementia. - PubMed - NCBI," *Sci. World J.*, vol. 2014, pp. 1–17, 2014, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25093211.

[32] M. Z. Suboh, R. Jaafar, N. A. Nayan, and N. H. Harun, "Shannon energy application for detection of ECG R-peak using bandpass filter and stockwell transform methods," *Adv. Electr. Comput. Eng.*, vol. 20, no. 3, pp. 41–48, 2020, doi: 10.4316/AECE.2020.03005.

[33] A. Eggerth, D. Hayn, and G. Schreier, "Medication management needs information and communications technology-based approaches, including telehealth and artificial intelligence," *Br. J. Clin. Pharmacol.*, vol. 86, no. 10, pp. 2000–2007, 2020, doi: 10.1111/bcp.14045.

Saturation and Vibration Behavior of High Curent-High Frequency Hybrid Core Structure Inductors

Funda Battal
Electrical Engineering
Bandirma Onyedi Eylul University
Bandirma/Balikesir, Turkey
fbattal@bandirma.edu.tr

Selami Balci
Electrical Electronics Engineering
Karamanoğlu Mehmetbey University
Karaman, Turkey
sbalci@kmu.edu.tr

Necmi Altın
Electrical-Electronic Engineering
Gazi University
Ankara, Turkey
naltin@gazi.edu.tr

İbrahim Sefa
Electrical-Electronic Engineering
Gazi University
Ankara, Turkey
isefa@gazi.edu.tr

***Abstract*— In this study, high-current high-frequency hybrid core design and presented. The proposed design is compared with the conventional single-material core KoolMu and nanocrystalline designs, and obtained results are compared. In the hybrid core design, the air-gaps in the nanocrystalline core are replaced with KoolMu block cores. Finite element analysis (FEA) based software is used to test performances of these core designs. First, saturation behavior of three core structures is compared, and then harmonic response analysis is performed to determine the cores vibration behaviors. The obtained results show that the magnetic and mechanical qualifications caused by air-gaps. Vibration amplitude of hybrid core design is lower than gapped nanocrystalline and KoolMu core structure inductors. Although the inductance value is small at low currents, when the current increases, the inductance value becomes higher than other core designs, and a softly roll-off value is obtained.**

Keywords— Inductor design, hybrid core, powder core material, nanocrystalline core material, vibration.

I. INTRODUCTION

The use of high-power and high-frequency inductors is common in the grid/load-side filter circuits in power electronics circuits of systems such as grid interfaces of renewable energy sources, AC motor drives and electric vehicles. Inductors used in high-power medium-/high-frequency converters are required to operate at high efficiency under predetermined operating conditions. Magnetic components in power converters are one of the important factors in achieving the desired power density with the smallest possible volume [1]. The cost of an inductor with these features is also one of the parameters to be considered in the design. The core material is one of the most important selection criteria at this point.

Generally, a single core material is used in the cores of inductors designed as one- or three-phase. An air-gap must be used in such inductor core structures to improve electromagnetic performance. However, these air-gaps used in the core can cause some undesired effects both electrically (such as power losses and temperature increases) and mechanical (such as vibration and acoustic noise) depending on their length.

In recent years, there are lots of core designs in the literature in which amorphous, ferrite and nanocrystalline magnetic materials are used as single core materials for inductors. However, in inductors designed with these core materials, which have different electromagnetic properties, the collapse in the inductance value (roll-off values) due to the increase in DC current exhibit different behaviors compared to each other. This becomes an even more important problem in high-power and high-frequency applications [2], [3].

The use of ferrite core in inductors provides lower copper losses than air-gapped core inductors and lower core losses than laminated-iron core inductors. In addition, due to the high electrical resistance of ferrite cores, the eddy current is limited, and they are suitable for operation at high frequencies. On the other hand, the use of ferrites is limited in achieving high power densities and is generally preferred in applications where small core size is required, such as high-frequency filter circuits or small-inductance reactors. For these reasons, air-gapped core and laminated-iron-core inductors are preferred in the medium/high power range. [4]. Nanocrystalline materials provides lower core loss without the need for air-gaps compared to other mid-/high-frequency core materials. However, in high current applications, the large air-gap requirement causes fringing flux, which results in gap loss at higher values than core losses [5–7].

Generally, air-gapped core designs are made with core materials such as amorphous, nanocrystalline and Si-Fe in high-power applications. In order to minimize fringing flux and the negative effects caused by it, an approach is adopted to combine air-gap cores with core materials with different permeability. For this purpose, special composites with distributed air gaps such as powder core material (Kool Mu or MPP) are used in some special designs. However, powder core materials have relatively high core loss, and their low relative permeability requires a higher number of turns to achieve the same inductance value. As a result, the size of the inductors increases, and a higher parasitic capacitance occurs. Sendust cores are preferred due to the advantages of high saturation flux value, relatively low power losses compared to other powder core

materials, low magnetostriction value, stable performance with temperature increase and can be produced in various shapes in order to minimize air-gap losses in inductors because they have distributed air-gapped core [8]. Since they have a higher energy storage capacity than ferrite cores with air-gaps, they can also be designed with smaller core sizes. The most important advantage of KoolMu core materials is that the air-gap loss problems are significantly avoided thanks to the distributed air-gaps [9]. In high-power and high-frequency inductors, the collapses due to roll-off behavior in the inductance value when using a hybrid core structure are not as sharp as when using a single core material. This also affects the electromagnetic behavior of the power electronic circuits in which the inductor is used.

In these cores, when the air-gap is replaced with a different core material, a smaller size inductor design can be designed since the large distance between the air-gapped core and the winding is no longer necessary. Various core designs have been presented for this purpose in the studies in [10-12]. In [10], for the design of the inductor, the core structures with air-gaps and powder core are investigated in terms of fringing flux, power losses and saturation conditions. The performances of the hybrid core structure, especially the different DC behaviors, were compared. They stated that when air-gap is used, fringing flux increases a lot and this increases power losses and, the design without air-gaps in such a core can be made with smaller size and better performance. In [11], a new design is proposed in which soft magnetic material and powder core material are used together to stabilize the saturation, which is an important design parameter to be considered in reactor cores, and to minimize power losses. It has been reported that the reactor has lower core losses thanks to the proposed structure. In [12], a hybrid core structure is proposed for the Power Factor Corrector (PFC) boost converter to reduce AC winding losses caused by high ripple currents and, fringing flux caused by the use of separated air-gaps in the core. The best possible core geometry and core material was determined compared to equivalent cores which is used air-gap, and better performance was reported owing to the hybrid core structure. In addition, additional core losses and vibration effects can occur in the core materials with distributed air-gaps. However, besides the advantage of reducing the winding losses, the result of increasing the core losses was encountered. For this reason, it is stated that it is necessary to pay attention to creating a balance between winding and core losses in the design of the composite core.

In the past literature, designs in which some of the core parts are replaced with a core material with different permeability were presented, especially in order to reduce power losses [13], [14], [15]. The use of core materials with different permeability in the composite core structure is preferred due to many advantages. Thanks to the core structure designed in this way in [14], both the eddy current losses caused by fringing flux were minimized and the inductance value did not decrease without changing the core size and number of turns. In another study on the use of hybrid core structure in inductors, core materials with different magnetostriction values were preferred and analyzes were carried out to analyze not only the saturation and power losses but also the audible noise of the core materials [13]. A hybrid core structure was proposed for the design of the toroidal core inductor in the high-frequency resonance converter in [15]. It was stated that the magnetic flux density distribution of the hybrid core, which designed as concentric, exhibits a more even and uniform distribution compared to the classical toroidal cores. It was also stated that significant reductions in both volume and core losses were achieved compared to their single-material equivalents. In these studies, although the magnetic performances of inductors using different core designs were examined in depth, the vibration conditions of these designs have not been investigated yet.

Selection of suitable core structure, core material type, air-gap design strategy and winding structure is important to obtain optimum performance from the magnetic component. However, the inductor design procedure becomes somewhat more complicated when topologies involve variable switching frequency and variable current amplitude. In this study, a hybrid core structure is proposed by placing a powder core material with lower flux density and distributed air gap feature on the core legs instead of inductors designed with air gap only with metal alloy or ferrite core material. In this way, both inductance stability will be ensured and fringing flux effects in air gaps will be minimized in high-frequency and high-power inductor design. In addition to these, high vibration levels caused by high electromagnetic forces occurring in air gaps will also be reduced at lower levels. Thus, it is aimed to reduce additional power losses, temperature increases and vibration effects in air gaps.

II. Theoretical Background on Inductor Design

Obtaining the desired inductance value and avoiding magnetic saturation are always fundamental design factors for inductor design. The air-gap in the legs and/or yokes portions of the core structure is one of the critical design factors used to achieve these goals. For example, when two different inductor designs share the same size and geometry, the core area and average length of both inductors are nearly the same. Accordingly, it means that it has a similar effective permeability according to Eq. (1) in providing the desired inductance value.

$$L = \frac{\mu_{eff}\mu_0 N^2 A_c}{l_c} \tag{1}$$

where μ_{eff} is the effective relative permeability of an inductor core without air gap; μ_0 permeability of space; N is the number of turns, l_c and A_c represent the average length and core area of the core, respectively. Also, the same effective magnetic permeability also means the same magnetomotive force (MMK) shown by Eq. (2).

$$Ni = H_g l_g = \Phi R_{eq} = \Phi \frac{l_g}{\mu_{eff}\mu_0 A_g} \tag{2}$$

where i is the current through the winding; H_c magnetic field strength; ϕ magnetic flux; R_{eq} is the effective reluctance (At/Wb) of the magnetic circuit [2].

However, in the design of the air-gapped core inductor, the desired inductance value can be written according to the effective permeability value. As seen in Eq. (3), this effective permeability (μ_{eff}) value can be expressed as a function of initial

permeability value, magnetic field strength and saturation flux density. In other words, the effective permeability of the core material can be adjusted by the air gaps in the core structure. The most well-known method to achieve this in inductors is the air-gapped core design. Owing to the air-gaps placed in the core structure, it is possible to provide the desired effective permeability value, to shift the saturation and to make the *BH* magnetization curve linear. Thus, the *BH* magnetization curve of the core material acquires a soft electromagnetic behavior. However, the distance of the air gaps in the core and the fringe flux effects of these air-gaps affect the electromagnetic performance.

$$\mu_{eff} = f\left(\mu_i, H_c, B_{sat}\right) \tag{3}$$

In large-current inductor designs (NI_{s2}), saturation is delayed by air-gaps in the core to obtain a soft saturation flux density characteristic as in Fig. 1 from the given *BH* curve for any ferromagnetic core material. However, although saturation can be delayed with air gaps, sharp saturation occurs in materials such as ferrite, amorphous, nanocrystalline and SiFe after a certain current value. Since the distributed air gap inherent exists in powder core materials such as Kool Mu, MPP and XFlux, there is no need to determine the air gap length in the core structures and a soft saturation occurs due to this feature of the material [3].

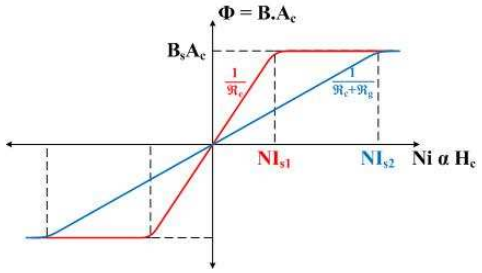


Fig. 1. Softening the B-H curve with air gaps in the core

In addition to this electromagnetic performance advantage provided by powder materials, thanks to its very low magnetostriction value, it will also ensure that vibrations arising from the ribbon structure are minimized when used as a gap filling material in ribbon wound cores.

III. Core Vibration of Inductors

As low vibration and acoustic noise as possible is desired in magnetic components operating at medium/high frequencies such as inductors. In order to keep these two parameters at the lowest levels, core materials with low magnetostriction values are preferred. Among the soft magnetic materials, core materials such as Sendust have near zero magnetostriction, so they can provide the desired performance in such applications [16].

The vibration in inductors generally occurs due to the interaction of three different forces. The first of these is the magnetostrictive forces arising from the magnetostriction value of the core material. Magnetostriction is due to the change in volume of the core material when it is in the magnetic field. Although there is a volumetric change in general, since the change in length is larger, it is only considered as the change in length of the material. Core magnetostriction can be expressed as given in Eq.4 according to the voltage applied to the windings and flux density.

$$\lambda = \frac{\Delta l}{l} = \frac{\lambda_S B_0^2}{B_S^2} cos^2(\omega t) = \frac{\lambda_S U_0^2}{(N_1 A \omega B_S)^2} cos^2(\omega t) \tag{4}$$

where λ_s and B_s are magnetostriction value and saturation flux density of the core material, respectively. As it is seen from equation, the magnetostriction varies proportionally with the square of the magnetic flux density (there is a nonlinear relation between them) [17, 18].

By using Eq.4, the vibration acceleration of the core caused by the magnetostriction can be derived as given in Eq.5.

$$a_c = \frac{v}{t} = \frac{d^2(\Delta l)}{dt^2} = -\frac{2\lambda_S l U_0^2}{(N_1 A \omega B_S)^2} cos^2(\omega t) \tag{5}$$

In Eq.5 which expresses vibration acceleration, ω indicates the angular frequency and can be written as $\omega = 2\pi f$. f is operating frequency of the inductor, and l is the total magnetic path length of the core. As it can be seen from Eq.5, the amplitude and frequency of the core vibration acceleration vary linearly with the amplitude of the voltage and the magnetostriction value. However, since it changes with the square of the angular frequency, the fundamental components of the vibration acceleration of the core and magnetostriction are twice the operating frequency [19, 20].

The second force that is the source of vibration in the inductor core is the Maxwell Force. Since a high amount of energy is stored in the air-gapped core, a higher level of force occurs compared to the powder core. In order to analyze the Maxwell forces occurring in the inductor core, firstly the Maxwell Stress Tensor given in Eq.6 is used [21];

$$\vec{T} = \frac{1}{\mu_0}\begin{bmatrix} B_x^2 - \frac{1}{2}B^2 & B_xB_y & B_xB_z \\ B_yB_x & B_y^2 - \frac{1}{2}B^2 & B_yB_z \\ B_zB_x & B_zB_y & B_z^2 - \frac{1}{2}B^2 \end{bmatrix}\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{6}$$

where, magnetic flux density can be written as $B^2 = B_x^2 + B_y^2 + B_z^2$. Using the Maxwell Stress Tensor, on a surface A the 3D Maxwell force occurring in the core can be written as in Eq.7;

$$\vec{F} = \int_A \vec{T} dA = \frac{1}{\mu_0}\int_A \begin{bmatrix} B_x^2 - \frac{1}{2}B^2 & B_xB_y & B_xB_z \\ B_yB_x & B_y^2 - \frac{1}{2}B^2 & B_yB_z \\ B_zB_x & B_zB_y & B_z^2 - \frac{1}{2}B^2 \end{bmatrix} \times \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} dA \tag{7}$$

As seen in the above equations, the forces generated in the core are directly related to the Maxwell Stress Tensor, that is, the components of the flux density in all three directions. Therefore, the core forces can be changed either by changing the size of the core, especially the gaps, by a redesign [22] or by changing the value and direction of the flux density. Thus, a decrease in the vibration level can be achieved by reducing the effect of these two forces [23].

The third force is the Lorentz forces, also known as the Laplace forces. The current flowing through the windings of an inductor causes a force to be generated on these windings. The generated force is expressed as the cross product of the current density and the magnetic flux density, as seen in Eq. 8:

$$f = J \times B \tag{8}$$

where J is the current intensity and B is the magnetic flux density [20].

As can be seen from these formulas given above, the magnetostrictive, Maxwell forces and Lorentz force are related to the magnetic flux density value. Therefore, it can be said that the flux density and its direction have an effect on these forces.

Since the displacements of the ribbon -structured core are in opposite directions, especially in the air-gapped core, the stiffness of the filling material placed in the air-gap region is important. However, adding a ferromagnetic material that will both reduce fringing flux and have a higher permeability than air will reduce excessive forces and flux density harmonics that may occur in this region.

The forces mentioned above can be written in the frequency domain as given in Eq. (9):

$$F(\omega) = F_{EM}(\omega) + F_{ms}(\omega) \tag{9}$$

Accordingly, the displacement in the frequency domain is calculated using as given in Eq. (10):

$$F_{EM}(\omega) + F_{ms}(\omega) = \boldsymbol{u}(\omega)(K - \omega^2 M) \tag{10}$$

where ω is the angular frequency and M is the mass matrix [24].

IV. SIMULATION STUDIES

In this study, as seen in Fig. 2, harmonic response analysis of three different inductor cores, KoolMu, nanocrystalline and hybrid cores, is performed. In Table 1, the properties of the inductors are given. Due to the excitation waveform of the inductors, the vibration harmonic spectrum is considered in two different frequency ranges, low frequency and high frequency range.

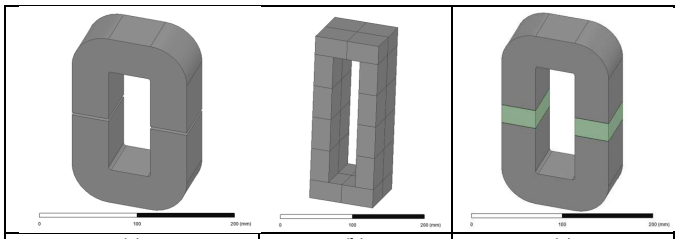


Fig. 2. Inductor core designs, a) nanocrystalline core, b) powder core (KoolMu), c) hybrid core

The incremental inductance stability of these inductors due to DC current is determined by Maxwell parametric analysis. Accordingly, when a comparison is made in terms of roll-off values, as can be seen in the graph given in Fig. 3, it has been determined that the hybrid core design has a softer roll-off value. Thus, a good roll-off value can be achieved in the hybrid core, which is a combination of air-gapped C-core with a sharp collapse and KoolMu core structures with a soft collapse.

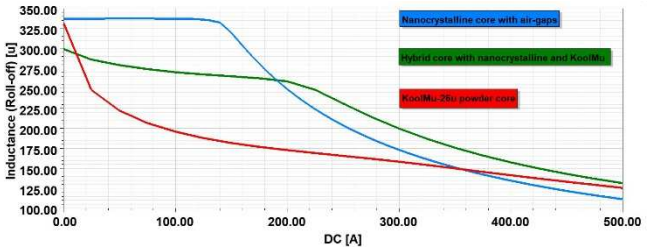


Fig. 3.Comparison of core structures according to roll-off values

When the flux distributions of the cores are investigated, the results given in Fig. 4 and obtained for maximum 1T are given. Fringing flux effects are observed in the gaps and the areas where saturation is reached in the air gap core. The saturation flux density value is 1.05 T for KoolMu and 1.24 T for nanocrystalline core material. Thus, it can be seen that the saturation flux density value is approached in the parts where there are air gaps in the core according to the flux distributions. On the contrary, the distributed air gap feature of the KoolMu core is an extremely good advantage for the powder core. For the hybrid core, the disadvantage in the air-gapped core structure has been eliminated and the effect of the combination with the powdered core is clearly visible.

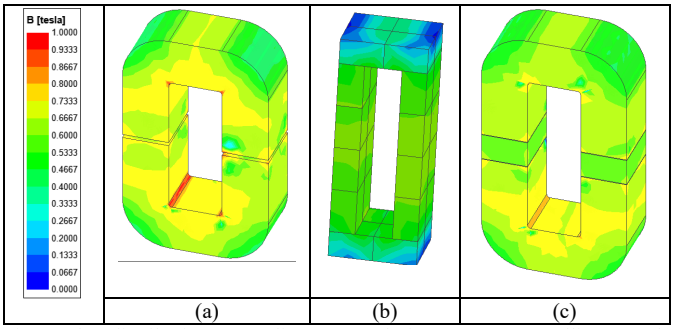


Fig. 4. Flux distributions according to core designs a) nanocrystalline core, b) powder core (KoolMu), c) hybrid core

TABLE I. PROPERTIES OF INDUCTORS

Parameter	**Value**
I_L	235A
$\Delta I_{ripple(p-p)}$	%20
L	200µH
f_{sw}	10kHz
f_0	50Hz
Coil Material	Aluminium
Core material	Powder material (KoolMu-26u)
	Nanocrystalline core (Vitroperm 250F [25])
	Hybrid (Nanocrystalline + KoolMu)

In Fig. 5, harmonic spectra of low-frequency vibration accelerations in the range of 0Hz-1000Hz are shown. In this range, KoolMu has the lowest vibration amplitude and harmonic content due to its core block structure. It is seen that

nanocrystalline core has higher amplitude and more harmonic components than other core structures due to its ribbon structure in all x-, y- and z-directions in the harmonic spectrum in this range. Due to the use of block core in the hybrid core, a case similar to KoolMu core is obtained. From these results, the most important advantage of using KoolMu core in gaps is that a hard filling material is used as a gap filling material, which can be considered as a material that can be damped vibration more, and that a material with a block structure is used instead of a ribbon structure. In both cases, the core vibration is more damped than when only the nanocrystalline core is used, resulting in a less complex spectrum. Another striking point in the low frequency acceleration harmonic spectrum is the following; In the z-direction vibration accelerations, the amplitude of the fundamental component was the lowest in the hybrid core. Therefore, it shows that the vibrations occurring in the vertical direction can be effectively damped by the block core. For example; if the acceleration values of the KoolMu core are taken as reference, the nanocrystalline core has an amplitude of 10 times and the hybrid core only 1.75 times higher amplitude for the component occurring at 25Hz in the x-direction. For fundamental components at 100Hz, the nanocrystalline core is 4.14 times, while the hybrid core has only 1.79 times higher amplitude. For this frequency range, it can be concluded that the hybrid core may have lower acoustic noise than the nanocrystalline core.

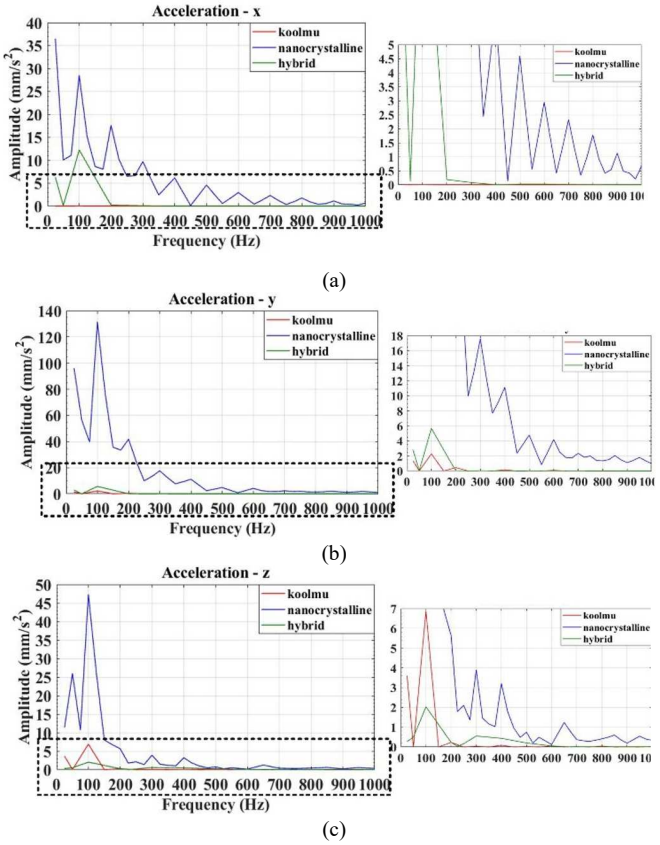
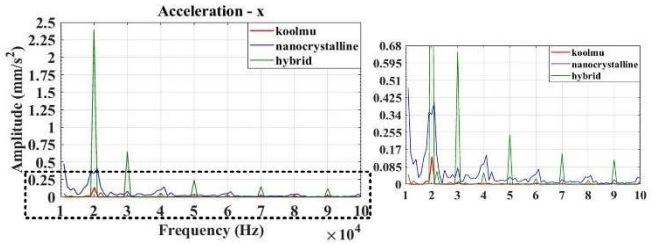


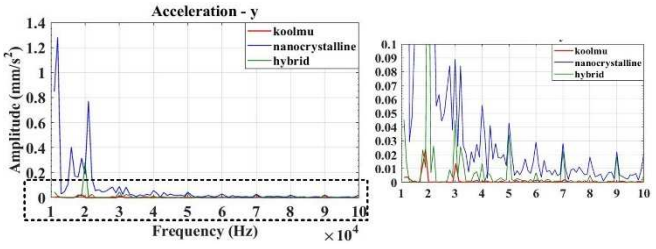
Fig. 5. Core vibrations at low frequency, a) Vibrations at x-direction, b) Vibrations at y-direction, c) Vibrations at z-direction.

High-frequency vibration acceleration harmonic spectra in the 10kHz-100kHz range are given in Fig. 6. In the acceleration harmonic components in the high-frequency range, which is quite different from the low-frequency range, the harmonic components of the hybrid core stand out in the x- and z-directions. It can be seen that in Fig.6a, in the x-direction acceleration components, the hybrid core has a prominent and the highest amplitude fundamental component relative to the other core structures. It can also be seen that there are components at frequencies of 30kHz, 50kHz, 70kHz and 90kHz, and components with very small amplitude values appear at double multiples of 20kHz. Although the nanocrystalline core has lower amplitudes of harmonic components in the x-direction compared to the hybrid core, it has significant peaks at inter-frequency values such as 11kHz, 21kHz and 41kHz. KoolMu has only the fundamental component in the core, as in the low frequency range. It can be seen that the nanocrystalline core has two peaks in the y-direction acceleration harmonic spectrum. These components also appear in the x-direction, as well as at uncharacteristic frequencies such as 12kHz and 21kHz. However, it is seen in Fig. 6b that the nanocrystalline core has harmonic components at many inter-frequencies, although the ones in the integer multiple of 10kHz are evident. However, the most striking point for the y-direction spectrum of the nanocrystalline core is its components at 12kHz and 21kHz. It can be seen that a different situation is exhibited in the harmonic spectrum in the z-direction given in Fig. 6c than in the y-direction. In the z-direction spectrum, no peaks other than 20kHz and 30kHz occurred in the hybrid core. In the nanocrystalline core, there are three distinct peaks, none of which are integer multiples of 10kHz. For the nanocrystalline core, the components at 30kHz, 50kHz, 70kHz, and 90kHz are prominent, but it can be seen to have many harmonic components at small amplitudes. Except for the fundamental component at 20kHz, the component whose amplitude is negligibly small is formed in the KoolMu core, but a simpler spectrum is obtained compared to other core structures.

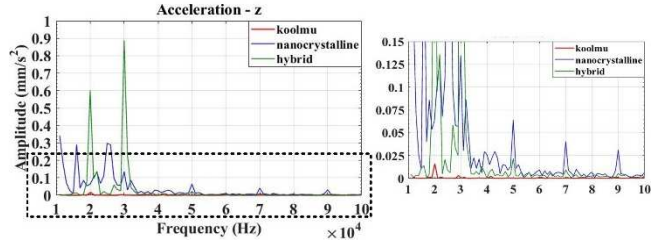
In general, the following conclusions can be drawn from the harmonic spectra of the vibration acceleration; Since the KoolMu core is in a block structure, it has negligible acceleration amplitudes compared to other cores. The nanocrystalline core, on the other hand, has the highest acceleration amplitudes, especially at frequencies after 20kHz, due to its ribbon structure. In addition, the combination of both magnetostriction and Maxwell forces led to the formation of an acceleration component at higher and many noncharacteristic frequencies. In the hybrid core, on the other hand, the ribbon structure of the nanocrystalline core minimizes the disadvantage of the KoolMu core and provides a simpler acceleration spectrum. Finally, the fact that the acceleration harmonic components in the vertical direction are lower than the nanocrystalline core shows that this core structure is more suitable from a mechanical point of view.



(a)



(b)



(c)

Fig. 6. Core vibrations at high frequency, a) Vibrations at x-direction, b) Vibrations at y-direction, c) Vibrations at z-direction.

CONCLUSION

In addition to air-gapped core structures using nanocrystalline and powder core materials are also popular in industrial power inductor design. Both show different advantages and disadvantages depending on their function within a circuit. Today, this type of power inductors is becoming widespread thanks to the many advantages provided by the core structures in which such materials with different permeability are used together. In this study, a hybrid core structure inductor is designed for high-current and high-frequency applications and the resulting core vibrations of KoolMu and nanocrystalline cores are compared. Core vibrations from both the ribbon-wound core material and the air-gap forces occur at lower levels in the hybrid core because a block-structured magnetic material is used in the air-gap parts. The magnetic core material in block structure both reduces fringing flux and acts as a hard gap filling material in the air-gap parts. According to the acceleration harmonic spectrum of the hybrid core structure, it can be concluded that it will have a lower acoustic noise compared to the pure nanocrystalline core.

REFERENCES

[1] Y. Liu, K.Y. See, K.J. Tseng, R. Simanjorang, J.S. Lai, "Magnetic Integration of Three-Phase LCL Filter With Delta-Yoke Composite Core," IEEE Transactions on Power Electronics, vol. 32, pp. 3835 – 3843, May 2017.

[2] X. Guo, L. Ran, P.Tavner, "Lessening gap loss concentration problems in nanocrystalline cores by alloy gap replacement," J. Eng., pp. 411-421, 2022.

[3] P. Arıkan , S. Balcı, F. Battal , "Determination of the roll-off value in the air-gapped inductor of a DC-DC boost converter circuit with FEA parametric simulations," Balkan Journal of Electrical and Computer Engineering, vol. 8, pp. 135-141, April 2020, doi:10.17694/bajece.664044.

[4] U. Reggiani, G. Grandi, G. Sancineto, and G. Serra. Comparison Between Air-Core and Laminated Iron-Core Inductors in Filtering Applications for Switching Converters, 7th IEEE International Power Electronics Congress. Technical Proceedings., p. 9-14, CIEP 2000.

[5] N. Kurita, K. Onda, K. Nakanoue, and K. Inagaki, "Loss estimation method for three-phase AC reactors of two types of structures using amorphous wound cores in 400-kVA UPS," IEEE Trans. Power Electron., vol. 29, pp. 3657–3668, Jul. 2014.

[6] R. Lee and D. S. Stephens, "Influence of core gap in design of currentlimiting transformers," IEEE Trans. Magn., vol. 9, pp. 408–410, Sep. 1973.

[7] H. Fukunaga, T. Eguchi, Y. Ohta, and H. Kakehashi, "Core loss in amorphous cut cores with air gaps," IEEE Trans.Magn., vol. 25, pp. 2694–2698, May 1989.

[8] Y. Liu, K.Y. See, S. Yin, R. Simanjorang, C. F. Tong, A. Nawawi, J.S. (Jason) Lai. "LCL Filter Design of a 50-kW 60-kHz SiC Inverter with Size and Thermal Considerations for Aerospace Applications," IEEE Transactions on Industrial Electronics, vol. 64, pp. 8321- 8333, October 2017.

[9] Magnetics Corp. https://www.mag-inc.com/Products/Powder-Cores/Kool-Mu-Cores/Learn-More-about-Kool-Mu-Cores.

[10] P. Winkler, W. Gunther, "Using powder materials to replace air-gaps for fringing flux reduction," PCIM Europe 2017, 16 – 18 May 2017, Nuremberg, Germany.

[11] M. Dai,K. Dai,J. Zhou, Shanghai, M. Zhou, and T. Liu, "Magnetic Core Structure and Electric Reactor," Shanghai (CN), Delta Electronics (Shanghai) CO.LTD and number US 9,281,117 B2, Mar. 8, 2016.

[12] V. Leonavicius, M. Duffy, U. Boeke, S. C. O. Mathuna, "Comparison of realization techniques for PFC inductor operating in discontinuous conduction mode," in IEEE Transactions on Power Electronics, vol. 19, no. 2, pp. 531-541, March 2004, doi: 10.1109/TPEL.2003.823249.

[13] J.M. Duràn, M. Esguerra, and U.M. Gibellini, "Hybrid core for power inductor," Falco Electronics Ltd. London and number EP 2 453 450 A1, May 16, 2012.

[14] S. Chandrasekaran, V. Mehrotra, and J. Sun, "Composite Magnetic Core for Switch-Mode Power Converters," ColdWatt, Inc., Austin, TX (US) and number US 6,980,077 B1, Dec. 27, 2005.

[15] Avala, S., Yalla, N., Agarwal, P. "Hybrid Magnetic Core for Downsizing the Inductor in LLC Converter," 2022 IEEE Texas Power and Energy Conference (TPEC), 2022.

[16] Yoo, K.Y., Lee, B.K., Kim, D.H. Investigation of Vibration and Acoustic Noise Emission of Powder Core Inductors, IEEE Transactions on Power Electronics, vol. 34, pp. 3633-3645, April 2019.

[17] B. Garcia, J. C. Burgos, A. Alonso, "Winding Deformations Detection in Power Transformers by Tank Vibrations Monitoring," Electric Power Systems Research 74, pp. 129-138, 2005.

[18] K. Hong, H. Huang, J. Zhou, "Winding Condition Assessment of Power Transformers Based on Vibration Correlation," IEEE Transactions on Power Delivery, vol. 30, pp. 1735 – 1742, 2015.

[19] B. X. Du, D. S. Liu, "Dynamic Behavior of Magnetostriction-Induced Vibration and Noise of Amorphous Alloy Cores," IEEE Transactions on Magnetics, vol. 51, 2015.

[20] Battal, F., Balcı, S., Sefa, İ. "An Analysis of High Power Inductor Vibration Behavior in Terms of Air Gaps," 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), 2019.

[21] H. R. Karshenas, H. Saghafi, "Basic Criteria in Designing LCL Filters for Grid Connected Converters," 2006 IEEE International Symposium on Industrial Electronics, p. 1996-2000, 2006.

[22] M. Rossi, J. L. Besnerais, "Vibration Reduction of Inductors Under Magnetostrictive and Maxwell Forces Excitation," IEEE Transactions on Magnetics, vol. 51, 2015.

[23] Yan, R., Gao, X., Zhu, L., Yang, Q., Ben, T., Li, Y., Yang, W., “Research on Three-Dimensional Stress Distribution of Reactor Core,” IEEE Transactions on Applied Superconductivity, vol. 26, 2016.

[24] Y. Gao, K. Muramatsu, M. J. Hatim, K. Fujiwara, Y. Ishihara, S. Fukuchi, T. Takahata, “Design of a Reactor Driven by Inverter Power Supply to Reduce the Noise Considering Electromagnetism and Magnetostriction,” IEEE Transactions on Magnetics, vol. 46, pp. 2179- 2182, 2010.

[25] Vitroperm 250F Magnetic Properties, https://vacuumschmelze.com/Nanocrystalline-Material

Follow our next events from
www. ecres.net
www.icmece.org