# ENP Engineering Science Journal

**Edited by Ecole Nationale Polytechnique, 10 Avenue des Frères Oudek, El-Harrach, 16200, Algiers, Algeria.**

## SCOPE

Ecole Nationale Polytechnique (ENP) is an academic and research institution where thousands of Engineers and PhDs have been educated during the last 6 decades. Its alumna and alumnus have had a positive impact on industry and science all over the world. ENP aims to disseminate high-quality academic work to contribute to technology and science advances for humanity.

The ENP Engineering Science journal (ENPESJ) is a peer-reviewed open-access journal. It publishes reviews and full original research papers covering all aspects of recent advances in engineering dealing with:

- Electrical and Computer Engineering
- Industrial Engineering
- Mechanical and Materials Engineering
- Civil and Hydraulic Engineering
- Chemical and Environmental Engineering

The journal would focus also on technology with industry applications, it plans to publish special issues on up-to-date emerging technologies and new research directions. ENPESJ promotes Open Science through truly free open access peer-reviewed journal articles.

**EDITOR-IN-CHIEF**

Adel BELOUCHRANI - Ecole Nationale Polytechnique, ALGERIA

**EDITORIAL OFFICE**

Mourad ADNANE - Ecole Nationale Polytechnique, ALGERIA

Ouiza Satour - Ecole Nationale Polytechnique, ALGERIA

Maliha BOUAYED - Ecole Nationale Polytechnique, ALGERIA

**ASSOCIATE EDITORS**

Karim ABED-MERAIM - Université d'Orléans, Polytech Orléans, FRANCE

Taous Meriem LALEG-KIRATI - KAUST, SAUDI ARABIA

Abdelouahab MEKHALDI - Ecole Nationale Polytechnique, ALGERIA

Riad HARTANI - Xona Partners, USA

Azzedine BOUSSEKSOU - CNRS LCC TOULOUSE, FRANCE

Ahmed BOUBAKEUR - Ecole Nationale Polytechnique, ALGERIA

Abderrahmane BEROUAL - Ecole Centrale De Lyon, FRANCE

Mohamed BOURI - Ecole Polytechnique Fédérale de Lausanne, SWITZELAND

Saadia BENMAMAR - Ecole Nationale Polytechnique, ALGERIA

Manu HADDAD - Cardiff University, UK

Enrique GONZALEZ SOSA - Autonomous University of Querétaro, MEXICO

El Hadi BENYOUSSEF - Ecole Nationale Polytechnique, ALGERIA

Amar KHENNANE - UNSW Canberra, AUSTRALIA

Mohamed OUZZANE - University of Madinah, SAUDI ARABIA

Abdeldjelil BELARBI - University of Houston, USA

Rachid IBTIOUEN - Ecole Nationale Polytechnique, ALGERIA

Nicolas ABATZOGLOU - University of Sherbrooke, CANADA

Lotfi TOUBAL - University of Quebec at Trois-Rivieres, CANADA

Jean Marc CADOU - Université Bretagne Sud, FRANCE

Mohamed HABOUSSI - University of Paris 13, FRANCE

Amina SERIR - University of Science and Technology Houari Boumediene, ALGERIA

Ahmed Reda YEDDOU - Ecole Normale Supérieure de Kouba, ALGERIA

# CONTENTS

# Challenges for Future Flexible Electric Power Systems

Ali Feliachi

**Abstract−This paper describes some of the challenges that face the operation of future electric power systems. These systems are becoming more flexible and agile. Their physical structures and connections are continuously changing as microgrids, electric vehicles, and other generation and storage devices are connected/disconnected from the grid, which result in new challenges for the operation, management, and control of the systems of the future that incorporate active participation of the consumers, and high penetration of intermittent nature renewable resources such as wind and solar.**

*Keywords−* **Electric Power Systems, Renewable Energy, Smart Grids.**

### NOMENCLATURE

| | |
|---|---|
| ACE | Area Control Error. |
| DG | Distributed Generator. |
| HVDC | High Voltage Direct Current. |
| ISO | Independent System Operator. |
| PCC | Point of Common Coupling. |
| PV | Photovoltaic. |
| US | United States of America |

## I. INTRODUCTION

The objective of this paper is to describe some of the challenges that face the operation and control of flexible and agile power systems. Advances in power electronics, renewable, storage, computing, communication and sensor technologies and their declining cost will increase their proliferation in the power system to make it a system of systems. Decarbonization, digitization and decentralization are the main drivers behind the accelerated pace at which the system is evolving towards a system of systems. Electrification was recognized as the Greatest Engineering Achievement of the twentieth century [1] by the US National Academy of Engineering, (ahead of Automobile, Airplane, Water Supply and Distribution, Electronics, Radio and Television, Agricultural Mechanization, and Computers). Electricity is a key factor in determining the standard of living in the world. Fig.1. shows the 2018 ranking of some countries [2] (top 9 countries, Algeria and similar countries in terms of electricity consumption per capita in kWh).

The resiliency of the aforementioned system will depend on how these subsystems are operated and managed taking advantages of the newest technologies. An agile and flexible system is defined here as an interconnection of many subsystems that can "come and go" based on contingencies due to weather, resource availability, malicious attacks, or fault conditions. It can be a single system, known as the grid, or it can collapse into several islanded microgrids, with the
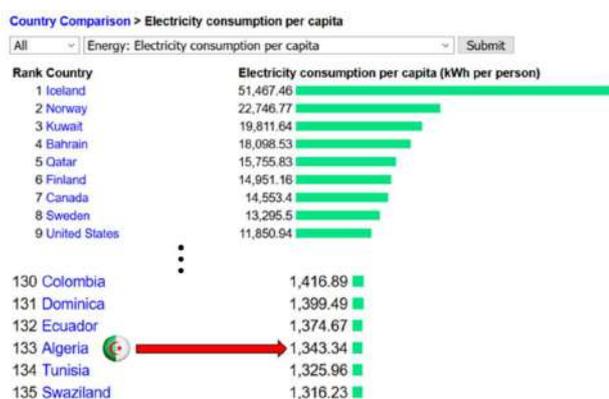
**Fig. 1:** Electricity Consumption per Capita (kWh per person)

possibility of morphing back to a single system.

The wealth of information available due to the proliferation of sensors, if utilized properly and communicated efficiently, along with automated switching, could be the key to the control and operation of smart grids of the future. However, the transition from idea to practice has proven to be quite challenging. Successful implementation requires an intelligent control system with an adaptive communication layer as well as a robust and scalable information network. The realization of such a large-scale flexible network poses some key challenges, but if addressed judiciously will result in an autonomous, resilient, and efficient power system.

This paper is organized as follows. Section II presents a sketch of the flexible power system and a smart grid architecture. Section III defines renewable resources. Challenges are presented in section IV.

## II. FLEXIBLE POWER SYSTEMS AND SMART GRIDS

A schematic of a flexible power system is shown in Fig. 1. A traditional power system is called the grid. Smart grids are defined as any additional subsystems that are added to the Grid through the Point of Common Coupling (PCC). Smart grid is loosely defined: it could be a Power Distribution Subsystem, a Consumer with storage and/or generation capabilities, etc.

A typical architecture of a Smart Grid is shown in Fig. 3. It consists of five layers. It first starts with a Power System with its physical power components and apparatusfollowed by the Communication layer which includes all the
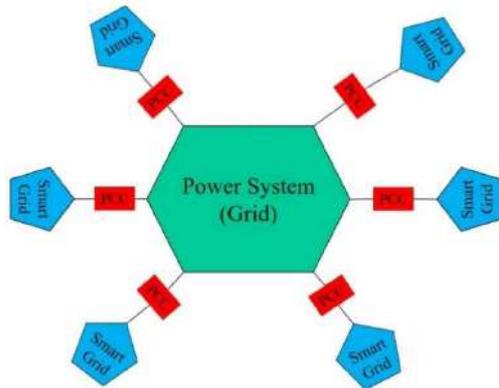
**Fig. 2:** Flexible Power System

communication devices and their protocols. The third is the Information layer that has the information flow, objects, and syntax. The fourth is the Function layer that describes the use cases, and has the supervisory control and data acquisition function, etc. The fifth is the Business layer that encompasses business cases, regulation, policies, etc. It is probably becoming necessary to add a sixth layer, a social layer, as social behavior and trends will affect this complex system.
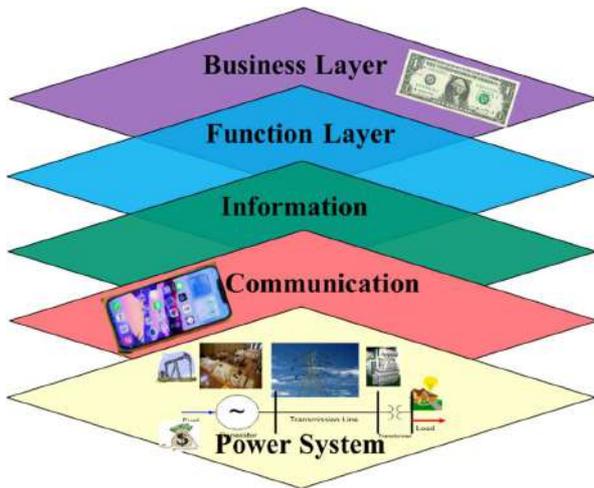


**Fig. 3:** Schematic of a Five-Layer Smart Grid Architecture

### III. RENEWABLE RESOURCES

Renewable Energy is defined as the energy from sources that are naturally replenishing but flow limited. They are virtually inexhaustible in duration but limited in the amount of energy that is available per unit of time. The major types of Renewable Resources (shown in Fig. 4a and Fig. 4b) are:

- Biomass: (plants/animals)
  - Wood and wood waste
  - Municipal solid waste
  - Landfill gas and biogas
  - Ethanol
  - Biodiesel
- Solar
- Wind
- Hydropower
- Geothermal (inside the earth)
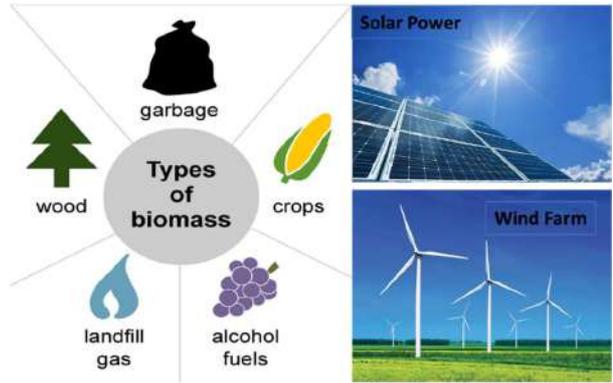


**Fig. 4a:** Renewable Energy Resources: Biomass, Solar and Wind
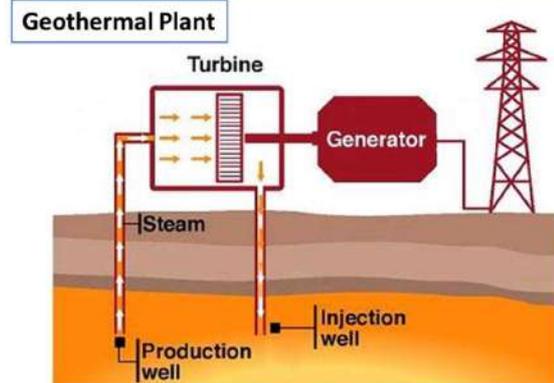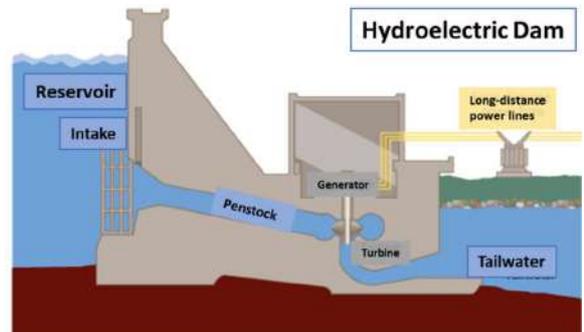


**Fig. 4b:** Renewable Energy Resources: Hydropower and Geothermal

These renewable resources are becoming a major challenge to traditional fossil fuel or nuclear units. Fig. 5. shows the US utility scale renewable generation in 2015 and 2019, especially in wind and solar, which are becoming a serious alternative to traditional fossil fuel generation.
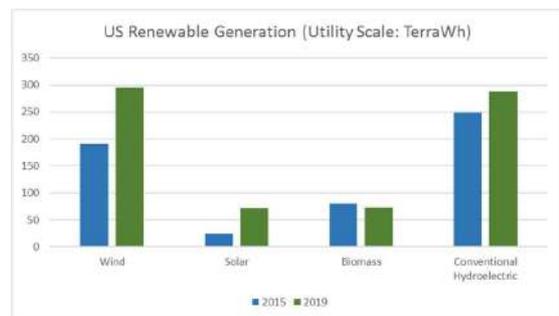


**Fig. 5.** US Renewable Generation in 2015 and 2019

The number of US PV installations is steadily increasing as shown in Fig. 6. This is due to government incentives and the decreasing cost of solar power which is now about $3 per Watt

installed [3]. The US has now surpassed two million solar installations [4].
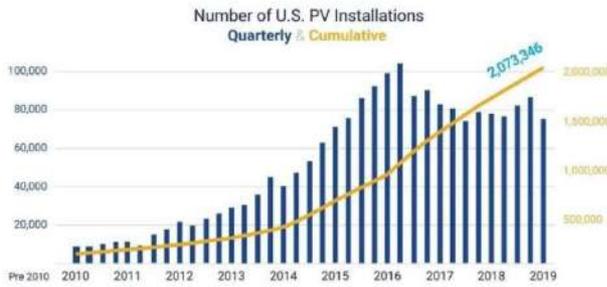


**Fig. 6.** Number of US PV Installations until 2019

## IV. CHALLENGES

Challenges facing future power systems are many and they are due to the following factors that incorporate advances in technology, economics, but also social and policy considerations:

- Economics, Deregulation, Policy, Environment
- Multiple Infrastructures
  – Electric Transportation
  – Communication, Internet, Information
- Computation
  – Data Analytics
  – Cyber Security
- Operation
  – Renewable Resources
    o Dispatch,
    o Intermittent Resources
    o Network Constraints
  – Units Ramp Up/Down: Duck Curve
- Control/Automation, Resiliency
  – Power Electronics, Sensors
  – Low Inertia, two-way power flow, switching
  – Frequency/Voltage Control

### A. Economics, Policies and Environmental Challenges

Economics, policies, and environmental issues play an important role in determining the penetration of renewable resources into the grid as well as the participation of different players in the generation and consumption of electricity. As an example, real time rates versus time of the day rates, versus fixed rates will affect consumer participation in demand response. Rooftop solar panels can become widespread if tax incentives are allowed. Electric vehicle owners can provide much needed storage to the grid if policies are in place to encourage it. Deregulation exists already in some form, but it is not common to all countries, or even amongst some states in the same country. These are challenging issues that will affect future power systems whose operation and resiliency depend on multiple market, social and engineering constraints.

A recent event in the state of Texas points to some real challenges that face the power grid. About 2 million homes in were left without electricity on Monday February 15, 2021 due to extreme winter conditions that witnessed a huge demand on the system, mainly for heating, and a diminishing supply. ERCOT, the Electric Reliability Council of Texas, the body

that manages the grid, initiated rotating outages the night before to avoid a complete collapse of the system. Many factors seem to have contributed to this event (1) ERCOT is not connected to the other two big US systems, the Western and the Eastern systems, (2) the wind generation did not produce the expected output as some wind turbines froze, (3) the gas supply was not adequate due to the freezing weather and so some gas-fired power plants were not in service, (4) the market structure and the deregulation policies might have also contributed to the problem. Some blame the inexistence of a capacity market because ERCOT has only an energy market. Hence, producers are paid for the power they generate, and so, they do not invest for extra capacity. Deregulation resulted in the creation of new energy companies that promised consumers lower prices. Unfortunately, some customers who signed up with these retail energy companies have seen bills as high as $15,000 for a month [5] instead of a bill in the $200 range! For whatever reasons, a serious problem occurred, and challenges need to be addressed to avoid similar future incidents. For this problem alone the challenges encompass economics and market structures, deregulation and policies, environment as well as multiple infrastructures such as gas, and renewable which is the case of wind here.

### B. Multiple Infrastructure Challenges

Some infrastructures have obvious effects on the power grid such as communication and information systems, hence the systems is referred to as a cyber-physical-system. But other infrastructures such as electric transportation, water, and gas are also important. In the event of Texas described in the previous section the lack of gas supply contributed to reduced production and aggravated the situation. During natural disasters, the loss of electricity has a profound effect on these other systems as most of the electric pumps for example cannot be used. Backup generation, which could represent as much as 15% of the total capacity in the US, is not used under normal conditions and if used judiciously could bring relief to the grid when the system is stressed.

In the following section, a scenario of an interaction between the grid operator and the electric transportation operator is presented. Fig. 7. is an illustration of this interaction. Each operator is responsible for unit commitments and scheduling of its own fleet. Such a problem can be optimized by exchanging price signals between the two infrastructures. Based on the load forecast, the grid operator commits generating units that can supply the load at minimum cost. The incremental cost or the price is provided to the electric fleet operator who schedules its charging and discharging electric vehicles and provides the information to the grid operator. Such an interaction could provide the most economical solution for both operators.

### C. Computation Challenges

The system, being a dynamic cyber-physical-system with potentially a huge amount of data due to the huge number of devices and sensors, requires tools to analyze and determine the situation awareness of the "Internet of Power Things" [6]. Data Analytics and Machine Learning are fields that are well suited for applications in this area, but the challenges are its unique characteristics, as for example the dynamics of the system make the availability of data sets for training basically inexistent. This poses a computational challenge, and solving it requires specialized computational tools and algorithms.

Another challenge is to expand Smart Contracts based on technologies like blockchain to advance direct deals between diverse players such as electric vehicle riders and a fleet operator along with energy traders following the scheme

depicted in Fig. 7. Blockchain has been used to link consumers and independent producers to buy and sell power. In New York City [7-8], as well as in Denmark, Blockchain is the technology adopted, and demonstrated, for rooftop solar energy trading by neighbors.
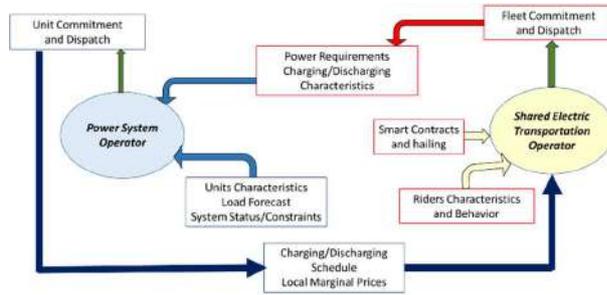


**Fig. 7.** Interaction Between Grid and Electric Fleet Operators

Finally, Cyber Security is a growing concern as daily attacks on the grid are becoming common. The Ukrainian example [9] of December 23, 2015, demonstrated that such attacks can happen. Hackers compromised information systems of three energy distribution companies in Ukraine and temporarily disrupted the electricity supply to consumers. It is the first known successful cyber attack on a power grid. Securing the system will remain a challenge as the grid becomes more flexible with a diverse number of players. This will increase the number of vulnerable points of attacks and will require novel cyber protection schemes.

### D.  Operation Challenges

#### D1. Renewable Resources Challenges

One of the challenges that result from the integration of renewable resources is their dispatchability; especially wind and solar resources; since their availability is intermittent even though their forecast is continuously improving.

Dispatchability of these renewable resources are summarized as follows:

- Dispatchable generators: those fueled by geothermal and biomass resources

- Non-dispatchable renewables: solar and wind energy. Dependent on the availability of naturally occurring resources.

- Hydroelectric generators fall between these categories; they can typically respond to dispatch signals, but they often have seasonal operational limitations that prevent them from being completely dispatchable.

Another challenge associated with renewable resources is their location. Wind farm for instance  can be either offshore or at faraway sites, such as on top of mountains where strong winds are often recorded. In order to get the power into the load centers, re-enforcement or building new lines are required. HVDC is an option that many countries are evaluating to transport the power produced by these resources.

On June 12, 2007, the *Los Angeles Times* [10] newspaper reported that *"California invested heavily in solar power. Now there's so much that other states are sometimes paid to take it. On 14 days during March, Arizona utilities got a gift from California: free solar power. Well, actually better than free. California produced so much solar power on those days that it paid Arizona to take excess electricity its residents weren't using to avoid overloading its own power lines."*

It is expected that Renewable Energy production will keep increasing as more countries are committed to this initiative. California has a mandatory 100% renewable electricity target for 2045, as well as several other measures such as storage capacity and an expected burst of electric vehicles to combat pollution. California regulators approved a historic plan to mandate rooftop solar panels on most new single-family homes that will add about $9,500 to the cost of building new houses. The solar mandate was expected to go into effect in 2020.

#### D2. Units Ramp Up/Down: Duck Curve

From an operation point of view, incorporating these intermittent resources demands sharp ramping up and down of the traditional units, for example, during sunrise when solar power becomes available and before sunset when solar power is diminishing. The shape of the load curve has the form of a "duck," hence named 'duck curve' as shown in Fig. 8.
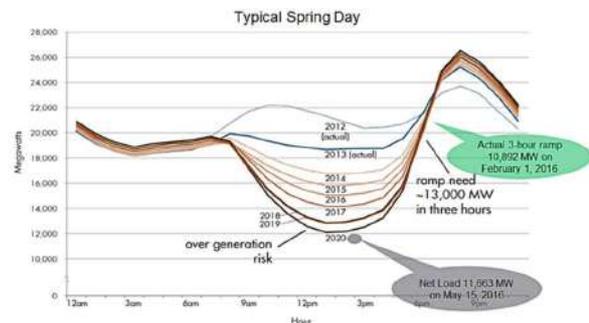


**Fig. 8.** Typical Duck Curve (California ISO)

Fig. 9. is an actual load curve from the California ISO (Independent System Operator) for the day of February 4, 2021 [11]. The green curve represents the load demand and the blue one is the electricity produced by the traditional units (net demand). The difference between the green and blue curves is the electricity produced by wind and solar generation. The average ramp-up in the afternoon for that day was very sharp, it was about 16 GW in 3 hours.
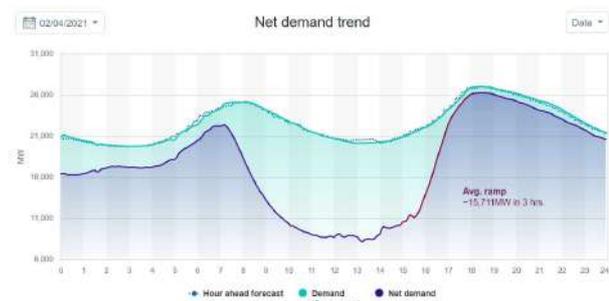


**Fig. 9.** California ISO Load Curve for February 4, 2021.

### E.  Control Challenges

The proliferation of renewable resources, which are mostly inverter-based generation, as shown in Fig. 10., results in low-inertia systems. This will affect several aspects of the control

and management of the grid including stability, load frequency control and voltage control [12-17].

In this paper load frequency control and volage control challenges are discussed.



Fig. 10. Inverter Based Generation [18]

For the load frequency control problem illustration, the utilized system consists of two areas as shown in Fig. 11. The areas are called Red and Green Areas. These colors are utilized in the figures below using red for Area 1 and green for Area 2. Blue is used for the interchange power.

Two cases are analyzed. In Case 1, Area 1 has a large inertia and Area 2 has a small inertia. In Case 2, both areas have similar inertias. The scenario under study is the effect of a load increase in Area 1.
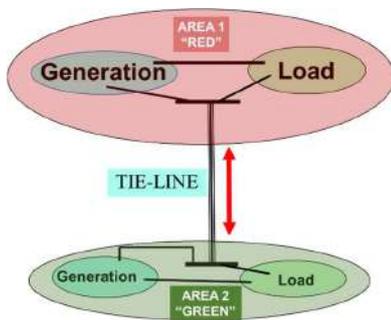


**Fig. 11**. A Two-Area Power System

An important quantity if the load frequency control problem is the Area Control Error (ACE). The statistics of ACE are a key to determining whether the areas are complying with the rules of operation so that the areas are helping each other during emergency, while each area is responsible for its own native load under normal conditions.

ACE for area (i) is defined as:

$$ACE_i = \sum_j \Delta P_{ij} + \beta_i \Delta f_i$$

Where $\Delta P_{ij}$ is the deviation from the nominal value of the power flow over the tie-line connecting area i to area j, $\beta_i$ is a bias or the frequency characteristic of area i, and $\Delta f_i$ is the deviation from the nominal value of the frequency in area i.

Simulations of the case studies described above are discussed here. Fig. 12a. shows that the frequency deviations in both

areas are unstable, while Fig. 12b. shows that the frequency goes back to its nominal value for Case 2.
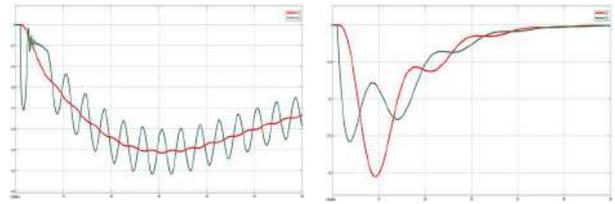


**Fig. 12a**. Frequency, Case 1          **Fig. 12b.** Frequency, Case 2

Fig. 13a. and Fig. 13b. show that the area control errors (ACE 1 and ACE 2) do not go back to zero as they are supposed to in Case 1, but they are converging, as required, in Case 2.
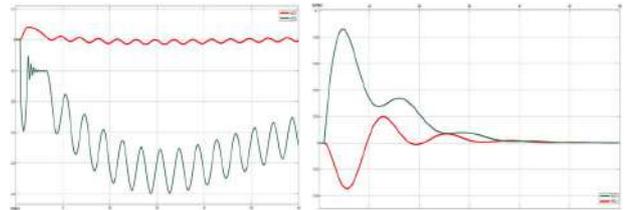


**Fig. 13a.** ACE, Case 1          **Fig. 13b.** ACE, Case 2

Fig. 14a. shows that the power exchange between the areas, $\Delta P_{12}$ does not go back to its scheduled value, but Fig. 14b. shows that the system is behaving like it is supposed to.
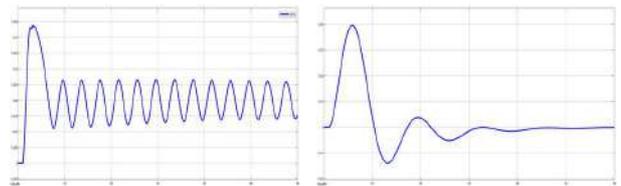


**Fig. 14a.** Tie-Line Power, Case 1          **Fig. 14b**. Tie-Line Power, Case 2

Finally, the change in mechanical power output is supposed to go to zero at steady state for area 2, which does not have a change in its demand. But Area 1 is supposed to increase its output to account for the increase of its load. Fig. 15a. and Fig. 15b. demonstrate that the results are as expected when both areas have similar sizes (Case 2), but when one area is larger than the other, in this case area 1, the results are not as expected. The challenge then is to reconsider the expectations when different size or inertia systems are interconnected. Traditional measures and expectations for the area frequency and control areas statistics need to be revisited, challenging the existing compliance performance measures.
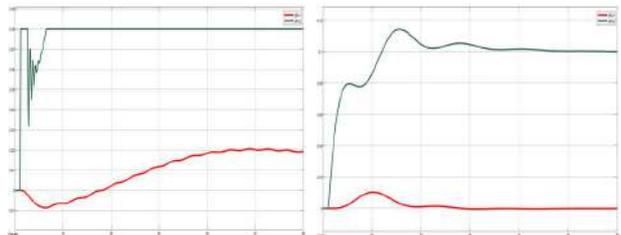


**Fig. 15a**. Mechanical Power Case 1          **Fig. 15b.** Mechanical Power Case 2

The last challenge presented in this paper is the voltage control in the presence of Distributed Generators (DGs). The DGs are mostly inverter-based generators dispersed throughout the microgrid. Therefore, the challenge addressed is to design practical distributed secondary voltage control. Because of advances in communication, one solution is to use/exploit this

infrastructure as illustrated using the system shown in Fig. 16. [19].
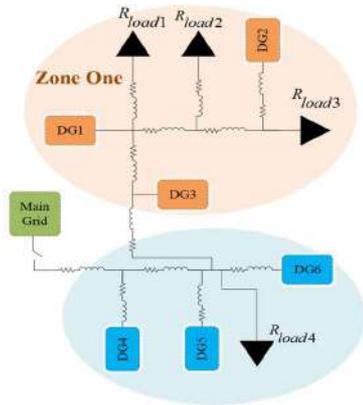


**Fig. 16.** A Six-DG-Two-Zone Microgrid

This problem has been solved using a consensus approach. Consensus algorithm is a distributed algorithm that can reach an agreement by sharing each node's voltage with its neighbors. A distributed cooperative tracking algorithm is proposed in [19]. In this algorithm, one of the DGs is selected as a leader node. This DG knows the reference value which will be communicated to the other DGs via a communication graph. The follower DGs will take that value and readjust their voltage controller set-points. All DGs must be connected to the leader via a connected graph. This algorithm works better in systems with DGs in close proximity. Instead of using the global consensus algorithm [19], the author and his colleagues have proposed a distributed or zonal consensus alternative [20].

The wireless communication network is shown in Fig. 17a. for the global consensus and in Fig. 17b. for the zonal consensus.
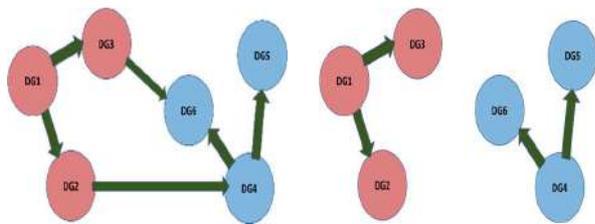


**Fig. 17a.** Global Consensus          **Fig. 17b.** Zonal Consensus

Fig. 18a. and Fig. 18b. show the controlled voltages, after a disturbance is applied, of the DGs in each zone using the distributed consensus algorithm. Zone 1 has DGs 1, 2 and 3. Zone 2 has DGs 4, 5 and 6. In both zones the voltages converge to their desired values.
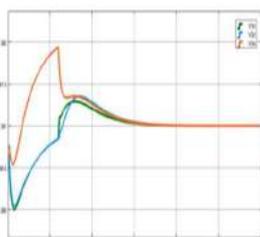


**Fig. 18a.** Zone 1 Voltages          **Fig. 18b.** Zone 2 Voltages
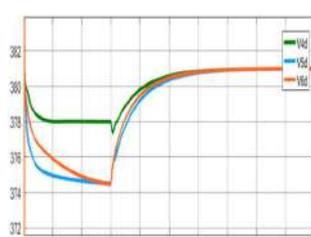
Fig. 19. shows the voltage of the six units when the global consensus algorithm is utilized. It can be concluded that similar results are obtained using the distributed algorithm, as shown in Fig. 18., an approach that uses less communication.
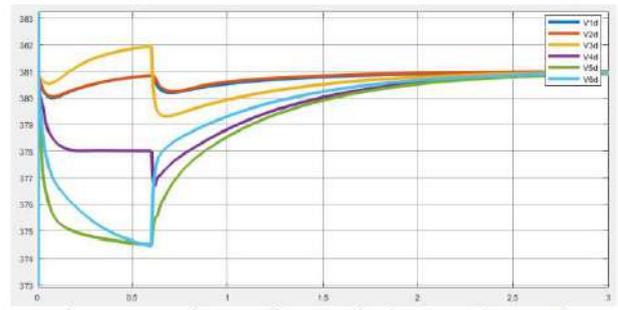


**Fig. 19.** Voltages of all Six Units using the Global Consensus Algorithm

Obviously, the challenges for voltage and frequency control are many. What is presented here is some work by the author and his colleagues and does not cover a lot of other innovations in the literature.

## V.  CONCLUSION

This paper describes some of the challenges that face the operation of future electric power systems. These systems are becoming more flexible and agile. Their physical structures and connections are continuously changing as microgrids, electric vehicles, and other generation and storage devices are connected/disconnected from the grid, which result in new challenges for the operation, management, and control of the systems of the future that incorporate active participation of the consumers, and high penetration of intermittent nature renewable resources such as wind and solar.

Additionally, numerous challenges are faced by future power systems and they are due to factors that incorporate advances in technology, economics, but also social, policy and environment factors. In this sense, this paper aimed to present some of the following challenges:

- Economics, Deregulation, Policy, Environment
- Multiple Infrastructures
    - Electric Transportation
    - Communication, Internet, Information
- Computation
    - Data Analytics
    - Cyber Security
- Operation
    - Renewable Resources
        - Dispatch,
        - Intermittent Resources
        - Network Constraints
    - Units Ramp Up/Down: Duck Curve
- Control/Automation, Resiliency
    - Power Electronics, Sensors
    - Low Inertia, two-way power flow, switching
    - Frequency/Voltage Control

The paper does not claim to be a complete presentation of the challenges facing flexible power systems, since these challenges are numerous and relate to a multitude of players and technologies. However, it should draw attention to this important and timely topic and help the reader understand the

nature of multidisciplinary aspects of this exciting field of research.

## REFERENCES

[1] https://en.wikipedia.org/wiki/National_Academy_of_Engineering#Greatest_Engineering_Achievements_of_the_20th_Century
[2] https://www.eia.gov/electricity/annual/
[3] https://solar-power-now.com/cost-of-solar/
[4] https://www.woodmac.com/news/feature/the-united-states-surpasses-2-million-solar-installations/
[5] https://theconversation.com/whats-behind-15-000-electricity-bills-in-texas-155822
[6] Abduladim Salem, "*Internet of Things in Smart Grids*," MS Project, West Virginia University, Dec 2015
[7] M.E. Peck, D. Wagman, "Energy Trading for Fun and Profit: Buy your neighbor's rooftop solar power or sell your own-solar-energy-trading-for-fun-and-profit — it'll all be on a blockchain," IEEE Spectrum, Oct 2017. https://doi.org/10.1109/MSPEC.2017.80488242
[8] https://spectrum.ieee.org/computing/networks/blockchains-will-allow-rooftophttps://seekingalpha.com/article/4262588-californias-mid-day-solar-power-glut-become-obvious
[9] https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf
[10] https://www.latimes.com/projects/la-fi-electricity-solar/
[11] http://www.caiso.com/TodaysOutlook/Pages/index.html
[12] F. Doost Mohammadi, H.K. Vanashi, A. Feliachi, "State Space Modeling, Analysis and Distributed Secondary Frequency Control of Isolated Microgrids," IEEE Transactions on Energy Conversion, Vol. 33, No. 1, pp. 155-165, March 2018. http://doi.org/10.1109/TEC.2017.2757012
[13] F. Doost Mohammadi, H. K. Vanashi, A. Dehghan Banadaki, A. Feliachi, "A novel cooperative distributed secondary controller for VSI and PQ inverters of AC microgrids," Heliyon, Elsevier Ltd., Vol. 5, Issue 6, pp. 1-14, June 2019. https://doi.org/10.1016/j.heliyon.2019.e01823
[14] A.H. Shahirina, A. Hajizadeh, D.C. Yu, A. Feliachi, "Control of a Hybrid Wind Turbine/Battery Energy Storage Power Generation System Considering Statistical Wind Characteristics," Journal of Renewable and Sustainable Energy, Vol. 4, Issue 5, 17 pages, 2012. https://doi.org/10.1063/1.4751472
[15] M.J. Ghorbani, M.A. Choudhry, A. Feliachi, "A Multi-Agent Design for Power Distribution Systems Automation," IEEE Transactions on Smart Grids, pp. 329-339, Vol. 7, No. 1, Jan. 2016. https://doi.org/10.1109/TSG.2015.2453884
[16] A. Hajizadeh, A. Feliachi, M.A. Golkar, "Intelligent Power Management of a Hybrid Fuel Cell/Energy Storage Distributed Generator," Chapter 8, pp. 265-285, Smart Power Grids 2011, Springer Verlag, Feb 2012. https://doi.org/10.1007/978-3-642-21578-0_8
[17] A. Feliachi, R. Belkacemi, "Intelligent Multi-Agent System for Smart Grid Power Management," Chapter 16, pp. 515-542, Smart Power Grids 2011, Springer Verlag, Feb 2012. https://doi.org/10.1007/978-3-642-21578-0_16
[18] https://www.cleanenergyreviews.info/blog/2014/5/4/how-solar-works
[19] A. Bidram, A. Davoudi, F.K. Lewis, SS Ge, "Distributed adaptive voltage control of inverter-based microgrids," IEEE Transactions on Energy Conversion, 2014. https://doi.org/ 10.1109/TEC.2014.2359934
[20] A. Dehghan Banadaki, A. Feliachi, V.K. Kulathumani, "Fully Distributed Secondary Voltage Control in Inverter-Based Microgrids," *IEEE PES T&D Conference & Exposition 2018*, April 16-19, 2018, Denver, CO. https://doi.org/ 10.1109/TDC.2018.8440329

## BIOGRAPHY

**Ali Feliachi** received the Diplôme d'Ingénieur en Electrotechnique from the Ecole Nationale Polytechnique of Algiers, Algeria, in 1976, and the MS and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, Atlanta, GA in 1979 and 1983 respectively. He has taught at the Ecole Supérieure de Chimie of the University of Algiers, at Lycée Ourida Meddad in El-Harrach, Algeria, and worked for SONELGAZ in 1976. At Georgia Tech, he was employed as a Graduate Teaching and Research Assistant, then a Postdoc after graduation. He worked, as a consultant, for Georgia Power Company before joining the Lane Department of Computer Science and Electrical Engineering at West Virginia University in 1984 where he is currently a Full Professor and the Director of the Advanced Power & Electricity Research Center (APERC). He also held the Electric Power Systems Endowed Chair Position for 15 years until 2017.

His research interests are modeling, control and simulation of smart grids and electric power systems. His work has been funded by electric utilities (Allegheny Power, Duquesne Light), and US DoE, US DoD, NSF, and EPRI (Electric Power Research Institute). He has published about 300 journal and conference articles in his field of expertise.

Dr. Feliachi is a *Life Senior Member of IEEE*, and a member of the honorary societies *Pi Mu Epsilon* (Math), *Eta Kappa Nu* (Electrical Engineering), and *Sigma Xi* (Research). He received the *ASEE* (American Society for Engineering Education) *North Central Section Dow Outstanding Young Faculty Award* in 1987, the following awards from the College of Engineering at West Virginia University: *Leadership Award* in 1989, *Outstanding Researcher Award* in 1991, 2004 and 2005, *Researcher of the Year* in 2005, *Outstanding Graduate Teacher Award* in 1991. In 1994 he received the Claude *Benedum Distinguished Scholar Award for the Sciences and Technology* from West Virginia University.

# Transient impedance of grounding system with impulse superimposed sinewave

Sherif S. M. Ghoneim, Ahdab M. Elmorshedy, and Rabah Y. Amer

**Abstract− Investigating the transient performance of grounding systems subject to lightning (impulse or impulse superimposed sinewave) is valuable for protecting the power system and maintaining the system operation. In this work, the grounding system's impedance is computed when an impulse superimposed sinewave is applied to the grounding grid's proposed lumped circuit andthe grounding system can be simulated as an inductance in series with resistance, and all of them are in parallel with capacitance based on Thione's assumption. Several variables were investigated to study their effects on the grounding system's behavior. The variables were the soil resistivity, soil permittivity, main wire length, grid conductor radius, grid side length, grid configurationand its mesh number. The grounding system configuration varied between square and rectangular shapes, which connects to the protecting rod via the main wire conductor. A 3.69 kA peak of impulse current was applied to avoid soil ionization. The results indicated the performance of the grounding system when subjecting to impulse current.**

*Keywords−* **transient behavior, grounding systems, impedance, step response, lightning protection**

## I. Introduction

Grounding systems are utilized as a part of a lightning protection system providing an easy path to discharging current to pass into the ground. In this sense,the impedance of the designed grounding system must be as low as possible to avoid the excessive voltage rise, which harms the equipment and individuals [1-4].

The grounding systems' analysis subjected to lightning strokes is very complicated, especially with grounding grids where much research has been addressed to explain the performance of grounding impedance of the grounding grid under lightning [5, 6]. The performance of grounding system when subjected to lightning is investigated through many techniques such as experimental works [7, 8], simplified computational methods [9], and numerical analysis [10, 11].

Some other approaches are analytical based on the circuit theory approach [12], which is based on replacing all conductor elements, including the lightning paths, with an equivalent electrical network. The circuit theory is a fast and straightforward calculation [13, 14]. The network analysis leads directly to the results in terms of currents and voltages for all interest points. The other approach is the field theory approach [15], based on the direct solution of the electromagnetic field equations about the energized conductors and all metallic structures nearby, whether directly energized or not.

In this paper, the transient impedance of the grounding grid is computed when applying the impulse superimposed sinewave current. Furthermore, the influence of some grid and soil variables on the grounding system behavior was investigated.

The variables were the soil resistivity, soil permittivity, main wire length, grid conductor radius, grid side length, grid configuration, and its mesh number.

## II. Grounding System Model

L. Thione observed that grounding conductors' response to an impulse current might be oscillatory [5]. As a result, the grounding systems' equivalent circuit should include an inductive part of the ground conductors' total inductance, directly in series to the ground resistance. Thus all of them are in parallel to ground capacitance, as in Fig. 1.
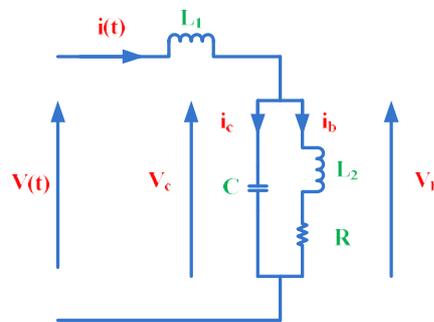


**Fig. 1.** Grounding system model.

Let the applied current, which is applied to the equivalent circuit in Fig. 1, is as follows,

$$i(t) = Ae^{-\lambda t} \sin \mu t, \qquad (1)$$

where,
$A = (\text{Imax} \times 6.45E8/3.69)kA,$
$\lambda = 1.21E5(1/s)$
and $\mu = 2.94E5 \, (rad/s)$

As the derivation in [16], the applied voltage in s domain is as follows,

$$V(S) = \frac{A\mu}{(S+\lambda)^2 + \mu^2} \times \left[ SL_1 + \frac{(SL_2 + R)}{(L_2CS^2 + RCS + 1)} \right]$$

where,

$$R = \frac{\rho}{4r} + \frac{\rho}{l} \quad \Omega$$

R refers to grounding resistance, ρ is the soil's resistivity, r is the equivalent radius of the grid, and l is the total grid length. The grid capacitance can be computed as follows,

$$C = \frac{\varepsilon\rho \times 10^{-9}}{36\pi R} \quad f$$

Where ε is the permittivity of the soil. The inductance of the grounding grid can be evaluated as follows,

$$L_2 = \frac{2l''}{3 \times 10^7} \ln \frac{4l''}{r'} \quad H$$

Where l″ is the side length of the grid and r′ is the grid radius. The inductance of the main wire can be determined as follows,

$$L_1 = \frac{2l'''}{3 \times 10^7} \ln \frac{2l'''}{r''} \quad H$$

Where l‴ refers to the main wire length and r″ is the main wire radius.

Equation (1) can be rewritten as in (2),

$$V(S) = \frac{A\mu SL_1}{(S+\lambda)^2 + \mu^2}$$
$$+ \frac{A\mu(SL_2 + R)}{((S+\lambda)^2 + \mu^2)(L_2CS^2 + RCS + 1)} \tag{2}$$

From (2), the first term can be rewritten as in (3),

$$\frac{A\mu SL_1}{(S+\lambda)^2 + \mu^2} = \left\{ \begin{array}{l} A\mu L_1\left\{\frac{(S+\lambda)}{(S+\lambda)^2 + \mu^2}\right\} \\ - A\mu L_1\left\{\frac{\mu}{(S+\lambda)^2 + \mu^2}\right\} \end{array} \right\} \tag{3}$$

Similarly, the second term in (2) can be rewritten as follows,

$$\frac{A\mu(SL2 + R)}{((S+\lambda)^2 + \mu^2)(L_2CS^2 + RCS + 1)}$$
$$= \frac{US + V}{((S+\lambda)^2 + \mu^2)} + \frac{XS + Y}{(L_2CS^2 + RCS + 1)} \tag{4}$$

where,

$$U = \frac{-A\mu RCL_2 - Y(1 - CL_2(\lambda^2 + \mu^2))}{RC - 2\lambda CL_2}, \tag{5}$$

$$Y = \frac{(RC - 2\lambda CL_2)(AL_2\mu - A\mu CR^2)}{W}$$
$$- \frac{(-A\mu RCL_2)(1 - CL_2(\lambda^2 + \mu^2))}{W} \tag{6}$$

$$W = (RC - 2\lambda CL_2)(2\lambda - CR(\lambda^2 + \mu^2))$$
$$- (1 - CL2(\lambda^2 + \mu^2))^2 \tag{7}$$

$$X = -CL_2 U \text{ and, } V = RA\mu - Y(\lambda^2 + \mu^2) \tag{8}$$

Then

$$\frac{US + V}{(S+\lambda)^2 + \mu^2} = \frac{US + V + U\lambda - U\lambda}{(S+\lambda)^2 + \mu^2}$$
$$= \frac{US + \lambda}{(S+\lambda)^2 + \mu^2} - \left(\frac{U\lambda}{\mu}\right)\left(\frac{\mu}{(S+\lambda)^2 + \mu^2}\right)$$
$$+ \left(\frac{V}{\mu}\right)\left(\frac{\mu}{(S+\lambda)^2 + \mu^2}\right) \tag{9}$$

And

$$\frac{XS + Y}{S^2 CL_2 + SCR + 1} = \frac{XS + Y}{CL_2 \times XX} \tag{10}$$

$$XX = S^2 + 2\zeta\omega_n S + \omega_n^2 \tag{11}$$

where,

$$\omega_n = \sqrt{\frac{1}{CL_2}} \quad \text{and} \quad \zeta = \sqrt{\frac{CR^2}{4L_2}}$$

Then,

$$\frac{XS + Y}{CL_2(S^2 + 2\zeta\omega_n S + \omega_n^2)} = \frac{X}{CL_2}\left(\frac{S}{XX}\right)$$
$$+ \frac{Y}{CL_2\omega_n^2}\left(\frac{\omega_n^2}{XX}\right) \tag{12}$$

hence,

$$V(S) = A\mu L_1\left(\frac{(S+\lambda)}{(S+\lambda)^2 + \mu^2}\right)$$
$$- A\mu L_1\left(\frac{\mu}{(S+\lambda)^2 + \mu^2}\right) + \frac{US + \lambda}{(S+\lambda)^2 + \mu^2}$$
$$- \frac{U\lambda}{\mu}\left(\frac{\mu}{(S+\lambda)^2 + \mu^2}\right) + \frac{V}{\mu}\left(\frac{\mu}{(S+\lambda)^2 + \mu^2}\right)$$
$$+ \frac{1}{CL_2}\left( \begin{array}{c} \frac{XS}{S^2 + 2\zeta\omega_n S + \omega_n^2} \\ + \frac{Y}{\omega_n^2}\left(\frac{\omega_n^2}{S^2 + 2\zeta\omega_n S + \omega_n^2}\right) \end{array} \right). \tag{13}$$

Then, the voltage can be determined using inverse Laplace as follows,

$$V(t) = (A\mu L + U_1)e^{-\lambda t}\cos\mu t$$
$$+ \left(\frac{V}{\mu} - \frac{U\lambda}{\mu} - A\mu L_1\right)e^{-\lambda t}\sin\mu t$$
$$+ \frac{1}{\beta}\left( \begin{array}{c} \frac{-Xe^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t - \varphi\right)}{\sqrt{1-\zeta^2}} \\ + \frac{Y}{\omega_n}\frac{e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t\right)}{\sqrt{1-\zeta^2}} \end{array} \right) \tag{14}$$

As shown in [5], the effective inductance of long grounding is

one-third of the grounding grid's total inductance, then $L_2$ should be modified to $L_2/3$ in (14). The impulse impedance is then identified as in [16] as follows,

$$Z_{imp} = \frac{\text{Crest value of Voltage}(V_m)}{\text{Crest value of Current}(I_m)} \qquad (15)$$

### III. TRANSIENT BEHAVIOR OF GROUNDING SYSTEM

Figures 2 and 3 show the applied current waveforms and the output voltage on the square and rectangular grids with a similar area. The circuit's behavior seems to be an inductive circuit where the current seems to be lag for the developed voltage.
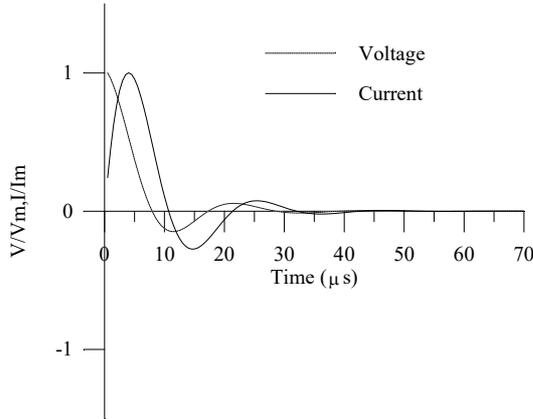


**Fig. 2.** Voltage and current waveshape for the square grid.
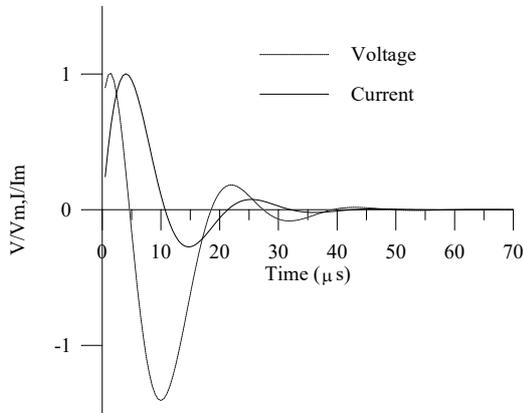


**Fig. 3.** Voltage and current waveshape for the rectangular grid.

### IV. INVESTIGATING THE INFLUENCE OF SOME PARAMETERS ON THE TRANSIENT IMPEDANCE

#### A. The Effect of Soil Characteristics

The influence of soil resistivity, permittivity on the ground system's transient impulse impedance was illustrated. The effect is shown in the following Figures 4 and 5.

The greater the resistivity, the greater the transient impulse impedance. The value of the impulse impedance as in (15) is high when the resistivity increases. The impulse impedance reaches to steady-state value after 10 μs in the case of a square grid but 8 μs in the rectangular grid as in Fig. 4.

Figure 5 shows no effect on the transient impulse impedance

with the change in permittivity, which refers to the water content.

#### B. The Effect of Grid Parameters

Figures from 6 to 10 explain the effect of the grid's side length, the radius of the conductor, the number of meshes, and the grid's width in the case of a rectangular grid.

The impulse impedance decreases as the grid side length increases. The impulse impedance with the variation of square grid length is lower than that in the side length variation of the rectangle grid.

The effect conductor radius on the impulse impedance is not significant, which is shown in Fig. 8. It is noted that the impulse impedance decreases moderately as an increase of the radius of the grid conductor (it reduces 25% when the conductor radius increases from 0.005m to 0.25m) [17].

As in Fig. 9, the results explain that the number of meshes' changes causes a slight shift in transient impulse impedance. For the rectangular grid in Fig. 10, the grid's width variation causes a significant difference in the transient impulse impedance.
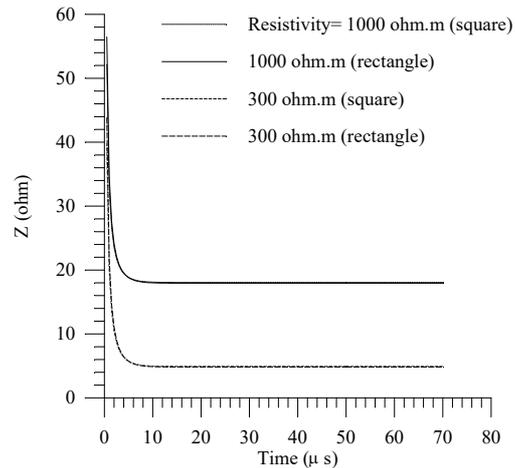


**Fig. 4.** Effect of the variation of resistivity on transient impulse impedance.
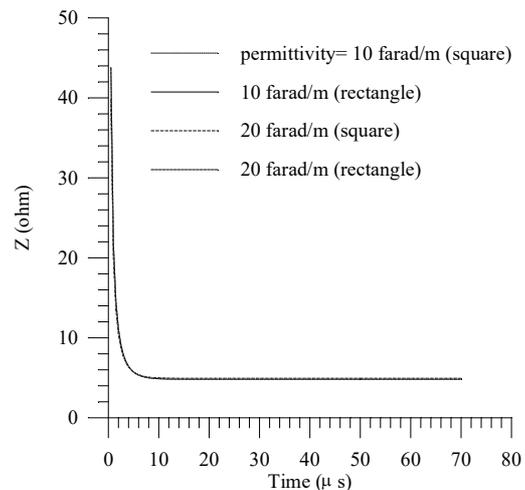


**Fig. 5.** Effect of the variation of permittivity on transient impulse impedance.
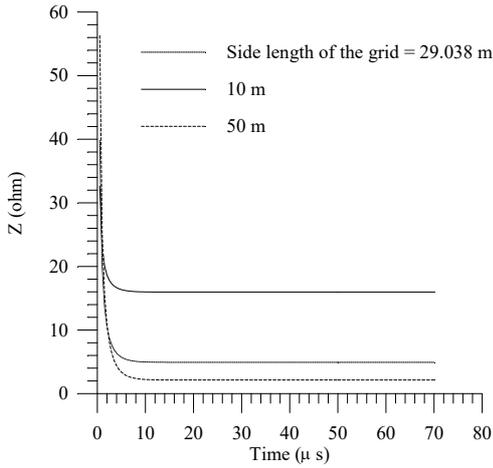
**Fig. 6.** Effect of the variation of side length on transient impulse impedance for square grid.



**Fig. 7.** Effect of the variation of side length on transient impulse impedance for the rectangular grid.



**Fig. 8.** Effect of the variation of grid conductor radius on transient impulse impedance.
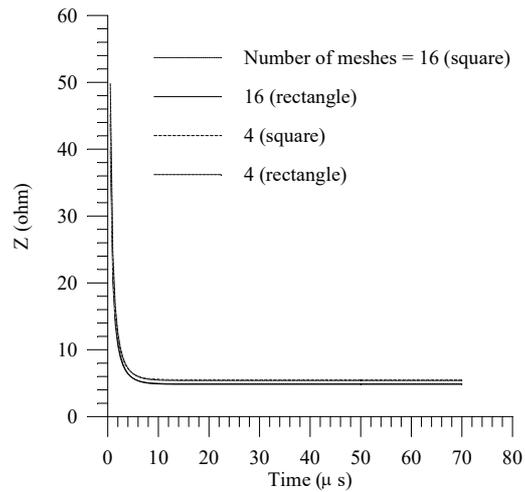


**Fig. 9.** Effect of the number of meshes on transient impulse impedance.
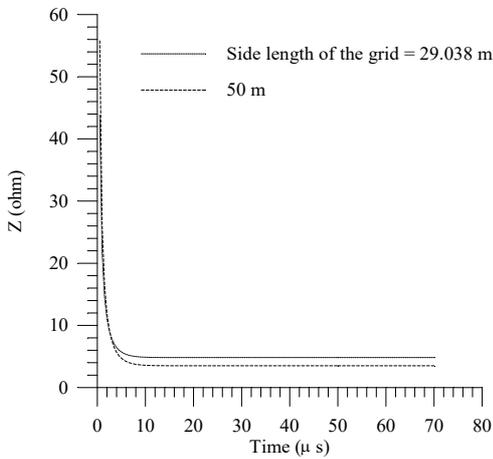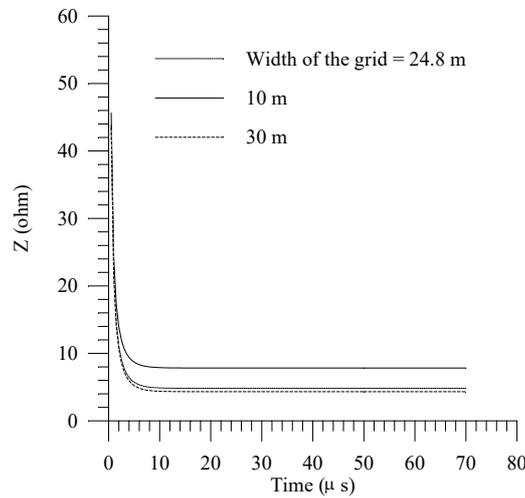


**Fig. 10.** Effect of the variation grid width on transient impulse impedance for the rectangular grid.
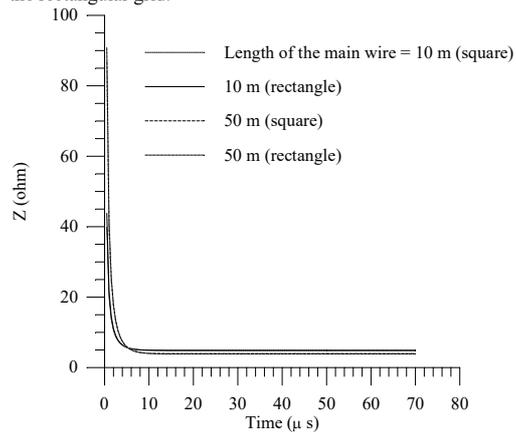


**Fig. 11.** Effect of the main wire length on transient impulse impedance.

## C. The Effect of Main Wire

The effect of the main wire is studied. The main wire is the wire that transfers the surge current to the ground system, and it is inductive, which leads to the enhancement of the transient impulse impedance of the grid. The effect of the main wire is shown in Fig.11. An increase in the length of the main wire leads to a decrease in the impulse impedance.

## V. STEP RESPONSE

The step response is used to express the grounding system's impulse impedance, which is convenient to understand the transient characteristics of the grounding system's impedance [9]. The following equation obtains the Step Response Zu(t) of a grounding impedance,

$$Z_u(s) = \frac{1}{s} \cdot \frac{L(V(t))}{L(i(t))} \qquad (16)$$
$$Z_u(t) = L - 1(Z_u(s))$$

V(t) is the measured grounding voltage, i(t) is the measured injected current, t is the time, s is the Laplace operator, and L and L-1 are the Laplace forward inverse transforms.
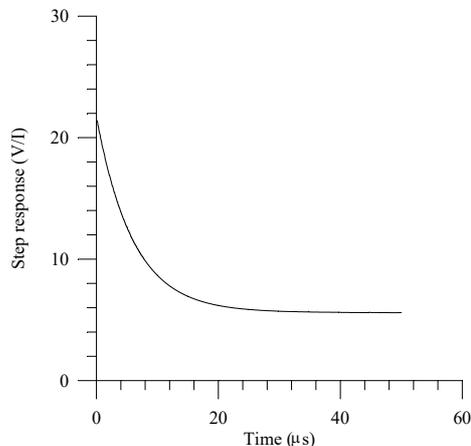


**Fig. 12.** Step response of grounding system.

## VI. CONCLUSION

The results obtained from the proposed circuit theory-based model help better understand the grounding systems' performance subjected to lightning (impulse superimposed sinewave). The step response is convenient to understand a transient characteristic of the impedance of the grounding system. When applying the proposed wave, the square grid configuration presents a low impedance for lightning in comparison to the rectangular one. The significant variation in transient impulse impedance occurs with the change of the soil's resistivity and the side length of the grid.

## REFERENCES

[1]     IEEE Standard Dictionary of Electrical and Electronics Terms, IEEE Std. 100-1992, Jan. 1993.
[2]     Ed.1: Protection Against Lightning—Part 3: Physical Damage to Structures and Life Hazard, IEC 62305-3, 2006.
[3]     IEEE Guide for Safety in AC Substation Grounding, IEEE Std.80-2000, May 2000.
[4]     R. Xiong, B. Chen, C. Gao, Y. Yi, and W. Yang, "FDTD Calculation Model for the Transient Analyses of Grounding Systems", *IEEE Trans. on Electromag. Compat.*, Vol. 56, No. 5, pp. 1155-1162, October 2014. https://doi.org/ 10.1109/TEMC.2014.2313918
[5]     R. Verma and D. Mukhedkar, "Fundamental Considerations and Impulse Impedance of Ground Grids", IEEE Trans. on Power App. and Sys., Vol. Pas-100, pp. 2053-2059, March 1981. https://doi.org/ 10.1109/TPAS.1981.316638
[6]     G. Parise, L. Parise, L. Martirano, "Intrinsically Safe Grounding Systems and Global Grounding Systems" IEEE Trans. Ind. App., Vol. 54, issue 1, pp. 25 - 31, 2018. https://doi.org/10.1109/TIA.2017.2743074
[7]     S. Visacro and R. Alipio, "Frequency dependence of soil parameters: Experimental results, predicting formula and influence on the lightning response of grounding electrodes," IEEE Trans. Power Del., vol. 27, no. 2, pp. 927–935, Apr. 2012. https://doi.org/10.1109/TPWRD.2011.2179070
[8]     Z. Zhang, Y. Dan, J. Zou, G. Liu, C. Gao, Y. Li, "Research on Discharging Current Distribution of Grounding Electrodes", IEEE Access, Vol. 7, pp. 59287 - 59298, 2019. https://doi.org/10.1109/ACCESS.2019.2914058
[9]     L. Grcev, "Modeling of grounding electrodes under lightning currents", IEEE Trans. on Electromag. Compat., Vol. 51, no. 3, pp. 559–571, Aug.2009. https://doi.org/10.1109/TEMC.2009.2025771
[10]   N. Permal, M. Osman, M. Z. Ab. Kadir, A. M. Ariffin, "Review of Substation Grounding System Behavior Under High Frequency and Transient Faults in Uniform Soil ," IEEE Access, Vol. 8, pp. 142468 - 142482, 2020. https://doi.org/10.1109/ACCESS.2020.3013657

[11]   R. Xiong, B. Chen, J. J. Han, Y. Y. Qiu,W. Yang, and Q. Ning, "Transient resistance analysis of large grounding systems using the FDTD method," *Progress Electromagn. Res.,* vol. 132, pp. 159–175, 2012. https://doi.org/10.2528/PIER12082601
[12]   O. Kherif, S. Chiheb, M. Teguar, A. Mekhaldi, N. Harid "Time-Domain Modelingof Grounding Systems' Impulse Response Incorporat ing Nonlinear and Frequency-Dependent Aspects", IEEE Transactions on Electromagnetic Compatibility, Vol. 60, No. 4, pp. 907 – 916, 2017. https://doi.org/10.1109/TEMC.2017.2751564
[13]   L. Grcev, M. Popov, "On high-frequency circuit equivalents of a vertical ground rod", IEEE Trans. Power Deliv. 20(2), 1598–1603, 2005. https://doi.org/10.1109/TPWRD.2004.838460
[14]   A. Taher, A. Said, T. Eliyan, A. Hafez, "Analysis and Mitigation of Ground Grid Lightning Potential Rise", Trans. on Elec. and Electronic Materials, pp. 1-11, 2020. https://doi.org/10.1007/s42341-020-00186-z
[15]   W. Xiong, F. P. Dawalibi, "Transient Performance of Substation Grounding Systems Subjected to Lightning and Similar Surge Currents" IEEE Trans. on Power Del., Vol. 9, No. 3, pp. 1412-1420, July 1994. https://doi.org/ 10.1109/61.311170
[16]   S. Sekioka, K. Tanemura, T. Sonpda, Y. Katyo, N. Nagaoka, and A. Ametani, "A Time- and Current-Dependent Grounding Impedance Model of a Grounding Net," International Conference on Electrical Engineering (ICEE'98) Vol. 1, pp. 832-835, 1998.
[17]   F. P. Dawalibi, D. Mukhedkar "Parametric analysis of grounding grids" IEEE Trans. on Power App. and Sys., Vol. PAS-98, No. 5, pp. 1659-1667, Sep/Oct 1979. https://doi.org/ 10.1109/TPAS.1979.319484
[18]   B. R. Gupta and V. K. Singh,"Impulse Impedance of Rectangular Grounding Grids, " IEEE Trans. on Power Del., Vol. 7, No. 1, pp. 214-218, January 1992. https://doi.org/ 10.1109/61.108910

**Sherif S. M. Ghoneim** received his B.Sc. and M.Sc. degrees from the Faculty of Engineering at Shoubra, Zagazig University, Egypt, in 1994 and 2000. Since 1996, he has been teaching at the Faculty of Industrial Education, Suez Canal University, Egypt. From 2005 to 2007, he was a guest researcher at the Institute of Energy Transport and Storage (ETS) of the University of Duisburg–Essen in Germany. In 2008, he earned his Ph.D. degree in electrical power and machines from the Faculty of Engineering, Cairo University (2008). After that, he joined Taif University as an associate professor in the Electrical Engineering Department, Faculty of Engineering. His research areas include grounding systems, dissolved gas analysis, breakdown in SF6 gas, and AI technique applications.

**Ahdab Elmorshedy** received the B.Sc., M.Sc., and Ph.D. in 1971, 1974, and 1978 respectively, in Electrical Engineering, from Cairo University, Egypt. Since 1971, she joined the faculty of Electrical Engineering at Cairo University, Egypt, as a Teaching staff. During the academic years of 1979 to 1981, she was a research scientist in the Department of Electrical Engineering, College of Engineering at Ohio State University, Columbus, Ohio, U.S.A. Since 1988 she was a Professor at the Department of Electrical Engineering, Cairo University. Her research activities include grounding, protection and safety of power systems, over-voltage transients, and pollution of insulators. She is also an IEEE Member.

**Rabah Amer** received the B.Sc., M.Sc. and Ph.D. in 1975, 1979, and 1983 respectively, in Electrical Engineering, from Cairo University, Egypt. Since 1975, he joined the faculty of Electrical Engineering at Cairo University, Egypt as a Teaching staff. Since 1994 he was a Professor at the Department of Electrical Engineering, Cairo University. His research activities include grounding, gas discharge, electric and magnetic field calculations, predicting polluted insulators flashover using laser methods, lightning protection of petroleum companies,  and overhead transmission line design and commissioning. Also, he is a consultant for petroleum, industrial and electrical companies.

# Conditional and Unconditional Deterministic Lower Bounds on the MSE of the Non-Uniform Linear Co-centered Orthogonal Loop and Dipole Array

Tao BAO and Mohammed Nabil EL KORSO

**Abstract**—The co-centered orthogonal loop and dipole (COLD) array exhibits some interesting properties, which makes it ubiquitous in the context of polarized source localization. In the literature, one can find a plethora of estimation schemes adapted to the COLD array. Nevertheless, their ultimate performance in terms of the so-called threshold region of mean square error (MSE), have not been fully investigated. In order to fill this lack, we focus, in this paper, on conditional and unconditional bounds that are tighter than the well known Cramér-Rao Bound (CRB). More precisely, we give some closed form expressions of the McAulay-Hofstetter, the Hammersley-Chapman-Robbins, the McAulay-Seidman bounds and the recent Todros-Tabrikian bound, for both the conditional and unconditional observation model. Finally, numerical examples are provided to corroborate the theoretical analysis and to reveal a number of insightful properties.

*Keywords*—Deterministic lower bounds, co-centered orthogonal loop and dipole array, mean square error, performance analysis, passive source localization, SNR threshold.

## I. INTRODUCTION

Nowadays, recent source localization systems need to operate in increasingly more crowded signal environments [2]. In this context, taking into account both the polarization diversity and the spatial diversity became ubiquitous in antenna array systems and their processing as wireless communication, radar, sonar systems, etc. [2–4]. Among different types of polarization sensitive arrays, the co-centered orthogonal loop and dipole array is commonly used since it exhibits numerous interesting properties [5–7] (e.g., the constant norm of the polarization vector, the insensibility of the polarization vector w.r.t. the source localization in the plan of the antenna etc.)

In the literature, one can find a plethora of estimation schemes adapted and/or designed particularly for the COLD array [6]. Nevertheless, their ultimate performance in terms of the mean square error (MSE), especially in the non-asymptotic region (meaning for low signal-to-noise ratio (SNR) or low observations), has not been fully investigated.

We can cite [8, 9], in which the authors derived closed-form expression of the approximated Cramér-Rao bound (CRB) for a sufficiently large number of sensors in the context of a COLD linear and uniform array. Whereas in [10,11] the authors derived, respectively, expressions of the CRB for a known single source and the resolution limit for two known sources, both for known polarization state parameters.

Nevertheless, to the best of our knowledge, no results concerning the breakdown prediction for the COLD linear array (possible non-uniform) can be found in the literature. To fill this lack, we focus, in this paper, on lower bounds that are tighter than the CRB. More precisely, we give some closed form expressions of the McAulay-Hofstetter (MCB), the Hammersley-Chapman-Robbins (HCRB), the McAulay-Seidman (MSB) bounds and a recently proposed Todros-Tabrikian bound (TTB), for both the commonly assumed conditional (i.e., when the signals are assumed to be deterministic) and unconditional (i.e., when the signals are assumed to be driven by a Gaussian random process) observation models with unknown direction of arrival (DOA) and unknown polarization state parameters. Such bounds are known to be efficient to delimit and predict the optimal operating zone of estimators [12, 13] which is given by the threshold or breakdown point, i.e., when the estimator's MSE increases dramatically.

Such deterministic lower bounds can be derived using one of the unifications given in [12, 14–16]. In this paper, we adopt the Todros and Tabrikian unification in which they propose a novel class of performance lower bounds by applying a proper integral transform [15]. Using an adequate choice of the kernel of the integral transform of the likelihood-ratio function, one obtains some well known lower bounds as the MCB, HCRB, MSB and TTB.

For the rest of this paper, the following notation will be used. A lowercase bold letter denotes a vector, and an uppercase bold letter denotes a matrix. Vectors are by default in column orientation unless specified. Upper scripts $^T$, $^C$ and $^H$ are, respectively, the transpose, the conjugate and the trans-conjugate of a matrix. The operators $tr\{.\}$, $|.|$, $\|.\|$ and $\mathcal{R}(.)$ represent, the trace, the determinant of a matrix, the Euclidean norm and the real part, respectively. $\odot$ and $\otimes$ are the Hadamard and the Kronecker product, respectively. $\boldsymbol{I}_L$ denotes the $L \times L$ identity matrix. $\boldsymbol{1}_L$ is the $L \times L$ matrix filled by ones. $[.]_i$ and $[.]_{i,j}$ denote the $i$-th element of the vector and the $i$-th row and the $j$-th column ele-
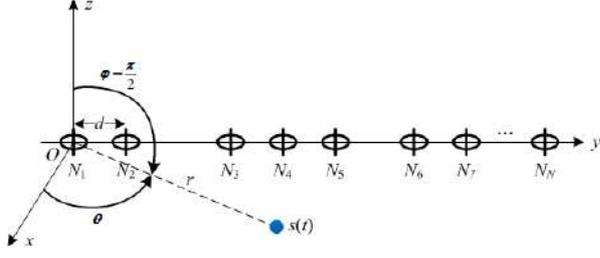
**Fig. 1**: A non-uniform linear COLD array representation in the presence of one far-field source

ment of the matrix. We define also the $n$-norm as $\sum \eta^n = ||\eta||_n$. Finally, $\hat{\boldsymbol{\xi}}$ is any unbiased estimate of $\boldsymbol{\xi}$.

## II. SYSTEM MODELS

Consider a non-uniform linear array composed of $N$ COLD pairs with inter-element spacing $d_1.d, d_2.d, \ldots, d_N.d$ that receives a signal emitted by a single far-field and narrow-band source. Let $d_n.d$ denotes the location of the $n$-th COLD sensor, in which $d$ is unit. The array is collinear with the $y$-axis of an $(O, xyz)$ coordinate system with its origin, $O$, in the center of the first pair of sensors (i.e., $d_1 = 0$). For each COLD sensor, the dipole is parallel to the $z$-axis and the loop is parallel to the $x - y$ plane.

Assume a narrowband far-field source which impinges on the array from direction described by the elevation angle $\phi$ and the azimuth angle $\theta$. In this paper, we suppose that the source is contained in the $x - y$ azimuthal plane, i.e. $\phi = \frac{\pi}{2}$, as illustrated in Fig.1. For a given polarized signal, the vertical and the horizontal components of the incoming signal electric field can be specified by polarization state parameters $\rho \in \left[0, \frac{\pi}{2}\right]$ and $\psi \in [-\pi, \pi]$. Consequently, the output of the $n$-th COLD sensor can be expressed by [6]

$$\boldsymbol{x}_n(t) = [x_{\text{loop}}, x_{\text{dipole}}]^T = s(t)\boldsymbol{u}(\psi, \rho)\, e^{j\omega d_n.d} + \boldsymbol{n}_n(t), \quad (1)$$

in which, $n = 1, \ldots, N, \quad t = 1, \ldots, T, x_{\text{loop}}$ and $x_{\text{dipole}}$ are the signals recorded on the small loop and the short dipole, respectively. The electrical angle $\omega = -\frac{2\pi d}{\lambda} \sin \theta$, where $\lambda$ is the signal wavelength. $T$ is the number of snapshots and $s(t)$ is the source signal. The random process $\boldsymbol{n}_n(t) = [n_{\text{loop}}, n_{\text{dipole}}]^T$ denotes a complex Gaussian circular noise with zero mean and a known covariance matrix $\boldsymbol{\Sigma}_{\text{noise}}$. The output vector received for the $t$-th snapshot can be written as

$$\begin{aligned} \boldsymbol{x}(t) &= [\boldsymbol{x}_1^T(t), \ldots, \boldsymbol{x}_N^T(t)]^T \\ &= s(t)\boldsymbol{e}(\omega) \otimes \boldsymbol{u}(\psi, \rho) + [\boldsymbol{n}_1^T(t), \ldots, \boldsymbol{n}_N^T(t)]^T \end{aligned}$$

where the steering vector is defined by $\boldsymbol{e}(\omega) = [1, e^{jd_2\omega}, \ldots, e^{jd_N\omega}]^T$ and the $2 \times 1$ polarization state vector is given by $\boldsymbol{u}(\psi, \rho) = [j2\pi \frac{A_{sl} \cos(\rho)}{\lambda}, -L_{sd} \sin(\rho)e^{j\psi}]^T$, in which $L_{sd}$ and $A_{sl}$ represent the length of the short dipole and the area of the small loop. From a modeling point of view, we can assume $L_{sd} = 2\pi \frac{A_{sl}}{\lambda} = 1$.

In the following, the unknown parameter vector is given by $\boldsymbol{\xi} = [\omega, \psi, \rho]^T$, whereas, $\boldsymbol{\xi}_0, \omega_0, \psi_0$ and $\rho_0$ denote the real value of the candidate parameters $\boldsymbol{\xi}, \omega, \psi$ and $\rho$, respectively. The joint probability distribution function (pdf) of the full observations

$\boldsymbol{\chi} = [\boldsymbol{x}^T(1), \ldots, \boldsymbol{x}^T(T)] \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\xi}_0), \boldsymbol{\Sigma}(\boldsymbol{\xi}_0))$ for a given $\boldsymbol{\xi}_0$, is expressed as follow

$$p(\boldsymbol{\chi}|\boldsymbol{\xi}_0) = \frac{1}{|\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)|\pi^{2NT}} e^{-(\boldsymbol{\chi}-\boldsymbol{\mu}(\boldsymbol{\xi}_0))^H \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}(\boldsymbol{\chi}-\boldsymbol{\mu}(\boldsymbol{\xi}_0))} \quad (2)$$

Let $E\left\{(\hat{\boldsymbol{\xi}}_0 - \boldsymbol{\xi}_0)(\hat{\boldsymbol{\xi}}_0 - \boldsymbol{\xi}_0)^T\right\}$ be the covariance matrix of an estimate of $\boldsymbol{\xi}_0$. Let us assume that $\hat{\boldsymbol{\xi}}$ is an asymptotically unbiased estimate of the true parameter vector $\boldsymbol{\xi}_0$, and define the CRB for the considered model [17–19]. The covariance inequality principle states that, under quite general/weak conditions [20],

$$\text{MSE} = E\left\{([\hat{\boldsymbol{\xi}}_0]_i - [\boldsymbol{\xi}_0]_i)^2\right\} \geq \text{CRB}([\boldsymbol{\xi}_0]_i), \ i = 1, 2, 3$$

where the CRB is given as the inverse of the Fisher information matrix (FIM) as $\text{CRB}([\boldsymbol{\xi}_0]_i) = [\textbf{FIM}^{-1}(\boldsymbol{\xi}_0)]_{i,i}$. Since we are working with a complex circular Gaussian observation model and using the Slepian-Bang formula [21, 22], the $i$-th, $k$-th element of the FIM for the unknown real parameter vector $\boldsymbol{\xi}_0$ can be written as

$$[\textbf{FIM}(\boldsymbol{\xi}_0)]_{i,k} =$$
$$tr\left\{\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_i}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_k}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right\}$$
$$+ 2\mathcal{R}\left\{\frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_i}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_k}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right\} \quad (3)$$

There exist two different models depending on the assumptions about the signal sources, the conditional (deterministic) and unconditional (stochastic) cases.

### A. The Conditional Case

The time-varying signal is modelled by $s(t) = a(t)e^{j(2\pi f_0 t + \gamma(t))}$, where $a(t)$ denotes the real amplitude, $\gamma(t)$ is the time-varying modulating shift phase and $f_0$ is the carrier frequency of the incident wave. Under this assumption, one has the following parameterized mean model given by $\boldsymbol{\chi} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\xi}_0), \boldsymbol{\Sigma}_{\text{noise}})$, and in which

$$\boldsymbol{\mu}(\boldsymbol{\xi}_0) = \boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \boldsymbol{u}(\psi_0, \rho_0)) \quad (4)$$

where $\boldsymbol{s} = [s(1), \ldots, s(T)]^T$. Consequently, the FIM in (3) reduces to

$$[\textbf{FIM}(\boldsymbol{\xi}_0)]_{i,k} = 2\mathcal{R}\left\{\frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_i}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_k}\bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right\}$$
$$\quad (5)$$
$$\forall i = 1, 2, 3, k = 1, 2, 3.$$

### B. The Unconditional Case

In the unconditional model, the signal is assumed to be complex circular Gaussian (with zero mean and variance $\sigma_s^2 \boldsymbol{I}$) independent from the noise. Consequently, the covariance parameterized observation model given by $\boldsymbol{\chi} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0))$, in which

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}_0) = \sigma_s^2(\boldsymbol{e}(\omega_0)\otimes\boldsymbol{u}(\psi_0, \rho_0))(\boldsymbol{e}(\omega_0)\otimes\boldsymbol{u}(\psi_0, \rho_0))^H + \boldsymbol{\Sigma}_{\text{noise}}$$
$$\quad (6)$$

Then, by applying (3), one obtains

$$[\mathbf{FIM}(\boldsymbol{\xi}_0)]_{i,k} = \tag{7}$$

$$Ttr\left\{\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_i}\Big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial[\boldsymbol{\xi}]_k}\Big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right\}$$

$$\forall i=1,2,3, k=1,2,3.$$

## III. DETERMINISTIC LOWER BOUNDS BACKGROUND AND DERIVATION

The unification presented in [15] gives analytical expressions of the McAulay-Seidman bound (MSB, $\boldsymbol{C}_{\text{MSB}}^{(L)}$), the Hammersley-Chapman-Robbins bound (HCRB, $\boldsymbol{C}_{\text{HCRB}}^{(L)}$), the McAulay-Hofstetter bound (MHB, $\boldsymbol{C}_{\text{MHB}}^{(L)}$) and the Todros-Tabrikian Bound (TTB, $\boldsymbol{C}_{\text{TTB}}^{(L)}$). Specifically, we can notice that:

$$\boldsymbol{C}_{\text{MSB}}^{(L)} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}^T \tag{8}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\xi}_1 - \boldsymbol{\xi}_0,\ldots,\boldsymbol{\xi}_L - \boldsymbol{\xi}_0]$ in which $\boldsymbol{\xi}_l$ denotes the $l$-th test point for $l = 1,\ldots,L$. Whereas, $[\boldsymbol{\Psi}]_{m,n} = E_{\boldsymbol{\chi}|\boldsymbol{\xi}_0}\{\nu(\boldsymbol{\chi},\boldsymbol{\xi}_m)\nu(\boldsymbol{\chi},\boldsymbol{\xi}_n)\}$ for $m = 1,\ldots,L$, $n = 1,\ldots,L$ and $\nu(\boldsymbol{\chi},\boldsymbol{\xi}_l)$ denotes the ratio-likelihood function, given by $\nu(\boldsymbol{\chi},\boldsymbol{\xi}_l) = \frac{p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)}{p(\boldsymbol{\chi}|\boldsymbol{\xi}_0)}$. Furthermore, the HCRB and the MHB are given by

$$\boldsymbol{C}_{\text{HCRB}}^{(L)} = \boldsymbol{\Phi}(\boldsymbol{\Psi} - \mathbf{1}_L\mathbf{1}_L^T)^{-1}\boldsymbol{\Phi}^T \tag{9}$$

and

$$\boldsymbol{C}_{\text{MHB}}^{(L)} = \boldsymbol{C}_{\text{CRB}} + \boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^T \tag{10}$$

where $\boldsymbol{C}_{\text{CRB}} = \mathbf{FIM}^{-1}$ and $\boldsymbol{Q} = \boldsymbol{C}_{\text{CRB}}\boldsymbol{D} - \boldsymbol{\Phi}$,

$$\boldsymbol{R} = \boldsymbol{\Psi} - \boldsymbol{D}^T\boldsymbol{C}_{\text{CRB}}\boldsymbol{D} \tag{11}$$

in which $\boldsymbol{D} = [\boldsymbol{d}(\boldsymbol{\xi}_1),\ldots,\boldsymbol{d}(\boldsymbol{\xi}_L)]$, and each element of $\boldsymbol{D}$ is given by

$$\boldsymbol{d}(\boldsymbol{\xi}_l) = \left(\frac{\partial\text{KLD}\left(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)\|p(\boldsymbol{\chi}|\boldsymbol{\xi})\right)}{\partial\boldsymbol{\xi}}\Big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\right)^T \tag{12}$$

The term $\text{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)\|p(\boldsymbol{\chi}|\boldsymbol{\xi}))$ denotes the Kullback-Leibler divergence [23] of $p(\boldsymbol{\chi}|\boldsymbol{\xi})$ from $p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)$. Finally, the TTB is given by

$$\boldsymbol{C}_{\text{TTB}}^{(L,J)} = \boldsymbol{C}_{\text{CRB}} + \boldsymbol{Q}\boldsymbol{W}^H\left(\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}^H\right)^{-1}\boldsymbol{W}\boldsymbol{Q}^T \tag{13}$$

in which, the three-dimensional-discrete-Fourier-transform (DFT) matrix is given by [15]

$$[\boldsymbol{W}]_{i,l} = e^{-j\boldsymbol{\Omega}_i^T\boldsymbol{\xi}_l} \tag{14}$$

in which $\boldsymbol{\Omega}_i$ is expressed for the $i$-th frequency test bin by

$$\boldsymbol{\Omega}_i^T = 2\pi\left[\frac{i_\omega}{\Delta_\omega L_\omega}, \frac{i_\psi}{\Delta_\psi L_\psi}, \frac{i_\rho}{\Delta_\rho L_\rho}\right],$$

in which $L_\omega, L_\psi$ and $L_\rho$ represent the number of test points w.r.t. the unknown parameters $\omega, \psi$ and $\rho$, and $L = L_\omega L_\psi L_\rho$. The uniform inter-test points w.r.t. the unknown parameters are represented by $\omega, \psi$ and $\rho$ $\Delta_\omega, \Delta_\psi$ and $\Delta_\rho$, respectively. Similarly, $i_\omega, i_\psi$ and $i_\rho$ denote the index of test-bin in the three-dimensional of the frequency domain, $i_\omega \in 1,\ldots,L_\omega$, $i_\psi \in 1,\ldots,L_\psi$ and $i_\rho \in 1,\ldots,L_\rho$. Consequently, the index $i$ is a unique combination of $i_\omega, i_\psi, i_\rho$ where the total number of these combinations is denoted by $J$.

In the remaining of this section, we derive analytical expressions of the MSB, the HCRB, the MHB and the TTB for conditional and unconditional observation model.

### A. *The Conditional Case*

Let us partition the FIM w.r.t. the signal parameter $\boldsymbol{\xi}_0$ as follow

$$\mathbf{FIM}(\boldsymbol{\xi}_0) = \begin{bmatrix} F_{\omega,\omega} & F_{\omega,\psi} & F_{\omega,\rho} \\ F_{\psi,\omega} & F_{\psi,\psi} & F_{\psi,\rho} \\ F_{\rho,\omega} & F_{\rho,\psi} & F_{\rho,\rho} \end{bmatrix} \tag{15}$$

in which, we used the notation $F_{u,v}$, where the lower script $u,v$ denotes the considered part of the **FIM** which corresponds to the derivation according to parameters $u$ and $v$ as shown in (3). Using (5), the entries of the FIM are given by

$$F_{\omega,\omega} = 2\Re[(\boldsymbol{s} \otimes (j\boldsymbol{e}'(\omega_0) \otimes \boldsymbol{u}(\psi_0,\rho_0)))^H\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s} \otimes (j\boldsymbol{e}'(\omega_0) \otimes \boldsymbol{u}(\psi_0,\rho_0)))]$$

$$F_{\psi,\psi} = 2\Re[(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi,\rho_0)}{\partial\psi}))^H|_{\psi=\psi_0}\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s}^H \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi,\rho_0)}{\partial\psi}))|_{\psi=\psi_0}]$$

and

$$F_{\rho,\rho} = 2\Re[(\boldsymbol{s}^H \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi_0,\rho)}{\partial\rho})^H)|_{\rho=\rho_0}\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi_0,\rho)}{\partial\rho}))|_{\rho=\rho_0}]$$

in which $\boldsymbol{e}'(\omega_0) = [d_1 e^{jd_1\omega_0},\ldots,d_N e^{jd_N\omega_0}]^T$. The cross terms are given by

$$F_{\psi,\rho} = F_{\rho,\psi} = 2\Re[(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi,\rho_0)}{\partial\psi}))^H|_{\psi=\psi_0}\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi_0,\rho)}{\partial\rho}))|_{\rho=\rho_0}]$$

$$F_{\omega,\rho} = F_{\rho,\omega} = 2\Re[(\boldsymbol{s} \otimes (j\boldsymbol{e}'(\omega_0) \otimes \boldsymbol{u}(\psi_0,\rho_0)))^H\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi_0,\rho)}{\partial\rho}))|_{\rho=\rho_0}]$$

and

$$F_{\omega,\psi} = F_{\psi,\omega} = 2\Re[(\boldsymbol{s} \otimes (j\boldsymbol{e}'(\omega_0) \otimes \boldsymbol{u}(\psi_0,\rho_0)))^H\boldsymbol{\Sigma}_{\text{noise}}^{-1}$$
$$(\boldsymbol{s} \otimes (\boldsymbol{e}(\omega_0) \otimes \frac{\partial\boldsymbol{u}(\psi_0,\rho)}{\partial\rho}))|_{\rho=\rho_0}]$$

In particular, if $\boldsymbol{n}_{\text{n}}(t)$ is a complex circular white Gaussian noise with zero-mean and unknown variance $\sigma_n^2$, assumed to be uncorrelated both temporally and spatially (i.e., $\boldsymbol{\Sigma}_{\text{noise}} = \sigma_n^2\boldsymbol{I}_{2NT}$). The above expressions simplifies to $F_{\omega,\omega} = \frac{2}{\sigma_n^2}||\boldsymbol{d}_N||_2^2||\boldsymbol{a}||_2^2$, $F_{\psi,\psi} = \frac{2N}{\sigma_n^2}\sin(\rho_0)^2||\boldsymbol{a}||_2^2$ and $F_{\rho,\rho} = \frac{2N}{\sigma_n^2}||\boldsymbol{a}||_2^2$. Whereas, the cross terms are given by $F_{\omega,\psi} = F_{\psi,\omega} = \frac{2}{\sigma_n^2}\sin(\rho_0)^2||\boldsymbol{d}_N||_1^2||\boldsymbol{a}||_2^2$ and $F_{\omega,\rho} = F_{\rho,\omega} = F_{\rho,\psi} = F_{\psi,\rho} = 0$ with $\boldsymbol{a} = \boldsymbol{s}$ and $\boldsymbol{d}_N = [d_1, d_2,\ldots,d_n]^T$.

Considering an identifiable situation in which $|\mathbf{FIM}| \neq 0$, one obtains after some calculus (the general case, i.e., $2\pi\frac{A_{sl}}{\lambda} \neq L_{sd}$, is given in the Appendix)

$$\mathbf{CRB}(\boldsymbol{\xi}_0) = \frac{\sigma_n^2}{2||\boldsymbol{a}||_2^2} \qquad (16)$$

$$\begin{bmatrix} \frac{N}{N||\boldsymbol{d}_N||_2^2 - \sin(\rho_0)^2||\boldsymbol{d}_N||_1^4} & -\frac{||\boldsymbol{d}_N||_1^2}{N||\boldsymbol{d}_N||_2^2 - \sin(\rho_0)^2||\boldsymbol{d}_N||_1^4} & 0 \\ -\frac{||\boldsymbol{d}_N||_1^2}{N||\boldsymbol{d}_N||_2^2 - \sin(\rho_0)^2||\boldsymbol{d}_N||_1^4} & \frac{||\boldsymbol{d}_N||_2^2}{N\sin^2(\rho_0)||\boldsymbol{d}_N||_2^2 - \sin(\rho_0)^2||\boldsymbol{d}_N||_1^4} & 0 \\ 0 & 0 & \frac{1}{N} \end{bmatrix}$$

On the other hand, in (12) the KLD for conditional case is given by

$$\mathrm{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)||p(\boldsymbol{\chi}|\boldsymbol{\xi})) = \int p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)\ln(\frac{p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)}{p(\boldsymbol{\chi}|\boldsymbol{\xi})})d\boldsymbol{\chi} \quad (17)$$

$$= (\boldsymbol{\mu}(\boldsymbol{\xi}) - \boldsymbol{\mu}(\boldsymbol{\xi}_l))^H \boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\mu}(\boldsymbol{\xi}_l) - \boldsymbol{\mu}(\boldsymbol{\xi}))$$

thus,

$$\boldsymbol{d}(\boldsymbol{\xi}_l) = -\left(\frac{\partial\mathrm{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)||p(\boldsymbol{\chi}|\boldsymbol{\xi}))}{\partial\boldsymbol{\xi}}\right)^T|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0} \qquad (18)$$

$$= -2\Re\left\{(\boldsymbol{\mu}(\boldsymbol{\xi}) - \boldsymbol{\mu}(\boldsymbol{\xi}_l))^H\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\xi})^H}{\partial\boldsymbol{\xi}}\right\}^T|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}$$

Using, (11), each element of $\boldsymbol{\Psi}$ is of the form

$$[\boldsymbol{\Psi}]_{m,n} = E_{\boldsymbol{\chi}|\boldsymbol{\xi}_0}\{\nu(\boldsymbol{\chi},\boldsymbol{\xi}_m)\nu(\boldsymbol{\chi},\boldsymbol{\xi}_n)\}$$

$$= \frac{1}{\pi^{2NT}|\boldsymbol{\Sigma}_{\mathrm{noise}}|}\int e^{\kappa_1}$$

$$= \alpha(\boldsymbol{\xi}_m,\boldsymbol{\xi}_n)\int\frac{1}{\pi^{2NT}|\boldsymbol{\Sigma}_{\mathrm{noise}}|}e^{\kappa_2}d\boldsymbol{\chi}$$

$$= \alpha(\boldsymbol{\xi}_m,\boldsymbol{\xi}_n) \qquad (19)$$

in which

$$\kappa_1 = -(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_m))^H\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_m))$$
$$- (\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_n))^H\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_n)) + (\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_0))^H$$
$$\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_0))d\boldsymbol{\chi}$$

and

$$\kappa_2 = -(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_m) - \boldsymbol{\mu}(\boldsymbol{\xi}_n) + \boldsymbol{\mu}(\boldsymbol{\xi}_0))^H$$
$$\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\chi} - \boldsymbol{\mu}(\boldsymbol{\xi}_m) - \boldsymbol{\mu}(\boldsymbol{\xi}_n) + \boldsymbol{\mu}(\boldsymbol{\xi}_0))$$

and

$$\alpha(\boldsymbol{\xi}_m,\boldsymbol{\xi}_n) = e^{2\Re\{(\boldsymbol{\mu}(\boldsymbol{\xi}_m)-\boldsymbol{\mu}(\boldsymbol{\xi}_0))^H\boldsymbol{\Sigma}_{\mathrm{noise}}^{-1}(\boldsymbol{\mu}(\boldsymbol{\xi}_n)-\boldsymbol{\mu}(\boldsymbol{\xi}_0))\}} \quad (20)$$

At last, plugging (4), (16)-(20) into (8)-(10) and (13) one obtains $\boldsymbol{C}_{\mathrm{C-MSB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{C-HCRB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{C-MHB}}^{(L)}$ and $\boldsymbol{C}_{\mathrm{C-TTB}}^{(L,J)}$, in which C stands stands for Conditional.

### B. The Unconditional Case

Let us consider the unconditional model. For simplicity we define $\boldsymbol{\Gamma}(\boldsymbol{\xi}_0) = \boldsymbol{e}(\omega_0) \otimes \boldsymbol{u}(\psi_0,\rho_0)$ and let us recall the following matrix properties $Tr(\boldsymbol{XY}) = vec(\boldsymbol{X}^H)^H vec(\boldsymbol{Y})$,

$vec(\boldsymbol{XYZ}) = (\boldsymbol{Z}^T \otimes \boldsymbol{X})vec(\boldsymbol{Y})$, which hold for any matrices $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$ [24]. Using these properties along with (7), we obtain

$$\frac{1}{T}\mathbf{FIM}(\boldsymbol{\xi}_0) = (\frac{\partial\boldsymbol{r}}{\partial\boldsymbol{\xi}^T})^H|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}(\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-T}\otimes$$

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1})(\frac{\partial\boldsymbol{r}}{\partial\boldsymbol{\xi}^T})|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}$$

$$= \begin{bmatrix} \boldsymbol{g}_\omega^H\boldsymbol{g}_\omega & \boldsymbol{g}_\omega^H\boldsymbol{g}_\psi & \boldsymbol{g}_\omega^H\boldsymbol{g}_\rho \\ \boldsymbol{g}_\psi^H\boldsymbol{g}_\omega & \boldsymbol{g}_\psi^H\boldsymbol{g}_\psi & \boldsymbol{g}_\psi^H\boldsymbol{g}_\rho \\ \boldsymbol{g}_\rho^H\boldsymbol{g}_\omega & \boldsymbol{g}_\rho^H\boldsymbol{g}_\psi & \boldsymbol{g}_\rho^H\boldsymbol{g}_\rho \end{bmatrix}$$

where

$$\boldsymbol{r} = vec(\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)) = \sigma_s^2(\boldsymbol{\Gamma}^c(\boldsymbol{\xi}_0) \otimes \boldsymbol{\Gamma}(\boldsymbol{\xi}_0)) + vec(\boldsymbol{\Sigma}_{\mathrm{noise}})$$

and

$$\boldsymbol{g}_\omega = vec\left(\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-\frac{1}{2}}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial\omega}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-\frac{1}{2}}\right).$$

If the sensor noise is both spatially and temporally white, as $\boldsymbol{\Sigma}_{\mathrm{noise}} = \sigma_n^2\boldsymbol{I}_{2N}$, using the matrix inversion lemma [25] into (6), one can obtain

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1} = \frac{1}{\sigma_n^2}\boldsymbol{I}_{2N} - \frac{\sigma_s^2}{\sigma_n^4 + N\sigma_s^2\sigma_n^2}\boldsymbol{\Gamma}(\boldsymbol{\xi}_0)\boldsymbol{\Gamma}^H(\boldsymbol{\xi}_0)$$

On the other hand, in (12) the KLD is obtained $\mathrm{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)||p(\boldsymbol{\chi}|\boldsymbol{\xi})) = E_{\boldsymbol{\chi}|\boldsymbol{\xi}_l}\{\boldsymbol{\chi}^H(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}))^{-1}\boldsymbol{\chi}\} - E_{\boldsymbol{\chi}|\boldsymbol{\xi}_l}\{\boldsymbol{\chi}^H(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}_l))^{-1}\boldsymbol{\chi}\} + T\ln|\boldsymbol{\Sigma}(\boldsymbol{\xi})| - T\ln|\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)|$ in which $E_{\boldsymbol{\chi}|\boldsymbol{\xi}_l}\{\boldsymbol{\chi}^H(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}))^{-1}\boldsymbol{\chi}\} = \sum_{i=1}^{2NT}\sum_{j=1}^{2NT}E_{\boldsymbol{\chi}|\boldsymbol{\xi}_l}\{[\boldsymbol{\chi}]_i^c[\boldsymbol{\chi}]_j^c[(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}))^{-1}]_{i,j}\} = Ttr\{\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}\}$ and similarly, one obtains $E_{\boldsymbol{\chi}|\boldsymbol{\xi}_l}\{\boldsymbol{\chi}^H\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)^{-1}\boldsymbol{\xi}\} = 2NT$. Consequently,

$$\mathrm{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)||p(\boldsymbol{\chi}|\boldsymbol{\xi})) = Ttr(\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}) - 2NT \quad (21)$$
$$+ T\ln(\frac{|\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)|}{|\boldsymbol{\Sigma}(\boldsymbol{\xi})|})$$

In addition, the $l$-th member of $\boldsymbol{D}$ in (12) is given by

$$\frac{\partial\mathrm{KLD}(p(\boldsymbol{\chi}|\boldsymbol{\xi}_l)||p(\boldsymbol{\chi}|\boldsymbol{\xi}))}{\partial\boldsymbol{\xi}} = Ttr\left\{\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}\right\}$$

$$+ Ttr\left\{-\boldsymbol{\Sigma}(\boldsymbol{\xi}_l)\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}\right\} \qquad (22)$$

and from (11), the element of $\boldsymbol{\Psi}$ is

$$[\boldsymbol{\Psi}]_{m,n} = E_{\boldsymbol{\chi}|\boldsymbol{\xi}_0}\{\nu(\boldsymbol{\chi},\boldsymbol{\xi}_m)\nu(\boldsymbol{\chi},\boldsymbol{\xi}_n)\}$$

$$= \frac{|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)|^2}{|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}_m)||\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}_n)|}\int e^{\kappa_3}p(\boldsymbol{\chi}|\boldsymbol{\xi}_0)d\boldsymbol{\chi}$$

$$= \frac{|\boldsymbol{\Sigma}(\boldsymbol{\xi}_0)||\boldsymbol{\Sigma}(\boldsymbol{\xi}_m)^{-1} + \boldsymbol{\Sigma}(\boldsymbol{\xi}_n)^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)^{-1}|}{|\boldsymbol{\Sigma}(\boldsymbol{\xi}_m)||\boldsymbol{\Sigma}(\boldsymbol{\xi}_n)|} \quad (23)$$

with

$$\kappa_3 = -\boldsymbol{\chi}^H((\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}\boldsymbol{\xi}_m))^{-1} + (\boldsymbol{I}_T\otimes$$
$$\boldsymbol{\Sigma}(\boldsymbol{\xi}_n))^{-1} - 2(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}(\boldsymbol{\xi}_0))^{-1})\boldsymbol{\chi}$$

Finally, $\boldsymbol{C}_{\mathrm{U-MSB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{U-HCRB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{U-MHB}}^{(L)}$, and $\boldsymbol{C}_{\mathrm{U-TTB}}^{(L,J)}$ are given by replacing (6), (21) - (23) into (8), (9), (10) and (13), in which U stands stands for Unconditional.

## IV. NUMERICAL INVESTIGATION

Numerical results are presented in this section for a non-uniform linear COLD array with $N = 8$ sensors. One narrowband far-field source is located according to $\theta_0 = 60°$ and the polarization state parameters are given by $\rho_0 = 30°$ and $\psi_0 = 45°$. Simulations are performed for $T = 15$ snapshots.

### A. Analytical and numerical analysis of the derived lower bounds

The aim of this part is to examine the usefulness of $\boldsymbol{C}_{\mathrm{MSB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{HCRB}}^{(L)}$, $\boldsymbol{C}_{\mathrm{MHB}}^{(L)}$ and $\boldsymbol{C}_{\mathrm{TTB}}^{(L,J)}$ to predict the SNR threshold. Using, (10), (11), (13) and (17), we plot the derived lower bounds for the parameter $\omega_0$ for both conditional and unconditional models in Fig.2 and Fig.3, respectively. These figures show that the derived bounds exhibit a threshold effect around $-9$dB. We note also that the new proposed TTB provides a better prediction of the SNR threshold as expected. Furthermore, for $J < L$, the computational cost of the TTB is lower in comparison to the MSB, HCRB and MHB. This is mainly due to the inversion of a $J \times J$ matrix due to presence of the discrete transform matrix $\boldsymbol{W}$ instead of the initial $L \times L$ matrix inversion.

In Fig.4 and 5, we plot the TTB w.r.t. $\omega$ vs. SNR for different polarization parameters $\psi_0$ and $\rho_0$ in the conditonal case (the same conclusion are noticed for the unconditional case). From Fig.4, we notice that, no matter how $\psi_0$ changes, the breakdown point is approximately fixed for a given $\omega_0$, meaning that the effect of the polarisation state parameter $\psi$ can be neglected in designing the COLD array. In Fig.5 we focus on the effect of $\rho$, in which, this figure shows that, for a different value of polarization parameter $\rho$ and for a fixed SNR, the higher the $\rho_0$ is, the better are the performances (w.r.t. the MSE in the asymptotic region but also breakdown point).

### B. Frequency test-bins and their effect on the TTB

One notes that increasing the number of sensors or test-points may improve the tightness of the bounds. This is also same for frequency-bins. However, this improvement comes at the expense of computational complexity. Considering there is an intuitive link between the performance of threshold prediction and the sequence of frequency test-bins. We give a sub-optimal method to design an optimal index of frequency test-bins in order to overcome an exhaustive search by computer over all possibilities. First, place only one frequency test-bin by minimizing the threshold SNR with respect to $J$ positions. Second, iterate the first step by placing the $n$-th frequency test-bin at once sequentially until $n = L$ with respect to $J - n + 1$ remaining positions. To illustrate and compare the accuracy and usefulness of the aforementioned approach, two numerical examples for TTB are obtained: case 1- With the same number of test-points $L = 32$, one notice that our sub-optimal method yields a very good agreement when compared with the searching solution performed by computer, as shown in Fig.6. Furthermore, this figure also shows us that increasing the number of frequency test-bins gives a considerable improvement in the SNR threshold prediction; case 2- In Fig.7, with the same number of test-points and frequency test-bins $L = J = 32$, the threshold prediction of optimal index is nearly 1dB better than the sequential one, which means that the index of frequency test-bins has an important effect on minimizing the ambiguity region threshold.
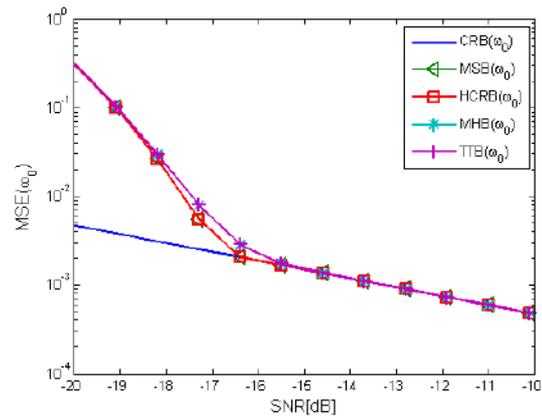


**Fig. 2**: Lower bounds on the mean square error (conditional case) w.r.t. $\omega$ for NULA-COLD array ($L = J = 8$).

As the number of snapshots $T$ increases, this advantage will be more apparent. Consequently, we notice that the proposed method with low complexity is useful for a threshold prediction problem, especially for the sensor arrays with large test-points.

### C. Designing a COLD array : Resolution factor vs threshold SNR

To better comprehend the system performance w.r.t the resolution factor (RF) and the threshold SNR as a function of array geometry we consider Fig.8 in which we compare all possible configuration for $N = 8$ and a given aperture $A = 23$ (more precisely, such context gives 74613 possible sensor array configurations). The term resolution factor is the minus curvature of the magnitude squared of the beampattern at the peak of its mainlobe and is determined by the array geometry [26] as:

$$RF = \frac{8d^2\pi^2}{N^2\lambda^2}\boldsymbol{d}_N^T(N\boldsymbol{I}_N - \boldsymbol{1}_N)\boldsymbol{d}_N$$

in which $\boldsymbol{d}_N = [d_1, \ldots, d_N]^T$ denotes the $N \times 1$ vector of array element index locations.

In general, we would like the threshold SNR to be as small as possible and the RF to be as large as possible [26]. We see that the best performance is achieved at the upper-left corner of the distribution given in Fig.8, where there is a tradeoff between resolution and threshold prediction.

As an example, eight types of array configuration are considered, as shown in Table 1. From comparison of simulation results, one can further notice that: 1) For the same array aperture and same number of sensors, the SNR threshold prediction and resolution factor are greatly affected by the array geometric configuration.

2) The configuration that puts two sensors at the extremity and the rest in the middle has the worst performance for resolution factor; Contrary, places two average parts of sensors on both sides seems to be the best geometry configuration (at least, it is the case for $N = 8$ and $A = 23$ sensors). From example, Type 1 to Type 4 arrays, which have the good performance for the threshold SNR, are nearly 3dB better than the relatively poor performance of type 7 and type 8, where this is due to the sensor geometry configuration.

**Fig. 3**: Lower bounds on the mean square error (unconditional case) w.r.t. $\omega$ for NULA-COLD array ($L = J = 8$).
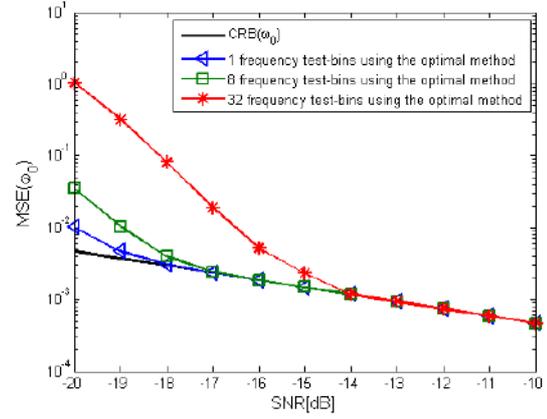


**Fig. 6**: Comparisons of different number of frequency test-bins for TTB threshold prediction with test-points $L = 32$.



**Fig. 4**: TTB on the mean square error (conditional case) w.r.t. $\psi$ for NULA-COLD array. ($L = J = 8$)



**Fig. 7**: Comparisons of different index of frequency test-bins for TTB threshold prediction with test-points $L = 32$, frequency test-bins $J = 32$.
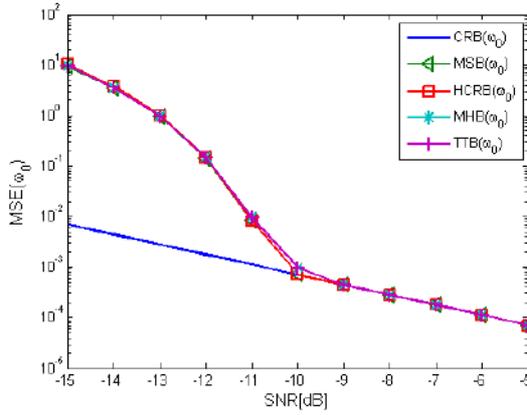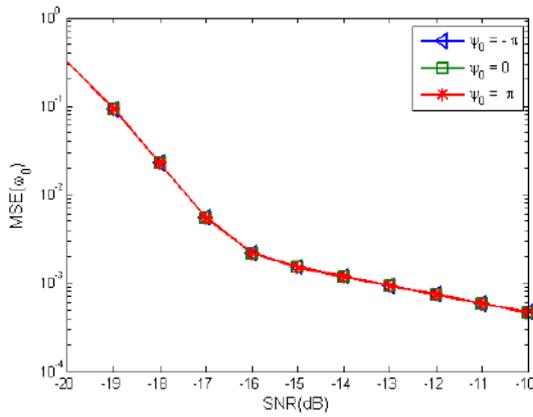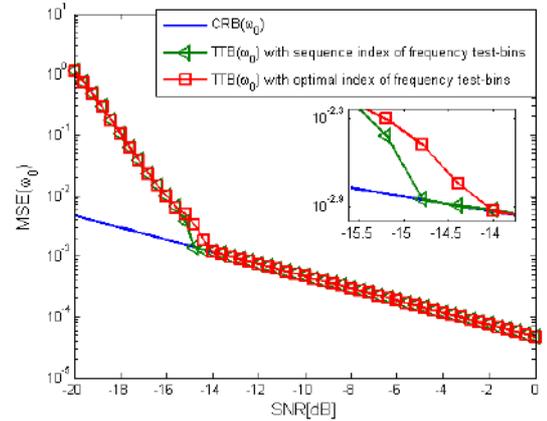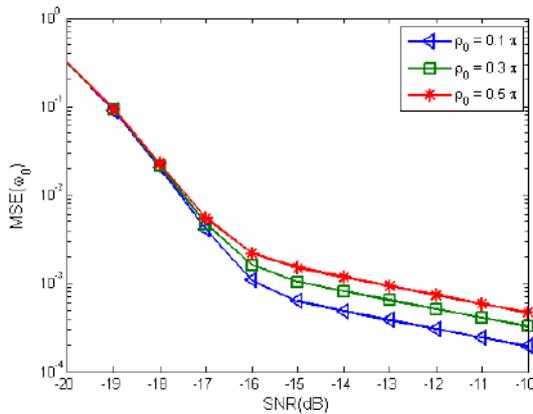


**Fig. 5**: TTB on the mean square error (conditionalc case) w.r.t. $\rho$ for NULA-COLD array. ($L = J = 8$)
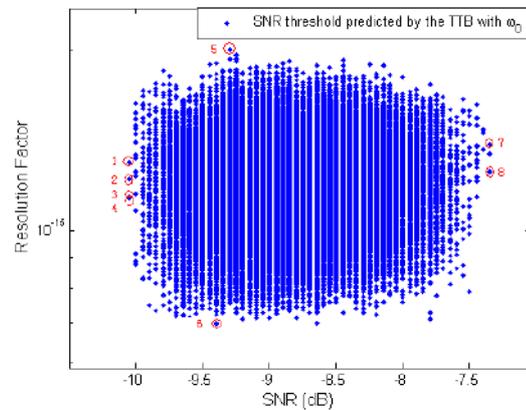


**Fig. 8**: Comparisons of different array geometric configurations for resolution factor w.r.t. SNR threshold predicted by the TTB (conditional case: $N = 8$, $A = 23$).

| Array Type | Geometric Configuration | Resolution Factor | Threshold SNR |
|---|---|---|---|
| Type 1 | ●●●○○○●○○○●○○○●○●●○○○○○●○● | | among the best configuration |
| Type 2 | ●○●●○○○○○●○●○○●○○○○●○○● | | among the best configuration |
| Type 3 | ●○●○●○○○○○●○●○○●○○●○○○● | the best configuration | |
| Type 4 | ●●○○○●○●○●○○○●○●○●○○○○● | | among the best configuration |
| Type 5 | ●●●●○●○○○○○○○○○○○○○○●●●● | the best configuration | |
| Type 6 | ●○○○○○○○●●●●●●○○○○○○○○○● | the worst configuration | |
| Type 7 | ●●○○○○○●○○○○●○●○●○●○○○●● | | among the worst configuration |
| Type 8 | ●○●○○○●○○○●○●○●○●○○○●● | | among the worst configuration |

**Table 1.** Different array geometric configurations among the 74613 possibilities for $N = 8$ and $A = 23$ (● and ○ represent the position of sensor and missing sensors, respectively.)

## V. CONCLUSION

In this paper, we derive explicit closed-form expressions of different deterministic lower bounds on the mean square error for the so-called non-uniform linear co-centered orthogonal loop and dipole arrays in a passive polarization source localization context. Taking advantage of these expressions, we analyse and characterize the performances in the asymptotic region and non-asymptotic region in terms of breakdown point prediction for the conditional and unconditional observation models. Finally, numerical simulations show the effect of each polarization parameters and the array geometry on the optimal designing strategy.

## VI. APPENDIX

The general case of the CRB, i.e., in the case of $2\pi \frac{A_{sl}}{\lambda} = L_{sd}$ is given as

$$\mathbf{CRB}(\boldsymbol{\xi}) = \frac{1}{D} \begin{bmatrix} F_{\psi,\psi} & -F_{\psi,\omega} & 0 \\ -F_{\omega,\psi} & F_{\omega,\omega} & 0 \\ 0 & 0 & \frac{D}{F_{\rho,\rho}} \end{bmatrix}$$

in which

$$D = F_{\omega,\omega} F_{\psi,\psi} - F_{\psi,\omega} F_{\omega,\psi}$$

$$F_{\omega,\omega} = \frac{2}{\sigma_n^2} ||\boldsymbol{s}||_2^2 ||\boldsymbol{d}_N||_2^2 (4\pi^2 \frac{A_{sl}^2}{\lambda^2} \cos^2(\rho) + L_{sd}^2 \sin^2(\rho))$$

$$F_{\rho,\rho} = \frac{2N}{\sigma_n^2} ||\boldsymbol{s}||_2^2 (4\pi^2 \frac{A_{sl}^2}{\lambda^2} \sin^2(\rho) + L_{sd}^2 \cos^2(\rho))$$

$$F_{\psi,\psi} = \frac{2N}{\sigma_n^2} ||\boldsymbol{s}||_2^2 L_{sd}^2 \sin^2(\rho)$$

$$F_{\psi,\omega} = F_{\omega,\psi} = \frac{2}{\sigma_n^2} ||\boldsymbol{s}||_2^2 ||\boldsymbol{d}_N||_1^2 L_{sd}^2 \sin^2(\rho)$$

## REFERENCES

[1] T. Bao, A. Breloy, M. N. El Korso, K. Abed-Meraim and H. Ouslimani , "Performance analysis of direction-of-arrival and polarization estimation using a non-uniform linear cold array," in *In 2017 Seminar on Detection Systems Architectures and Technologies (DAT), IEEE*, vol. 1, 2017, pp. 1–5. https://doi.org/10.1109/DAT.2017.7889165

[2] M. Haardt, M. Pesavento, F. Roemer, and M. N. E. Korso, "Subspace methods and exploitation of special array structures," Electronic Reference in Signal Processing: Array and Statistical Signal Processing (M. Viberg, ed.), vol. 3, pp. 651-717, Academic Press Library in Signal Processing, Elsevier Ltd., Chapter 2.15, ISBN 978-0-12-411597-2, 2014.

[3] H. L. VanTrees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, vol. 1.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Mag.*, vol. 13, no. 4, pp. 67–94, 1996. https://doi.org/10.1109/79.526899

[5] J. Li, P. Stoica, and D. Zheng, "Efficient direction and polarization estimation with a cold array," *IEEE Trans. Antennas Propagat.*, vol. 44, no. 4, pp. 539–547, Apr. 1996. https://doi.org/10.1109/8.489306

[6] J. Li and R. Compton, "Angle and polarization estimation using esprit with a polarizationsensitive array," *IEEE Trans. Antennas Propagat.*, vol. 39, no. 9, pp. 1376–1383, Sep. 1991. https://doi.org/10.1109/8.99047

[7] T. Bao, M. N. El Korso and H. Ouslimani, "Cramer-Rao bound and statistical resolution limit investigation for near-field source localization," in *Digital Signal Processing, Elsevier*, vol. 48, No. 1, 2016, p. 137–147. https://doi.org/10.1016/j.dsp.2015.09.019

[8] R. Boyer and S. Miron, "Study of the asymptotic cramer-rao bound for the cold uniform linear array," in *Physical Communication, Elsevier*, 1982. https://doi.org/10.1016/j.phycom.2012.02.004

[9] R. Boyer, "Analysis of the COLD uniform linear array," in *Proc. IEEE Int. Work. Signal Processing, Wireless Communications*, Perugia, Italy, 2009. https://doi.org/10.1109/SPAWC.2009.5161848

[10] J. Tabrikian, R. Shavit, and D. Rahamim, "An efficient vector sensor configuration for source localization," *IEEE Signal Process. Letters*, vol. 11, pp. 690–693, Aug. 2004. https://doi.org/10.1109/LSP.2004.831682

[11] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Statistical resolution limit of the uniform linear cocentered orthogonal loop and dipole array," *IEEE Trans. Signal Processing*, vol. 59, no. 1, pp. 425–431, Jan. 2011. https://doi.org/10.1109/TSP.2010.2083657

[12] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal, "A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators," *IEEE Trans. Signal Processing*, vol. 56, no. 11, pp. 5319–5333, Nov. 2008. https://doi.org/10.1109/TSP.2008.927805

[13] E. Chaumette, A. Renaux, and M. N. El Korso, "A class of weiss-weinstein bounds and its relationship with the bobrovsky-mayer-wolf-zakai bounds," in *IEEE Transactions on Information Theory*, vol. 63, No. 4, 2017, pp. 2226–2240. https://doi.org/10.1109/TIT.2017.2671883

[14] P. Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, Orlando, Fl, 2002. https://doi.org/10.1109/ICASSP.2002.5744000

[15] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation Part I: Non-bayesian bounds for unbiased estimators," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5045–5063, Oct. 2010. https://doi.org/10.1109/TIT.2010.2059850

[16] M. N. E. Korso, A. Renaux, R. Boyer, and S. Marcos, "Deterministic performance bounds on the mean square error for near field source localization," *IEEE Trans. Signal Process.*, vol. 61, pp. 871–877, Apr. 2013. https://doi.org/10.1109/TSP.2012.2229990

[17] S. T. Smith, "Statistical resolution limits and the complexified Cramér Rao bound," *IEEE Trans. Signal Processing*, vol. 53, pp. 1597–1609, May 2005. https://doi.org/10.1109/TSP.2005.845426

[18] J.-P. Delmas, M. N. El Korso, H. Gazzah and M. Castella, "Crb analysis of planar antenna array for optimizing near-field source localization," in *Signal Processing Journal, Elsevier*, vol. 127, No. 1, 2016, p. 117–134. https://doi.org/10.1016/j.sigpro.2016.02.021

[19] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Statistical resolution limit for the multidimensional harmonic retrieval model: Hypothesis test and Cramér-Rao bound approaches," *EURASIP Journal on Advances in Signal Processing*, vol. special issue on Advances in Angle-of-Arrival and Multidimensional Signal Processing for Localization and Communications, pp. 1–14, Jul. 2011. https://doi.org/10.1186/1687-6180-2011-12

[20] H. Cramér, *Mathematical Methods of Statistics*. New York: Princeton University, Press, 1946.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing : Estimation Theory*. NJ: Prentice Hall, 1993, vol. 1.

[22] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir and A. Renaux, "Slepian-bangs-type formulas and the related misspecified cramér-rao bounds for complex elliptically symmetric distributions," in *Elsevier Signal Processing Journal*, vol. 142, 2018, pp. 320–329. https://doi.org/10.1016/j.sigpro.2017.07.029

[23] S. Kullback and R. A. Leibler, "On information and sufficiency," in *Ann. Math. Stat.*, vol. 22, No. 1, 1951, pp. 79–86. https://doi.org/10.1214/aoms/1177729694

[24] A. Graham, "Kronecker products and matrix calculus with applications," in *Wiley, New York*, 1982.

[25] G. H. Golub and C. F. V. Loan, *Matrix Computations*. London: Johns Hopkins, 1989.

[26] K. Bell, Y. Ephraim, and H. L. V. Trees, "Ziv Zakai lower bounds in bearing estimation," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, Detroit, Michigan, 1995, pp. 2852–2855. https://doi.org/10.1109/ICASSP.1995.479888

**Tao Bao** was born in Bayannur, Inner Mongolia Autonomous Region, China, in 1983. She received the B.S., M.S. and Ph.D. degrees in information and communication engineering from Northwestern Polytechnical University, China, in 2005, 2008, and 2010, respectively. From 2010, she has been with the provincial electronics experiment teaching demonstration center, Northwestern Polytechnical University, where she is currently a deputy director. From 2013 to 2014, she was a visiting scholar at the Paris Ouest Nanterre La Defense University. Her major research interests include statistical signal processing, array signal processing and space–time adaptive processing in wireless communications.

**Mohammed Nabil El Korso** was born in Oran, Algeria. He received the M.Sc. in Electrical Engineering from the National Polytechnic School, Algeria in 2007. He obtained the Master Research degree in Signal and Image Processing from Paris-Sud XI University/Supelec, France in 2008. In, 2011, he obtained his Ph.D. degree from Paris-Sud XI University. Between 2011 and 2012, he was a research scientist in the Communication Systems Group at Technische Universitat Darmstadt, Germany. He was a temporary assistant professor at l'école normale supérieure de Cachan, between 2012 and 2013. Currently, he is assistant professor at University of Paris Ouest Nanterre la Défense and a member of LEME (EA4416) laboratory. His research interests include statistical signal processing, estimation/detection theory with applications to array signal processing.

# 2-D Steady-State Heat Transfer Prediction in Rotating Electrical Machines Taking into account Materials Anisotropy: Thermal Resistances Network, Exact Analytical and Hybrid Methods

Kamel Boughrara and Frédéric Dubas

**Abstract**− **This paper presents two-dimensional (2-D) thermal resistances network (TRNM), exact analytical (AM) and hybrid (HM) methods for calculating steady-state temperature and heat flux distribution in rotating electrical machines considering materials anisotropy (i.e., different thermal conductivities in both directions). They are based on the thermal equivalent circuit (TEC), the improved exact subdomain (SD) technique where the solution and thermal conductivities depend on both directions ($r$, $\theta$) and the coupling between the two methods. TRNM is known as a semi-analytical method that can predict the heat transfer in the machine in less time than finite element method (Fem). The implementation of TRNM by considering the difference between the thermal conductivities in ($r$, $\theta$) using its equivalence with Fem is presented. The SD technique is improved to consider the difference between thermal conductivities in the directions ($r$, $\theta$). It is known that the SD technique with non-homogeneous boundary conditions (BCs) is very sensitive to the dimensions of SDs where the harmonics number and the accuracy are lower in small subdomains. Hence, the HM from the TRNM and AM is given to answer these inaccuracies especially in electrical machines with a high number of stator slots and rotor poles. The heat sources are volumetric power losses due to hysteresis, eddy-current, Joule losses and windage losses in all the regions of the machine obtained by a simplified method. The studied problem is conductive with conductive interface conditions (ICs) and convective heat transfer between the machine and the external air and at the rotor internal air. The semi-analytical results are compared between them as well as with those obtained by Fem.**

**Keywords**−**Anisotropic materials, conductive heat transfer, convection, exact subdomain technique, thermal resistances network.**

| TRNM | Thermal Resistances Network Method. |
| AM | Analytical Method. |
| HM | Hybrid Method. |
| TEC | Thermal Equivalent Circuit. |
| Fem | Finite element method. |
| SD | Subdomain. |
| BCs | Boundary Conditions. |
| PM | Permanent Magnet. |
| PDEs | Partial Differential Equations. |
| ICs | Interface Conditions. |

## I. INTRODUCTION

Thermal modeling is used to design the insulation system of electrical machines. Currently, the Fem and TEC are the most used methods [1]-[6]. Some of them take into account materials

anisotropy especially in the *z*-direction with a three-dimensional (3-D) study. Recently, Boughrara *et al.* (2018) [7] introduced a new 2-D exact SD technique able to predict steady-state heat transfer in rotating electrical machines without considering the anisotropy of thermal conductivities. This model is based on the Dubas' superposition technique [8]-[9] developed for the prediction of the magnetic field in air- or iron-cored coil. This method is very accurate and can be used for different topologies of synchronous and asynchronous machines.

The thermal modeling of electrical machines using TEC in steady-state and/or transient is fast with acceptable accuracy compared to Fem and AM, especially when the number of nodes and thermal resistances is low. TRNM is more accurate than TEC with higher time consumption [10]-[11]. However, in reality, it is not widely used for thermal design. The numerical thermal model is used in a second stage of the design to verify the temperature distribution given by semi-analytical methods. In various methods, the power losses are used as heat sources or coupled directly to electromagnetic analysis. Fem is also used with computational fluid dynamics to model convective problems inside the electrical machine and the different type of cooling. Recently, there are a few references that used an HM. TRNM of the stator and rotor are coupled to an exact AM only in the air-gap [10]-[11].

In this paper, we present three semi-analytical methods, viz., i) TRNM, ii) AM and iii) HM. TRNM is considered as Fem where the steps of meshing, materials definition with different thermal conductivities in the directions ($r$, $\theta$), elementary matrix/vector and introduction of BCs exist. The model presented in [7], based on the exact SD technique, is improved to consider the materials anisotropy in the directions ($r$, $\theta$) for the prediction of

steady-state temperature and heat flux in rotating electrical machines. AM is suitable for the SDs with high dimensions where the harmonics number can be high to achieve very good accuracy.
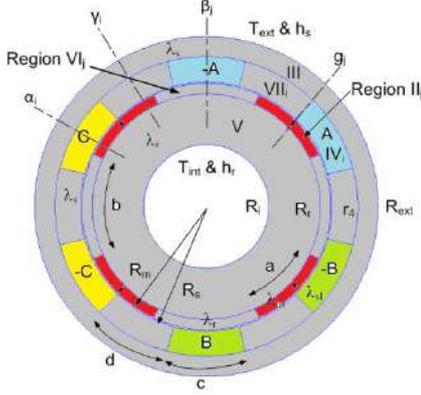


**Fig. 1:** Studied inset-PM machine [7].

This is the case of electrical machines with a low number of stator slots and rotor poles. In electrical machines with a high number of poles and stator slots, it is interesting to use the HM (i.e., the coupling between TRNM and AM) to simplify the thermal model and improve the accuracy. The used HM also considers the materials anisotropy. TRNM is used to model the stator slots and teeth SDs and rotor permanent-magnets (PMs) and teeth SDs while the AM is used to model the stator, rotor and air-gap. The coupling between TRNM and AM is achieved by using the discrete Fourier series at the ICs to satisfy the temperature continuity [12].

The developed semi-analytical methods are used to determine the heat transfer in inset-PM machines. Although, it is valid for most rotating electrical machines [7].

To determine the heat sources, a simple method is used in this paper to determine the power losses for the studied inset-PM machine [13]. Although, AM and magnetic resistances network method (MRNM) can be used [14]-[16]. The semi-analytical results are compared between them and with those obtained by Fem [17].

## II. TEMPERATURE CALCULATION METHODS

The steady-state heat transfer studied in this paper is conductive with volumetric power sources and convective at the ambient air and the rotor shaft. The model is adopted with the following assumptions:

- The materials are considered anisotropic having different constant thermal conductivities in both directions without any variation with temperature;
- The stator and rotor slots have radial sides;
- The heat sources are volumetric, uniform and constant in each SD;
- The radiation outside stator and inside the rotor is ignored;
- The interfaces between all regions are considered perfect without any contact resistance.

The analyzed inset-PM machine has 6-slots/4-poles [see Fig. 1].

### A. Thermal Resistance Network Method (TRNM)

The studied inset-PM machine is modeled using TRNM with the following steps:

*1- Meshing*

The step of meshing is performed in the same way as in Fem with the use of circular elements having a node in the middle of each element. First, we define $ms+1$ radii $wrs(i)$ with $i$ vary from 1 to $ms+1$ and $ns+1$ angles $wt(j)$ with $j$ vary from 1 to $ns+1$. The number of nodes which is equal to the number of elements is $ms \times n1$. The numbering and coordinates of nodes are given by

*For i from 1 to ms do*
*For j from 1 to ns do*
$$nods(i,j) = j+(i-1)\cdot ns ; \qquad (1)$$
$$l = nods(i,j);$$
$$x(l) = \frac{wrs(i+1)+wrs(i)}{2}\cos\left(\frac{wt(j+1)+wt(j)}{2}\right); \quad (2)$$
$$y(l) = \frac{wrs(i+1)+wrs(i)}{2}\sin\left(\frac{wt(j+1)+wt(j)}{2}\right); \quad (3)$$
*end do*
*end do*

In Fig. 2, we can show an example of mesh with 32 elements and nodes (i.e., $ms=4$ and $ns=8$). In this example, each element has a thickness of $wrs(i+1)-wrs(i)$ and opening of $wt(j+1)-wt(j)$. This example is introduced for clarity in the implementation of TRNM:

$$wrs = 10^{-3}.\begin{bmatrix} 30 & 60 & 90 & 120 & 150 \end{bmatrix} \qquad (4)$$

$$wt = \begin{bmatrix} 0 & \dfrac{\pi}{4} & \dfrac{\pi}{2} & 3.\dfrac{\pi}{4} & \pi & 5.\dfrac{\pi}{4} & 3.\dfrac{\pi}{2} & 7.\dfrac{\pi}{4} & 2\pi \end{bmatrix} \quad (5)$$

As it can be shown in (1), the numbering is done starting with 1 in the *θ*-direction than after in the *r*-direction. From the two first radii, we have 8 elements and 8 nodes numbered from 1 to 8 from right to left. The next two radii are numbered from 9 to 16 and the same for the other radii.

*2- Connectivity Matrix*

From Figs. 2 ~ 4, we can show that the node 9 is connected to 4 nodes (10, 1, 16, 17). This is the case for the entire mesh with elements having 4 thermal resistances. Supplementary nodes are added to the first and last radii to represent the BC between the stator and the external air and the rotor with the shaft. The connection between the nodes permits to obtain a matrix called connectivity matrix $isks(5,k)$ as in Fem. For internal elements that are not situated in the boundaries, we define the matrix $isks$ as

*For i from 2 to ms−1 do*
*For j from 2 to ns−1 do*
$$k = nods(i,j);$$
$$isks(1,k) = nods(i,j-1); \qquad (6)$$
$$isks(2,k) = nods(i,j); \qquad (7)$$
$$isks(3,k) = nods(i,j+1); \qquad (8)$$
$$isks(4,k) = nods(i-1,j); \qquad (9)$$
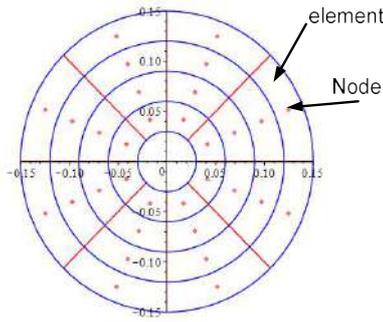$$isks(5,k) = nods(i+1,j); \qquad (10)$$
*end do*
*end do*

**Fig. 2:** Example of meshing with circular elements.



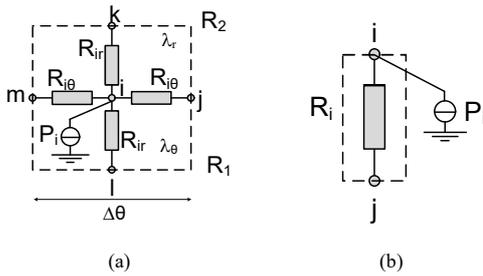(a)                                         (b)

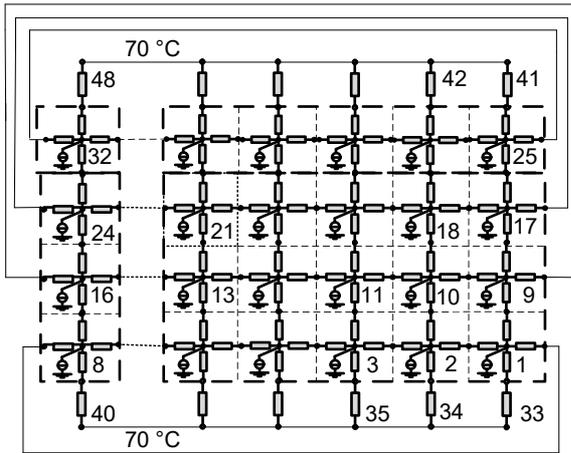**Fig. 3:** Thermal resistances of an element.



**Fig. 4:** TRNM representation of the example mesh.

Boundary elements have some nodes without connectivity and the connectivity value is zero. Generally, two types of elements can be used in TRNM and can have 1, 2, 3, 4 or 5 nodes [see Figs. 3(a) and (b)]. The elements can have 4, 3, 2 or 1 thermal resistance.

The thermal resistance for elements with 1 resistance can be radial or tangential. However, for elements with 4 thermal resistances, we have 2 radial and 2 tangential resistances. The element shown in Fig. 3 is used to represent thermal conduction [see Fig. 3(a)] and convection [see Fig. 3(b)] respectively. In this paper, the internal elements with a node in the center and 4 thermal resistances are used to represent the thermal conduction by

$$R_{ir} = \frac{1}{2} \frac{\ln\left(\frac{R_2}{R_1}\right)}{\lambda_r \Delta\theta L_u} = \frac{1}{2} \frac{\ln\left(\frac{wrs(i+1)}{wrs(i)}\right)}{\lambda_r \left(wt(j+1) - wt(j)\right) L_u} \quad (11)$$

$$R_{i\theta} = \frac{1}{2} \frac{\Delta\theta}{\lambda_\theta L_u \ln\left(\frac{R_2}{R_1}\right)} = \frac{1}{2} \frac{wt(j+1) - wt(j)}{\lambda_\theta L_u \ln\left(\frac{wrs(i+1)}{wrs(i)}\right)} \quad (12)$$

where $\lambda_r$ and $\lambda_\theta$ are the thermal conductivity in the $r$- and $\theta$-direction; $R_1$, $R_2$ and $\Delta\theta$ are respectively the internal, the external radii and the opening of element; and $L_u$ the axial length of the machine.

The conductivity in anisotropic material is a tensor with

$$[\lambda] = \begin{bmatrix} \lambda_r & 0 \\ 0 & \lambda_\theta \end{bmatrix} \quad (13)$$

It is interesting to note that the conductivities $\lambda_{r\theta}$ and $\lambda_{\theta r}$ are considered null for the studied machine. In polar coordinates, the thermal conductivity tensor $[\lambda]$ can be obtained from the thermal conductivity tensor in Cartesian coordinates as [18]

$$\begin{bmatrix} \lambda_{rr} & \lambda_{r\theta} \\ \lambda_{r\theta} & \lambda_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_{xx} & \lambda_{xy} \\ \lambda_{xy} & \lambda_{yy} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (14)$$

It is important to note that Fem uses the Cartesian coordinate's tensor for the thermal conductivity [17] and [19]. For this, the validation of the results considering the materials anisotropy in polar coordinates is done with TRNM and AM.

The boundary elements [see Fig. 4] with 1 node at fixed temperature (viz., 70°C) represent the thermal convection resistance as follows

$$R_{rk} = \frac{1}{h_s L_u \left(wt(j+1) - wt(j)\right) wrs(ms+1)} \quad (15)$$

$$R_{rk} = \frac{1}{h_r L_u \left(wt(j+1) - wt(j)\right) wrs(1)} \quad (16)$$

where $h_s$ and $h_r$ are respectively the convection coefficients at the external radius and at the shaft of the machine.

In the proposed method, the Dirichlet's condition (without convection) can be introduced by setting the thermal resistance of convection to zero.

For the mesh example in Fig. 4, there are 48 equations corresponding to 48 elements and nodes. Each conductive element has an internal volumetric source of heat $P_i$ and connected to four nodes. All elements can be represented similarly to equations of elements 1 and 2 by

$$P_1 = \frac{T_1 - T_8}{R_{1\theta} + R_{8\theta}} + \frac{T_1 - T_2}{R_{1\theta} + R_{2\theta}} + \frac{T_1 - T_{33}}{R_{1r} + R_{33}} + \frac{T_1 - T_9}{R_{1r} + R_{9r}} \quad (17)$$

$$P_2 = \frac{T_2 - T_1}{R_{2\theta} + R_{1\theta}} + \frac{T_2 - T_3}{R_{2\theta} + R_{3\theta}} + \frac{T_2 - T_{34}}{R_{2r} + R_{34}} + \frac{T_2 - T_{10}}{R_{2r} + R_{10r}} \quad (18)$$

The 16 equations of convection at boundaries are similar to those of elements 41 and 34, e.g.,

$$0 = \frac{T_{41} - T_{25}}{R_{25r} + R_{41}} \qquad (19)$$

$$0 = \frac{T_{34} - T_{2}}{R_{2r} + R_{34}} \qquad (20)$$

where $T_{41} = T_{34} = 70 \,°C$.

In the case of Dirichlet's conditions, the thermal resistances of convection are fixed at infinity (i.e., $R_{41} = R_{34} = \infty$). A fixed heat flux BC can be considered. In this case, (16) and (17) can be modified as follows

$$P_{41} = \frac{T_{41} - T_{25}}{R_{25r}} \qquad (21)$$

$$P_{34} = \frac{T_{34} - T_{2}}{R_{2r}} \qquad (22)$$

where $P_{34}$ and $P_{41}$ are imposed heat fluxes. The periodicity condition is satisfied by connecting nodes 1 and 8, 9 and 16, 17 and 24, 25 and 32.

*3- Global Matrix*

The 48 equations are represented in matrix form $gm(48,48)$ without replacing the known temperatures at the BCs. The second member of the system represents the entire values of $P_i$ with a vector $f(48)$ which represent the power losses in the machine. We begin the assembly gm with the internal elements of the conductive problem, then the equations of thermal convection at the rotor shaft and the stator ambient air. The introduction of BCs by convection is given by the thermal resistances and the fixed temperature (viz., 70°C) at the rotor shaft and the external air of the machine. The fixed temperatures are introduced into the global matrix and vector using high diagonal numbers in the global matrix and vector. These steps for constructing the global matrix are given in Appendix A.

The global matrix and vector for the 48 unknown temperatures are solved by direct method.

$$T = gm^{-1}.f \qquad (23)$$

The obtained solution vector $T(48)$ allows us to calculate the density of heat flux. The radial density of heat flux for each element can be calculated by

$$q(isks(4,k),k) = \frac{T_k - T_{isks(4,k)}}{S_{kr}\left(R_{kr} + R_{isks(4,k)r}\right)} \qquad (24)$$

$$q(isks(5,k),k) = \frac{T_k - T_{isks(5,k)}}{S_{kr}\left(R_{kr} + R_{isks(5,k)r}\right)} \qquad (25)$$

The tangential density of heat flux for each element is obtained by

$$q(isks(1,k),k) = \frac{T_k - T_{isks(1,k)}}{S_{k\theta}\left(R_{k\theta} + R_{isks(1,k)\theta}\right)} \qquad (26)$$

$$q(isks(3,k),k) = \frac{T_k - T_{isks(3,k)}}{S_{k\theta}\left(R_{k\theta} + R_{isks(3,k)\theta}\right)} \qquad (27)$$

where $S_{kr}$ and $S_{k\theta}$ are respectively the surface of an element in the $r$- and $\theta$-direction.

For the analyzed inset-PM machine with TRNM, $ms = 36$, $ns = 360$, the total number of elements and nodes is 12,960. The number of additional nodes to consider convection heat transfer at the external radius is 360 and at the shaft is 360. The dimensions of the global matrix and vector are respectively $13,680 \times 13,680$ and $13,680$.

### B. Analytical Method (AM)

*1- Problem Description, Assumptions and Partial Differential Equations (PDEs)*

In this section, we have improved the AM developed in [7] to consider the materials anisotropy in both directions $(r,\theta)$. The machine is subdivided into 7 regions, viz., Region I for the air-gap, Region IIj for the PMs, Region III for the stator yoke, Region IVi for the stator slots, Region V for the rotor yoke, Region VIj for the rotor teeth, and Region VIIi for the stator teeth.

In steady-state, PDEs representing the temperature distribution in each region are given by
- in Region I by

$$\frac{\lambda_{er}}{r}\frac{\partial}{\partial r}TI + \lambda_{er}\frac{\partial^2}{\partial r^2}TI + \frac{\lambda_{e\theta}}{r^2}\frac{\partial^2}{\partial \theta^2}TI = -p_e \qquad (28)$$

where $\lambda_{er}$ and $\lambda_{e\theta}$ are respectively the thermal conductivities (in $W/mK$) of the air-gap in the $r$- and $\theta$-direction, and $p_e$ is the windage loss density (in $W/m^3$).

- in Regions IIj by

$$\frac{\lambda_{mr}}{r}\frac{\partial}{\partial r}TII_j + \lambda_{mr}\frac{\partial^2}{\partial r^2}TII_j + \frac{\lambda_{m\theta}}{r^2}\frac{\partial^2}{\partial \theta^2}TII_j = -Pm_j \qquad (29)$$

where $\lambda_{mr}$ and $\lambda_{m\theta}$ are respectively the thermal conductivities (in $W/mK$) of PMs in the $r$- and $\theta$-direction, and $Pm_j$ the power loss density of the $j^{th}$ Region II (in $W/m^3$).



(a) Non-homogenous BCs.

$q_r^r = L(r)$    $T^\theta = g(\theta)$

$q_r^\theta = 0$    $T^r = 0$

$\Delta T^r + \Delta T^\theta = 0$

$T^\theta = f(\theta)$    $q_r^r = R(r)$

$T^r = 0$    $q_r^\theta = 0$

(b) Principle of superposition.

**Fig. 5:** Region with non-homogenous BCs [7].

- in Region III by

$$\frac{\lambda_{sr}}{r}\frac{\partial}{\partial r}TIII + \lambda_{sr}\frac{\partial^2}{\partial r^2}TIII + \frac{\lambda_{s\theta}}{r^2}\frac{\partial^2}{\partial\theta^2}TIII = -p_s \tag{30}$$

where $\lambda_{sr}$ and $\lambda_{s\theta}$ are respectively the thermal conductivities (in $W/mK$) of the stator yoke in the $r$- and $\theta$-direction, and $p_s$ the power loss density in the stator iron (in $W/m^3$). This power loss is considered uniform and constant in the whole stator iron.

- in Regions IVi by

$$\frac{\lambda_{sltr}}{r}\frac{\partial}{\partial r}TIV_i + \lambda_{sltr}\frac{\partial^2}{\partial r^2}TIV_i + \frac{\lambda_{slt\theta}}{r^2}\frac{\partial^2}{\partial\theta^2}TIV_i = -Psl_i \tag{31}$$

where $\lambda_{sltr}$ and $\lambda_{slt\theta}$ are respectively the thermal conductivities (in $W/mK$) of stator slot in the $r$- and $\theta$-direction, and $Psl_i$ the Joule and proximity losses densities (in $W/m^3$). This power is considered uniform and constant in each stator slot.

- in Region V by

$$\frac{\lambda_{rr}}{r}\frac{\partial}{\partial r}TV + \lambda_{rr}\frac{\partial^2}{\partial r^2}TV + \frac{\lambda_{r\theta}}{r^2}\frac{\partial^2}{\partial\theta^2}TV = -p_r \tag{32}$$

where $\lambda_{rr}$ and $\lambda_{r\theta}$ are respectively the thermal conductivities (in $W/mK$) of rotor yoke in the $r$- and $\theta$-direction, and $p_r$ the power loss density in the rotor iron (in $W/m^3$). This power loss is considered uniform and constant in the rotor iron.

- in Regions VIj by

$$\frac{\lambda_{ar}}{r}\frac{\partial}{\partial r}TVI_j + \lambda_{ar}\frac{\partial^2}{\partial r^2}TVI_j + \frac{\lambda_{a\theta}}{r^2}\frac{\partial^2}{\partial\theta^2}TVI_j = -Pdr_j \tag{33}$$

where $\lambda_{ar}$ and $\lambda_{a\theta}$ are respectively the thermal conductivities (in $W/mK$) of rotor tooth in the $r$- and $\theta$-direction, and $Pdr_j$ the power loss density in the rotor tooth (in $W/m^3$). This power loss is considered uniform and constant in the rotor tooth.
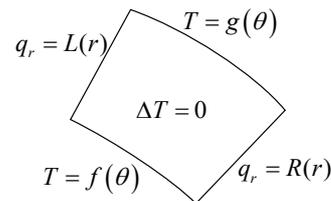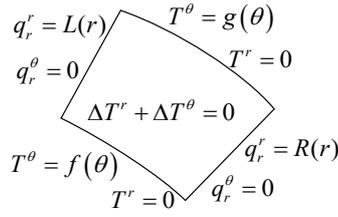
- in Regions VIIi by

$$\frac{\lambda_{dsr}}{r}\frac{\partial}{\partial r}TVII_i + \lambda_{dsr}\frac{\partial^2}{\partial r^2}TVII_i + \frac{\lambda_{ds\theta}}{r^2}\frac{\partial^2}{\partial\theta^2}TVII_i = -Pds_i \tag{34}$$

where $\lambda_{dsr}$ and $\lambda_{ds\theta}$ are respectively the thermal conductivities (in $W/mK$) of the stator tooth in the $r$- and $\theta$-direction, and $Pds_i$ the power loss density in the stator tooth (in $W/m^3$). This power loss is considered uniform and constant in the stator tooth.

Using $\mathbf{q} = -[\lambda]\cdot\nabla\mathbf{T}$, the heat flux density components (in W/m²) in polar coordinates are defined as

$$q_r = -\lambda_r\frac{\partial T(r,\theta)}{\partial r} \tag{35}$$

$$q_\theta = -\frac{\lambda_\theta}{r}\frac{\partial T(r,\theta)}{\partial\theta} \tag{36}$$

where $\lambda_r$ and $\lambda_\theta$ are respectively the thermal conductivities in the $r$- and $\theta$-direction.

*2- Temperature Solution in each SD*

The steady-state heat transfer in the inset-PM machine is studied using the improved 2-D exact SD technique presented in [7]. The general solutions of the above PDEs in non-homogenous BCs [see Fig. 5(a)] are deduced by applying the superposition principle [8]-[9] [see Fig. 5(b)] and using the Fourier's series as well as the separation of variables method. The Laplace's equations in Region I, III and V have homogeneous BCs and Region II, IV, VI and VII present non-homogeneous BCs. The solution of Laplace's equation

$$\frac{\lambda_r}{r}\frac{\partial}{\partial r}T(r,\theta) + \lambda_r\frac{\partial^2}{\partial r^2}T(r,\theta) + \frac{\lambda_\theta}{r^2}\frac{\partial^2}{\partial\theta^2}T(r,\theta) = 0 \tag{37}$$

Using the separation of variables method by replacing $T(r,\theta) = R(r)\Theta(r,\theta)$ gives

$$\frac{r\lambda_r\dfrac{dR(r)}{dr}}{R(r)} + \frac{r^2\lambda_r\dfrac{d^2R(r)}{dr^2}}{R(r)} = \mu^2 \text{ and/or } -\mu^2 \tag{38}$$

$$\frac{-\lambda_\theta\dfrac{d^2\Theta(\theta)}{d\theta^2}}{\Theta(\theta)} = \mu^2 \text{ and/or } -\lambda^2 \tag{39}$$

where $\mu^2$ is positive constant.

For a positive constant equal to $\mu^2$, the solutions are

$$R_1(r) = C1.r^{\frac{\mu}{\sqrt{\lambda_r}}} + C2.r^{-\frac{\mu}{\sqrt{\lambda_r}}} \tag{40}$$

$$\Theta_1(\theta) = C3\sin\left(\frac{\mu\theta}{\sqrt{\lambda_\theta}}\right) + C4\cos\left(\frac{\mu\theta}{\sqrt{\lambda_\theta}}\right) \tag{41}$$

For a negative constant equal to $-\mu^2$, the solutions are

$$R_2(r) = E1.\cos\left(\frac{\mu\ln(r)}{\sqrt{\lambda_r}}\right) + E2.\sin\left(\frac{\mu\ln(r)}{\sqrt{\lambda_r}}\right) \tag{42}$$

$$\Theta_2(\theta) = E3\sinh\left(\frac{\mu\theta}{\sqrt{\lambda_\theta}}\right) + E4\cosh\left(\frac{\mu\theta}{\sqrt{\lambda_\theta}}\right) \tag{43}$$

For the constant equal to zero, the solutions are

$$R_3(r) = B1 + B2\ln(r) \tag{44}$$

$$\Theta_3(\theta) = A1.\theta + A2 \tag{45}$$

In Region I, III and V with the homogenous BCs and periodicity equal to $2\pi$, the constant A1=0 and $\mu = n\sqrt{\lambda_\theta}$ (with $n$ is a positive integer). The periodic regions II, IV, VI and VII in the $r$- and $\theta$-direction with the non-homogeneous BCs have the constant A1=0 and $\mu = \dfrac{n\pi}{a}\sqrt{\lambda_\theta}$ in the $\theta$-direction for PMs region where $a$ is the PM-opening angle and $\mu = \dfrac{n\pi}{\ln\left(\dfrac{r_2}{r_1}\right)}\sqrt{\lambda_r}$ in the $r$-direction. The particular solution of Poisson's equations (28) to (34) in each SD is given by

$$T_p = -pr^2 / 4\lambda_r \tag{46}$$

where $p$ is a volumetric constant power loss in each SD.

The solution of (28) in the air-gap with the homogenous BCs is given for each harmonic $n$ by

$$TI(r,\theta) = \begin{vmatrix} R_3(r)\Theta_3(r,\theta) + R_1(r)\Theta_1(r,\theta) \\ \cdots + T_p \end{vmatrix} \tag{47}$$

and can be written as

$$TI(r,\theta) = -pr^2/(4\lambda_{er}) + A1_0 + A2_0\ln(r)$$
$$L + \sum_{n=1}^{nn}\left( A1_n\left(\frac{r}{R_s}\right)^{n.\tau_e} + A2_n\left(\frac{r}{R_m}\right)^{-n.\tau_e} \right)\sin(n\theta)$$
$$L + \sum_{n=1}^{nn}\left( A3_n\left(\frac{r}{R_s}\right)^{n.\tau_e} + A4_n\left(\frac{r}{R_m}\right)^{-n.\tau_e} \right)\cos(n\theta) \tag{48}$$

where $\tau_e = \sqrt{\lambda_{e\theta}/\lambda_{er}}$, and $p_e$ the windage loss in the air-gap.

The solution of (29) in Region IIj with the non-homogeneous BCs is given for each harmonic by

$$TII_j = \begin{vmatrix} R_3(r)\Theta_3(r,\theta) + R_1(r)\Theta_1(r,\theta) + R_2(r)\Theta_2(r,\theta) \\ \cdots + T_p \end{vmatrix} \tag{49}$$

and can be reduced using BCs presented in Fig. 5 as

$$TII_j(r,\theta) = B1_{j,0} + B2_{j,0}\ln(r) - Pm_j r^2/(4\lambda_{mr})$$
$$+\sum_{m=1}^{mm}\left( B1_{j,m}\left(\frac{r}{R_m}\right)^{fra_m.\tau_m} + B2_{j,m}\left(\frac{r}{R_r}\right)^{-fra_m.\tau_m} \right)\cos\left(fra_m\left(\theta - \theta1_j\right)\right)$$
$$+\sum_{k=1}^{kk}\begin{pmatrix} B3_{j,k}\dfrac{sh\left(\dfrac{fr_k}{\tau_m}\left(\theta-\theta a1_j\right)\right)}{sh\left(\dfrac{fr_k}{\tau_m}a\right)} \\ L + B4_{j,k}\dfrac{sh\left(\dfrac{fr_k}{\tau_m}\left(\theta-\theta a2_j\right)\right)}{sh\left(\dfrac{fr_k}{\tau_m}a\right)} \end{pmatrix}\sin\left(fr_k\ln\left(\frac{r}{R_r}\right)\right)$$
$$\tag{50}$$

where $fr_k = \dfrac{k\pi}{gg}$, $gg = \ln\left(\dfrac{R_m}{R_r}\right)$, $fra_m = \dfrac{m\pi}{a}$, $\theta a1_j = g_j - \dfrac{a}{2}$,

$\theta a2_j = g_j + \dfrac{a}{2}$ and $\tau_m = \sqrt{\lambda_{m\theta}/\lambda_{mr}}$.

The stator yoke represented by Region III has homogenous BCs and the solution of (30) is given by

$$TIII(r,\theta) = A6_0 + A5_0\ln(r) - p_s r^2/(4\lambda_{sr})$$
$$L + \sum_{n=1}^{nn}\left( A5_n\left(\frac{r}{R_{ext}}\right)^{n.\tau_s} + A6_n\left(\frac{r}{r_4}\right)^{-n.\tau_s} \right)\cos(n\theta)$$
$$L + \sum_{n=1}^{nn}\left( A7_n\left(\frac{r}{R_{ext}}\right)^{n.\tau_s} + A8_n\left(\frac{r}{r_4}\right)^{-n.\tau_s} \right)\sin(n\theta) \tag{51}$$

where $\tau_s = \sqrt{\lambda_{s\theta}/\lambda_{sr}}$.

The $Q_s$ stator slots represented by Region IVi has non-homogenous BCs, the solution of (31) is given by

$$TIV_i(r,\theta) = C1_{i,0} + C2_{i,0}\ln(r) - Psl_i r^2/(4\lambda_{slr})$$
$$+\sum_{m1=1}^{mm1}\left( C1_{i,m1}\left(\frac{r}{r_4}\right)^{fsc_{m1}.\tau_{sl}} + C2_{i,m1}\left(\frac{r}{R_s}\right)^{-fsc_{m1}.\tau_{sl}} \right)\cos\left(fsc_{m1}\left(\theta - \theta c1_i\right)\right)$$
$$+\sum_{k1=1}^{kk1}\begin{pmatrix} C3_{i,k1}\dfrac{sh\left(\dfrac{fs_{k1}}{\tau_{sl}}\left(\theta-\theta c1_i\right)\right)}{sh\left(\dfrac{fs_{k1}}{\tau_{sl}}c\right)} \\ L + C4_{i,k1}\dfrac{sh\left(\dfrac{fs_{k1}}{\tau_{sl}}\left(\theta-\theta c2_i\right)\right)}{sh\left(\dfrac{fs_{k1}}{\tau_{sl}}c\right)} \end{pmatrix}\sin\left(fs_{k1}\ln\left(\frac{r}{R_s}\right)\right)$$
$$\tag{52}$$

where $fs_{k1} = \dfrac{k1\pi}{ff}$, $ff = \ln\left(\dfrac{r_4}{R_s}\right)$, $fsc_{m1} = \dfrac{m1\pi}{c}$, $\theta c1_i = \alpha_i - \dfrac{c}{2}$,

$\theta c2_i = \alpha_i + \dfrac{c}{2}$, and $\tau_{sl} = \sqrt{\lambda_{sl\theta}/\lambda_{slr}}$.

The rotor yoke represented by Region V has homogenous BCs, the solution of (32) is given by

$$TV(r,\theta) = A10_0 + A9_0\ln(r) - p_r \cdot r^2/(4\lambda_{rr})$$
$$L + \sum_{n=1}^{nn}\left( A9_n\left(\frac{r}{R_r}\right)^{n.\tau_r} + A10_n\left(\frac{r}{R_i}\right)^{-n.\tau_r} \right)\cos(n\theta)$$
$$L + \sum_{n=1}^{nn}\left( A11_n\left(\frac{r}{R_r}\right)^{n.\tau_r} + A12_n\left(\frac{r}{R_i}\right)^{-n.\tau_r} \right)\sin(n\theta) \tag{53}$$

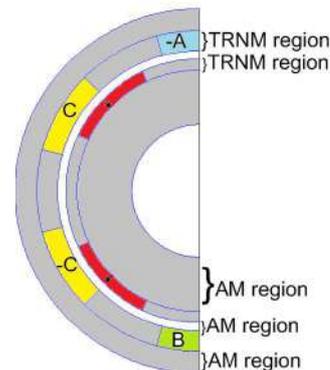where $\tau_r = \sqrt{\lambda_{r\theta}/\lambda_{rr}}$.



Fig. 6: Distribution of AM and TRNM regions in the HM (air-gap zoomed).

In Region VIj with the non-homogenous BCs, the solution of (33) is given by

$$TVI_j(r,\theta) = B6_{j,0} + B5_{j,0}\ln(r) - Pdr_j \cdot r^2/(4\lambda_{ar})$$

$$+\sum_{m=1}^{mm}\left( B5_{j,m}\left(\frac{r}{R_m}\right)^{frb_m\cdot\tau_{dr}} + B6_{j,m}\left(\frac{r}{R_r}\right)^{-frb_m\cdot\tau_{dr}} \right)\cos\left(frb_m\left(\theta - \theta b1_j\right)\right)$$

$$+\sum_{k=1}^{kk}\begin{pmatrix} \dfrac{B7_{j,k}\,sh\left(\dfrac{fr_k}{\tau_{dr}}\left(\theta - \theta b1_j\right)\right)}{sh\left(\dfrac{fr_k}{\tau_{dr}}b\right)} \\ \\ L + \dfrac{B8_{j,k}\,sh\left(\dfrac{fr_k}{\tau_{dr}}\left(\theta - \theta b2_j\right)\right)}{sh\left(\dfrac{fr_k}{\tau_{dr}}b\right)} \end{pmatrix}\sin\left(fr_k\ln\left(\frac{r}{R_r}\right)\right)$$

$$(54)$$

where $\quad frb_m = \dfrac{m\pi}{b}, \quad \theta b1_j = \beta_j - \dfrac{b}{2}, \quad \theta b2_j = \beta_j + \dfrac{b}{2} \quad$ and $\tau_{dr} = \sqrt{\lambda_{dr\theta}/\lambda_{drr}}$ .

In Region VIIi representing the stator teeth with $Pds_i$ power losses, we have

$$TVII_i(r,\theta) = C5_{i,0} + C6_{i,0}\ln(r) - Pds_i \cdot r^2/(4\lambda_{dsr})$$

$$+\sum_{m1=1}^{mm1}\left( C5_{i,m1}\left(\frac{r}{r_4}\right)^{fsd_{m1}\cdot\tau_{ds}} + C6_{i,m1}\left(\frac{r}{R_s}\right)^{-fsd_{m1}\cdot\tau_{ds}} \right)\cos\left(fsd_{m1}\left(\theta - \theta d1_i\right)\right)$$

$$+\sum_{k1=1}^{kk1}\begin{pmatrix} C7_{i,k1}\dfrac{sh\left(\dfrac{fs_{k1}}{\tau_{ds}}\left(\theta - \theta d1_i\right)\right)}{sh\left(\dfrac{fs_{k1}}{\tau_{ds}}d\right)} \\ \\ L + C8_{i,k1}\dfrac{sh\left(\dfrac{fs_{k1}}{\tau_{ds}}\left(\theta - \theta d2_i\right)\right)}{sh\left(\dfrac{fs_{k1}}{\tau_{ds}}d\right)} \end{pmatrix}\sin\left(fs_{k1}\ln\left(\frac{r}{R_s}\right)\right)$$

$$(55)$$

where $fs_{k1} = \dfrac{k1\pi}{ff}$ , $ff = \ln\left(\dfrac{r_4}{R_s}\right)$ , $fsd_{m1} = \dfrac{m1\pi}{d}$ , $\theta d1_i = \gamma_i - \dfrac{d}{2}$ ,

$\theta d2_i = \gamma_i + \dfrac{d}{2}$ and $\tau_{ds} = \sqrt{\lambda_{ds\theta}/\lambda_{dsr}}$ .

The anisotropy coefficient can be defined as

$$\psi = \lambda_\theta/\lambda_r \tag{56}$$

*3- ICs and their Development*

To determine the unknown coefficients of temperature in each SD, there are 18 ICs, viz., 14 ICs are in the $\theta$-direction and 4 ICs in the $r$-direction [7]. The development of ICs permits to obtain an equations system whose unknowns are the coefficients of Fourier's series solution in each SD. The solving of this system gives the temperature and heat flux distribution in the whole machine.

For the studied inset-PM machine with the harmonics number in each SD: $mml = 40$ , $kkl = 40$ , $mm = 50$ , $kk = 50$ and $nn = 200$ , the dimensions of global matrix and vector are

respectively $5,966 \times 5,966$ and $5,966$ .
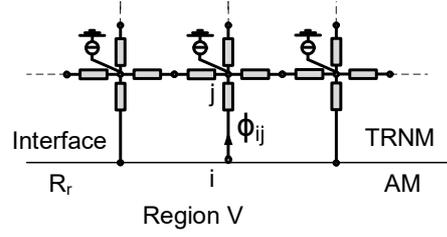


**Fig. 7:** Heat flux density IC between TRNM and AM at $R_r$ .

*C. Hybrid Method (HM)*

It is well known that AM is more accurate with low computational time than TRNM. However, when the number of SDs is high (i.e., the ICs number in both directions is also high) and their dimensions are small, the harmonics number can be very low and then the accuracy of AM can be very low. This situation can be found in the case of inset-PM machine with high number of stator and rotor slots. For this, we propose in this paper to model the stator and rotor slots regions (also the stator and rotor teeth) with TRNM and the other regions with AM [see Fig. 6].

To release the coupling between AM and TRNM, nodes of coupling are added to TRNM presented in Section A at the radii separating the two methods (ICs between AM and TRNM at $R_r$ , $R_m$ , $R_s$ and $r_4$ ).

For example, at the radius $R_r$ , the continuity conditions of AM are given by

$$TV\left(R_r,\theta\right) = TII_j\left(R_r,\theta\right) \tag{57}$$

$$TV\left(R_r,\theta\right) = TVI_j\left(R_r,\theta\right) \tag{58}$$

$$qV_r\left(R_r,\theta\right) = \begin{cases} qII_{r,j}\left(R_r,\theta\right) & for\ \theta \in \left[g_j + \dfrac{a}{2}; g_j - \dfrac{a}{2}\right] \\ qVI_{r,j}\left(R_r,\theta\right) & for\ \theta \in \left[\beta_j + \dfrac{b}{2}; \beta_j - \dfrac{b}{2}\right] \end{cases} \tag{59}$$

In the HM, the Region IIj and VIj are modeled by TRNM and Region V by AM. To satisfy the ICs (57) and (58), the nodes temperatures of TRNM at $R_r$ are written as a $2\pi$ periodic function using discrete Fourier series [12] as

$$T_{TRNM\_Rr}\left(\theta\right) = \frac{a_0}{2} + \sum_{n=1}^{nn}a_n\cos\left(n\theta\right) + b_n\sin\left(n\theta\right) \tag{60a}$$

$$a_0 = \frac{2}{2\cdot nn}\cdot\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)}T_k \tag{60b}$$

$$a_n = \frac{2}{2\cdot nn}\cdot\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)}T_k\cos\left(n\theta_k\right)\ \forall n \tag{60c}$$

$$b_n = \frac{2}{2\cdot nn}\cdot\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)}T_k\sin\left(n\theta_k\right)\ \forall n \tag{60d}$$

where $nn = ns/2$ , $T_k$ are the temperatures at the boundary nodes $nodb\left(1,i,1\right)$ to $nodb\left(1,i,ns\right)$ situated at $R_r\left(i = 1\right)$ and

$\theta_k$ the angular position of boundary nodes at the radius $R_r$. It is important to note that $nn$ is also the total harmonics number of the AM solution in the Region V.

This Fourier series function permits to replace the ICs (57) and (58) by

$$TV(R_r,\theta) = T_{TRNM\_Rr}(\theta) \qquad (61)$$

This above equation permits to get 3 equations as

$$-\frac{1}{4}\frac{q_r R_r^2}{\lambda_{rr}} + A9_0 \ln(R_r) + A10_0 = \frac{1}{2.nn}\cdot\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)} T_k \qquad (62)$$

$$A9_n + A10_n\left(\frac{R_r}{R_i}\right)^{-n.\tau_r} = \frac{1}{nn}\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)} T_k \cos(n\theta_k) \qquad (63)$$

$$A11_n + A12_n\left(\frac{R_r}{R_i}\right)^{-n.\tau_r} = \frac{1}{nn}\sum_{k=nodb(1,1,1)}^{nodb(1,1,ns)} T_k \sin(n\theta_k) \qquad (64)$$

To satisfy the heat flux density IC at $R_r$, i.e., (59), each boundary node is considered receiving a radial heat flux $\phi_{ij}$ from the AM region [see Fig. 7] as [20]-[21]

$$\phi_{ij} = \frac{T_i - T_j}{R_{ij}} = -\lambda_{rr} R_r L_u \int_{\theta_k-d\theta}^{\theta_k+d\theta} \frac{\partial TV(r,\theta)}{\partial r} d\theta \qquad (65)$$

where $\theta_k$ is the angular position of interface nodes and $d\theta$ is half opening angle of an element equal to 0.5° (each element of TRNM has an angular opening equal 1° in TRNM).

The development of (65) gives

$$\phi_{ij} = \left(-L_u\lambda_{rr}A9_0 + \frac{R_r^2}{2}L_u q_r\right)\int_{\theta_k-d\theta}^{\theta_k+d\theta} d\theta$$

$$\cdots + L_u\lambda_{rr}n\tau_r\left(-A9_n + A10_n\left(\frac{R_r}{R_i}\right)^{-n\tau_r}\right)\int_{\theta_k-d\theta}^{\theta_k+d\theta}\cos(n\theta)d\theta$$

$$\cdots + L_u\lambda_{rr}n\tau_r\left(-A11_n + A12_n\left(\frac{R_r}{R_i}\right)^{-n\tau_r}\right)\int_{\theta_k-d\theta}^{\theta_k+d\theta}\sin(n\theta)d\theta \quad (66)$$

In the HM, $ms=21$ and $ns=360$ with 11 radii for TRNM rotor region and 11 radii for TRNM stator region. The elements and nodes number of TRNM stator region is $nsgs=3,600$ and TRNM rotor region is $nsgr=3,600$. In TRNM rotor region, we have added 2 boundaries additional nodes at $R_r(i=1)$ and $R_m(i=11)$ represented by vectors $nodb(1,1,j)$ and $nodb(2,11,j)$ where $j$ varies from 1 to $ns$. For TRNM stator region, we have added 2 boundaries nodes at $R_s(i=12)$ and $r_4(i=22)$ represented by vectors $nodb(3,12,j)$ and $nodb(4,22,j)$. The nodes of TRNM connected to the 4 boundary nodes $nodb(1,1,j)$, $nodb(2,11,j)$, $nodb(3,12,j)$, $nodb(4,22,j)$ are $Bdrti(j)=nods(1,j)$, $Bdrto(j)=nods(10,j)$, $Bdsti(j)=nods(12,j)$ and

$Bdsto(j)=nods(ms,j)$ respectively. The total elements and nodes number of TRNM mesh considering the nodes number at the 4 BCs is equal to $nsgb=8,640$. There are 3 AM regions in the HM where each AM region necessitates $2+4\cdot nn$ unknowns. The total nodes number of TRNM regions and AM regions is $nsgb1=10,806$.

After defining the nodes connectivity as described in Section A, the assembly of the global matrix from TRNM and AM is performed. We start with the TRNM nodes representing the rotor slots and teeth. The introduction of equations (62) to (64) in the global matrix $gm$ is performed as given in Appendix B.

There are 360 equations (66) to be introduced in the global matrix depending on the number of boundary nodes at $R_r$. The equations (62) to (64) and (66) related to the ICs (57) to (59) at $R_r$ have been introduced in the global matrix $gm$ and the global vector $f$ of the HM. The coupling is performed between the AM temperature of rotor $TV(r,\theta)$ (53) and TRNM at $R_r$. In the same way, the coupling between $TI(r,\theta)$ (48) and rotor TRNM at $R_m$ and $R_s$, and the coupling between $TIII(r,\theta)$ (50) and stator TRNM at $r_4$ can be established and assembled to the global system. The introduction of convection heat transfer at the external radius and shaft is done using AM.

**TABLE. I**
PARAMETERS OF STUDIED INSET-PM MACHINE.

| Symbol | Parameters | Value |
|---|---|---|
| $B_{rm}$ | Remanence flux density of PMs | 1.3 $T$ |
| $\mu_{rm}$ | Relative permeability of PMs | 1.0277 |
| $N_c$ | Number of conductors per stator slot | 23 |
| $I_m$ | Peak phase current | 7 $A$ |
| $Q_s$ | Number of stator slots | 6 |
| $c$ | Stator slot-opening | 30 $deg.$ |
| $a$ | PM-opening | 40 $deg.$ |
| $p$ | Number of pole pairs | 2 |
| $R_{ext}$ | Radius of the external stator surface | 110 $mm$ |
| $r_4$ | Outer radius of stator slot | 97 $mm$ |
| $R_s$ | Radius of the stator inner surface | 80.5 $mm$ |
| $R_m$ | Radius of the rotor outer surface at the PM | 79.7 $mm$ |
| $R_r$ | Radius of the rotor inner surface at the PM | 73 $mm$ |
| $g$ | Air-gap length | 0.8 $mm$ |
| $L_u$ | Axial length | 40 $mm$ |
| $\Omega$ | Mechanical speed | 500 $rpm$ |

**TABLE. II**
PARAMETERS OF THE THERMAL MODEL.

| Symbol | Parameters | Value |
|---|---|---|
| $\lambda_e$ | Thermal conductivity of air-gap | 0.03 W/(m K) |
| $\lambda_a$ | Thermal conductivity of air | 0.03 W/(m K) |
| $\lambda_m$ | Thermal conductivity of PMs | 9 W/(m K) |
| $\lambda_s$ | Thermal conductivity of stator iron | 55 W/(m K) |
| $\lambda_r$ | Thermal conductivity of rotor iron | 55 W/(m K) |
| $\lambda_{sl}$ | Thermal conductivity of stator slot coil | 1.73 W/(m K) |
| $p_s$ | Stator core losses | 4.07 W |
| $p_r$ | Rotor core losses | 0.31 W |
| $p_m$ | PM losses | 7.94 W |
| $p_{sl}$ | Stator slot losses | 18.12 $W$ |
| $P_e$ | Windage losses in the air-gap | 5 W |
| $h_r$ | Convection coefficient inside the rotor | 100 W/(m² K) |
| $T_{int}$ | Temperature inside the rotor | 70 °C |
| $h_s$ | Convection coefficient outside the stator | 100 W/(m² K) |
| $T_{ext}$ | Temperature outside the stator | 70 °C |

| $\psi$ | Anisotropy coefficients | 0.5/1/1.5 |
|---|---|---|

### III. TEMPERATURE AND HEAT FLUX RESULTS

The parameters and dimensions of the studied inset-PM machine are given in Table I. The machine has a simple distributed 4 poles winding. The power losses of the inset-PM machine at 500 *rpm* as well as the thermal conductivities, convection coefficients and ambient temperatures used in the thermal model are listed in Table II. The harmonics number of AM is $nn = 200$, $mm = 50$, $kk = 50$, $mm1 = 40$ and $kk1 = 40$. These harmonics numbers provide very good accuracy compared to Fem with a reasonable computation time. The average elements and nodes number of the Fem calculation [15] are respectively 109,168 and 55,484.

#### A. TRNM Results Without Materials Anisotropy and Validation with Fem

For the 6-slots/4-poles inset-PM machine, the temperature distribution at speed of 500 *rpm* in the whole machine without taking into account the materials anisotropy is shown in Fig. 8. We can observe that the temperature is higher inside the stator slots where power loss is higher [see Table II]. The directions of heat flux are represented with vectors oriented to inside and outside the machine. This is due to convections coefficients imposed outside and inside the machine. In the middle of the air-gap, the distribution of temperature and heat flux components calculated by the developed TRNM and Fem using the parameters and power losses in Table II are given in Fig. 9.
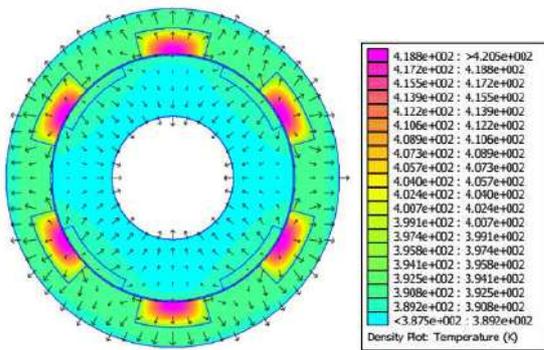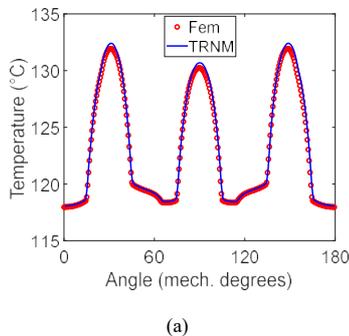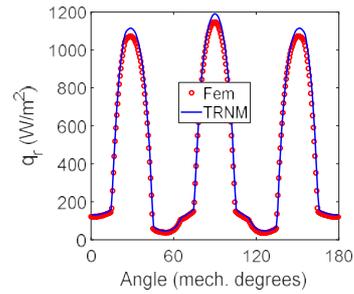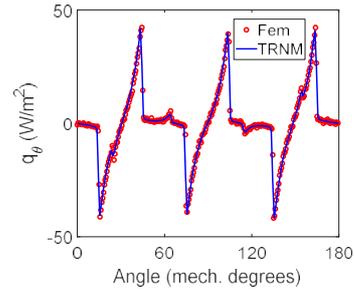


**Fig. 8:** Temperature and flux distribution obtained using Fem.



(a)



(b)



(c)

**Fig. 9:** Temperature and heat flux components distribution in the air-gap at the radius 80.05 mm.
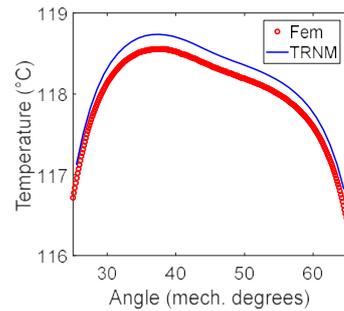


(a) In the $\theta$-direction at 76.77 mm.
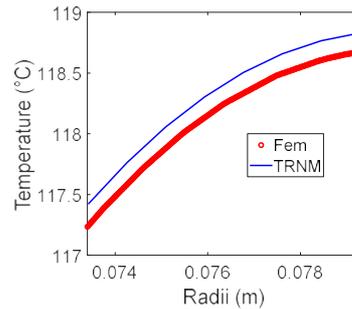


(b) In the $r$-direction at 45.5 degrees.

**Fig. 10:** Temperature in the middle of the first PM.
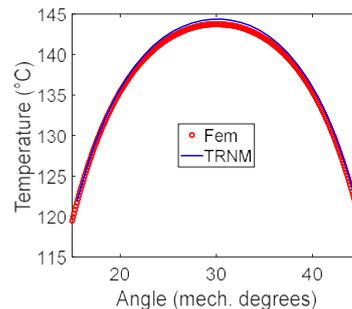


(a) In the $\theta$-direction at 87.72 mm.
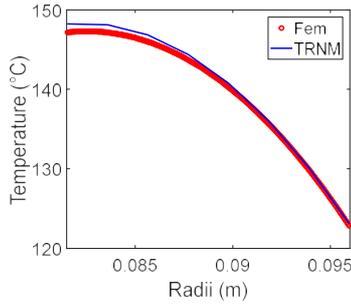
(b) In the *r*-direction at 30.5 degrees.

**Fig. 11:** Temperature in the middle of the first stator slot.

To show the ability of the TRNM to predict the temperature distribution in the PMs and the stator slots, the temperature curves in the $\theta$- and $r$-direction obtained using TRNM are shown in Figs. 10 ~ 11 and compared with Fem. The TRNM results are in good agreement with the Fem results. A small difference exist between TRNM and Fem results in the PM region. This difference is due to the meshing of machine which can be improved as done in TRNM [22]-[23]. A parametric analysis with variation of the convective coefficients $h_s$ and $h_r$ is also performed. Fig. 12 shows the temperature and heat flux distribution in the inset-PM machine when the convective coefficients $h_s = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_r = 100\,W/\!\left(m^2 \cdot K\right)$. The vecors of heat flux are oriented to inside and outside the machine. It can be seen that the heat flux oriented to inside the rotor is higher than the heat flux oriented to outside the stator. The corresponding air-gap temperature distribtion is shown in Fig. 13. It can be observed that the temperature is higher than in the case with $h_s = 100\,W/\!\left(m^2 \cdot K\right)$.

For the case with $h_r = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_s = 100\,W/\!\left(m^2 \cdot K\right)$, Figs. 14 ~ 15 show the temperature distribution in the machine obtained using Fem and temperature distribution in the middle of the air-gap obtained using the developed TRNM and Fem. It can be seen from Fig. 14 that the heat flux oriented to outside the stator is higher than the heat flux oriented to inside the rotor. Also, in this case, the TRNM results are very close to those of Fem. The variation of the temperature in the middle of the PM and stator slot when the convective coefficients $h_s$ and $h_r$ varies is shown in Figs. 16 ~ 17. The comparison of the TRNM results with those obtained by Fem confirms the validity of the proposed TRNM to predict the temperature and heat flux distribution in the inset-PM machine with a very good accuracy.



**Fig. 13:** Temperature distribution in the middle of the air-gap for $h_s = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_r = 100\,W/\!\left(m^2 \cdot K\right)$.



**Fig. 14:** Temperature and heat flux distribution in the inset-PM machine for $h_r = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_s = 100\,W/\!\left(m^2 \cdot K\right)$.



**Fig. 15:** Temperature distribution in the middle of air-gap for $h_r = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_s = 100\,W/\!\left(m^2 \cdot K\right)$.



(a) Temperature at the center of the first PM.





**Fig. 12:** Temperature distribution and heat flux in the inset-PM machine for $h_s = 20\,W/\!\left(m^2 \cdot K\right)$ and $h_r = 100\,W/\!\left(m^2 \cdot K\right)$.
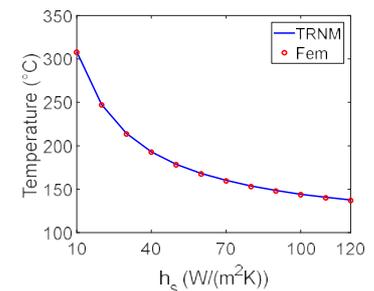
(b) Temperature at the center of the first stator slot.

**Fig. 16:** Temperature variation with varying $h_s$ and $h_r = 100\,W/\left(m^2 \cdot K\right)$ in a point at the center of PM and stator slot.
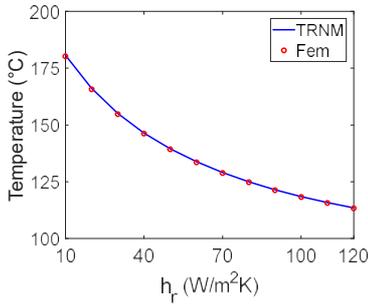
(a) Temperature at the center of the first PM.
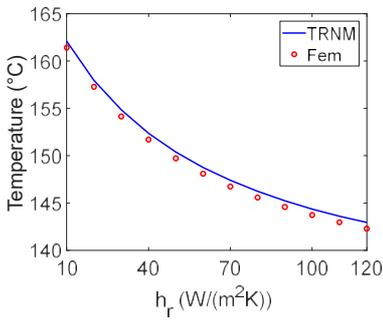
(b) Temperature at the center of the first stator slot.

**Fig. 17:** Temperature variation with varying $h_r$ and $h_s = 100\,W/\left(m^2 \cdot K\right)$ in a point at the center of PM and stator slot.

### B. AM Thermal Results with Materials Anisotropy and Validation with TRNM

The distribution of temperature and heat flux components in the middle of the air-gap obtained with AM and TRNM taking into account the materials anisotropy is shown in Fig. 18. We can observe a very good agreement between the AM and TRNM results. The temperature distribution in the middle of the first PM and the stator slot in the $\theta$- and $r$-direction [see Figs. 19 ~ 20] obtained analytically and with TRNM confirm the accuracy of the proposed AM. Again, we can show a small difference due to the mesh size adopted in TRNM.

When the cooling outside the inset-PM machine is not sufficient, i.e., $h_s = 20\,W/\left(m^2 \cdot K\right)$, the heat is not evacuated and the temperature is very high in the air-gap [see Fig. 21]. The same observation can be done in the case of insufficient cooling in the rotor shaft with $h_r = 20\,W/\left(m^2 \cdot K\right)$ [see Fig. 22]. In this case, the rotor temperature is high but lower than the case of low value of $h_s$. The variation of temperature in the middle of the first PM and the stator slot with the convection coefficient $h_s$ and $h_r$ is shown in Figs. 23 ~ 24. Those curves are very important for the design of stator winding insulation and PMs whose characteristics depend on temperature.
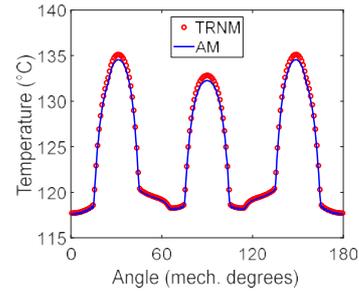
(a)
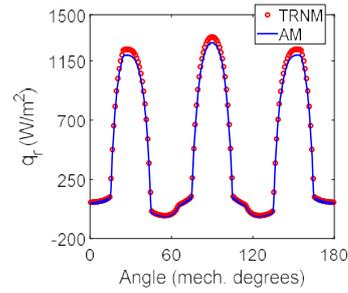
(b)
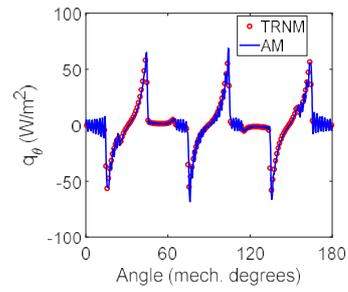
(c)

**Fig. 18:** Temperature and heat flux components distribution in the middle of the air-gap.
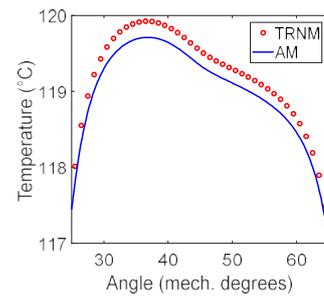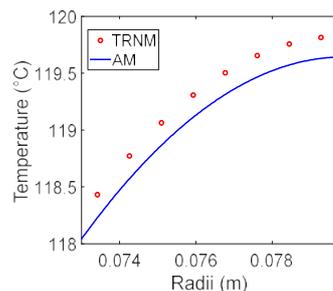
(a) In the $\theta$-direction.

(b) In the $r$-direction.

**Fig. 19:** Temperature distribution in the middle of the first PM.
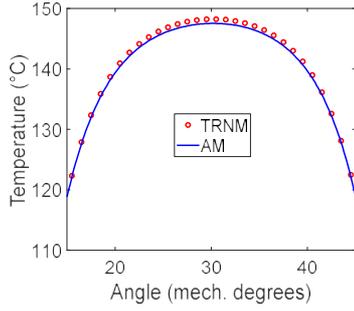


(a) In the $\theta$-direction.



(b) In the $r$-direction.

**Fig. 20:** Temperature in the middle of the first stator slot.
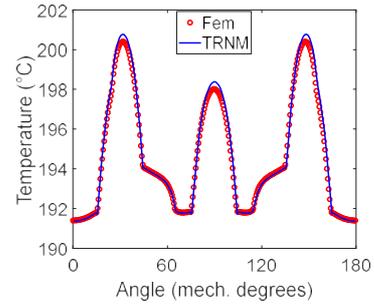


**Fig. 21:** Temperature distribution in the middle of the air-gap for $h_s = 20 \, W/(m^2 \cdot K)$ and $h_r = 100 \, W/(m^2 \cdot K)$.



**Fig. 22:** Temperature distribution in the middle of the air-gap for $h_r = 20 \, W/(m^2 \cdot K)$ and $h_s = 100 \, W/(m^2 \cdot K)$.
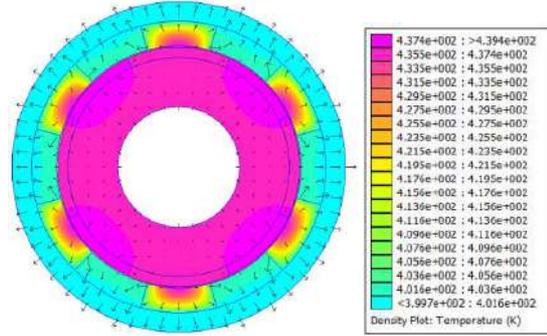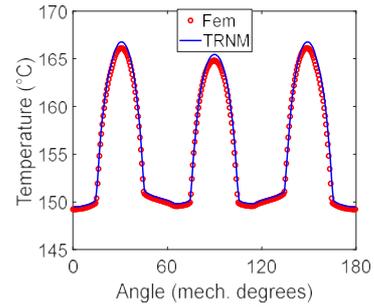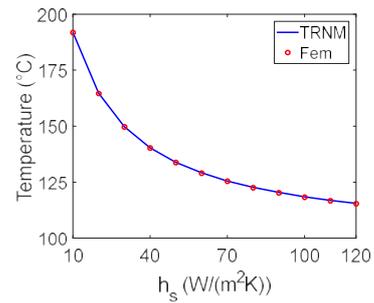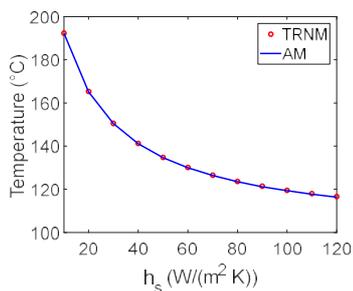


(a) Temperature at the center of the first PM.
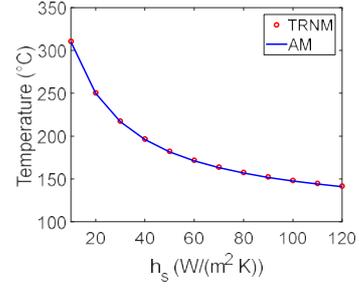


(b) Temperature at the center of the first stator slot.

**Fig. 23:** Temperature variation with varying $h_s$ and $h_r = 100 \, W/(m^2 \cdot K)$ in a point at the center of PM and stator slot.
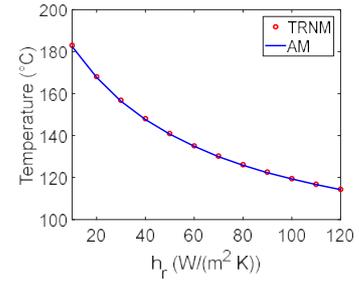


(a) Temperature at the center of the first PM.
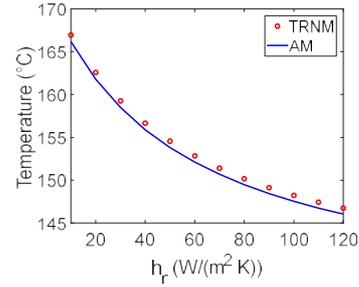


(b) Temperature at the center of the first stator slot.

**Fig. 24:** Temperature variation with varying $h_r$ and $h_s = 100 \, W/(m^2 \cdot K)$ in a point at the center of PM and stator slot.

### C. HM Thermal Results and Validation with TRNM and Fem

It is not easy to use AM in heat transfer prediction for rotating electrical machines with high number of stator slots and rotor poles. The ICs in the $r$- and $\theta$-direction are important and the dimensions of SDs are small, requiring a small harmonics number and thus lower accuracy. For this, it is appropriate to use HM. In this section, we apply the HM in both cases, with isotropic and anisotropic materials. For isotropic materials, the validation of results can be performed with Fem.

### 1- Isotropic Materials

The temperature and heat flux distribution in the middle of the air-gap is shown in Fig. 25. The HM results are very close to those from Fem. The temperature distribution in the $\theta$- and $r$-direction in the middle of the first PM and the first stator slot is shown in Figs. 26 ~ 27. The accuracy of HM is established also in those SDs where it is important to know the heat transfer for the insulation design. A small difference can be observed in the PM region between HM and Fem. The mesh size of the TRNM rotor has affected the HM results. For this, it is necessary to optimize the TRNM parts using them in HM.

The effect of cooling outside the inset-PM machine and inside the rotor shaft is represented with the convective coefficients $h_s$ and $h_r$ respectively. In Fig. 28, for $h_s = 20\,W/(m^2 \cdot K)$ which is small, we represent the temperature distribution in the middle of the air-gap. The temperature is higher compared to $h_s = 100\,W/(m^2 \cdot K)$.

For $h_r = 20\,W/(m^2 \cdot K)$ compared to $h_r = 100\,W/(m^2 \cdot K)$, the temperature distribution in the air-gap [see Fig. 29] is higher than the case with $h_r = 100\,W/(m^2 \cdot K)$. Low values of convective coefficients represent a barrier for heat transfer outside the stator and inside the rotor.
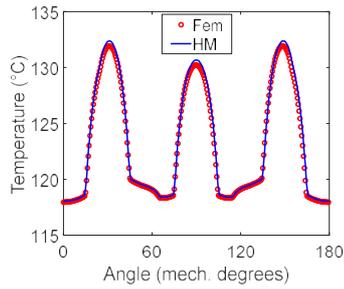


(a)
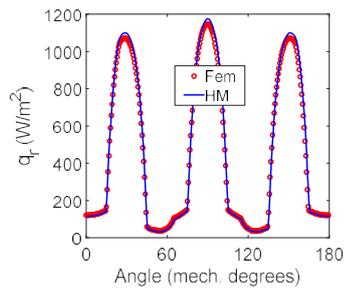


(b)
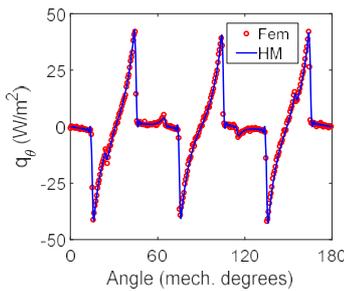


(c)

**Fig. 25:** Temperature and heat flux components distribution in the middle of the air-gap.
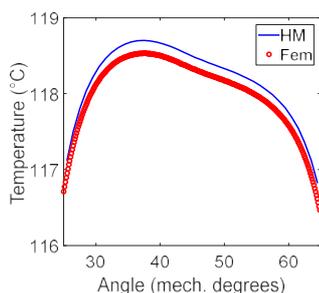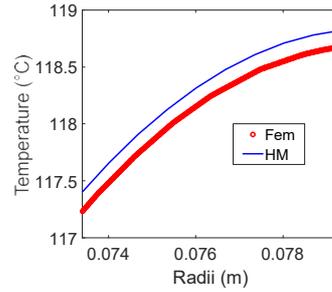


(a) In the $\theta$-direction.



(b) In the $r$-direction.

**Fig. 26:** Temperature in the middle of the first PM.
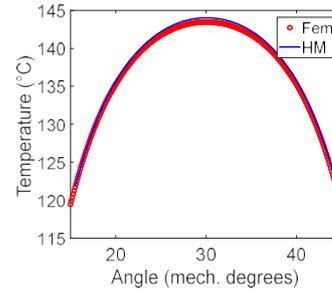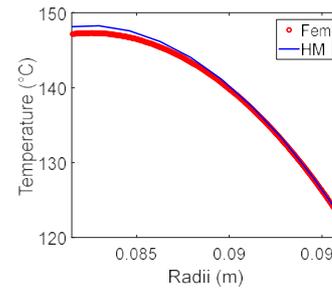


(a) In the $\theta$-direction.



(b) In the $r$-direction.

**Fig. 27:** Temperature in the middle of the first stator slot.
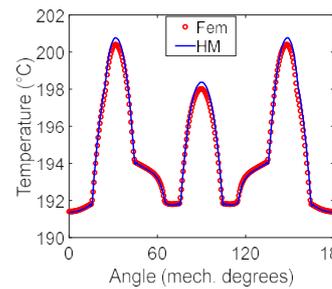


**Fig. 28:** Temperature distribution in the middle of the air-gap for $h_s = 20\,W/(m^2 \cdot K)$ and $h_r = 100\,W/(m^2 \cdot K)$.
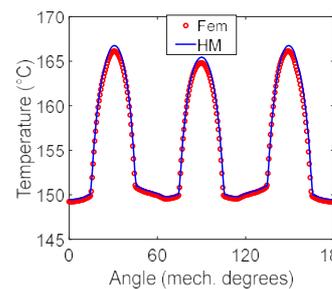


**Fig. 29:** Temperature distribution in the middle of the air-gap for $h_r = 20\,W/(m^2 \cdot K)$ and $h_s = 100\,W/(m^2 \cdot K)$.

## 2- Anisotropic Materials

In the analysis of heat transfer in the inset-PM machine taking into account the materials anisotropy, TRNM is used for the validation of results. This is due to Fem which use Cartesian representation of thermal conductivities [17] and [19]. For the dimensions and parameters of the studied machine [see Tables I and II] and an anisotropy coefficient equal to 0.5, we represent on Fig. 30 the distribution of temperature and heat flux in the middle of the air-gap. It can be seen that the temperature in the air-gap is higher than in the case of isotropic materials and the comparison between the two methods gives very good agreement. It is important to note that the anisotropy coefficient of materials is applied for rotor and stator iron, slots and PM without air-gap.

A parametric study is performed in this section as a function of the anisotropy coefficient (viz., $\psi = 0.5, 1, 1.5$). When $\psi = 1$, the materials are isotropic and when $\psi = 0.5$ the tangential value of thermal conductivity is smaller than the radial value. For $\psi = 1.5$, the tangential thermal conductivity is higher than the radial conductivity. The temperature distribution at the middle of the air-gap, the middle of the PM and the middle of the stator slot for different values of $\psi$ is shown in Figs. 31 ~ 33. It can be observed that the heat transfer in the machine is better when the tangential thermal conductivity of materials is higher than the radial conductivity. In this case, the temperature is lower. This remark is valid for the studied case where all materials of the machine have the same anisotropy coefficient, which is not true. A more realistic study should take into consideration consider the real values of thermal conductivity anisotropy in each region (i.e., slots, stator and rotor iron cores, PM). Moreover, it is important to note that the conductivity in the $z$-direction of rotating electrical machines is mostly affected by materials anisotropy and a 3-D study is appropriate.
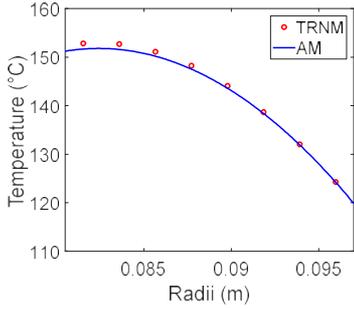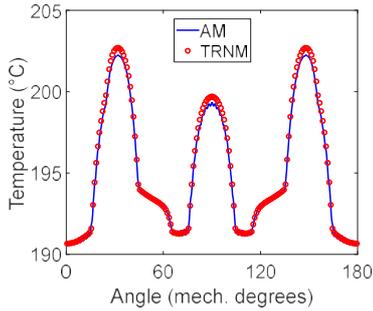


(c)
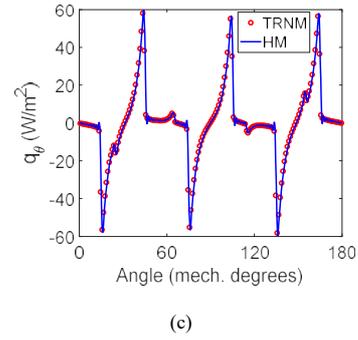
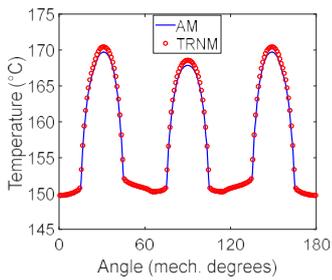**Fig. 30:** Temperature and heat flux components distribution in the middle of the air-gap.
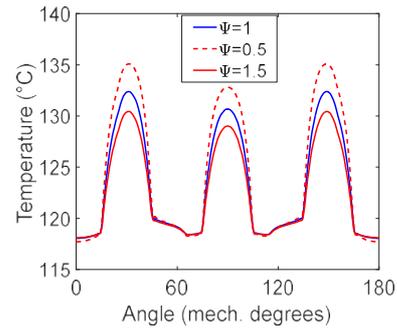


**Fig. 31:** Temperature distribution in the middle of the air-gap using HM.
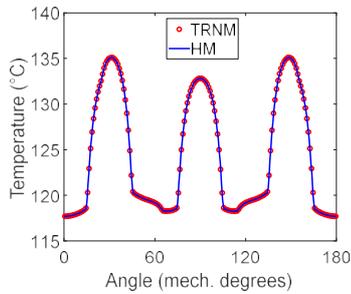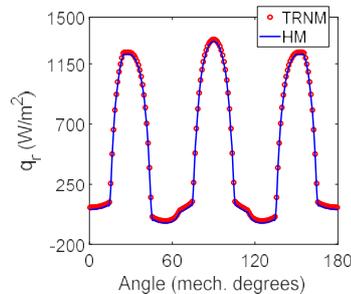


(a)
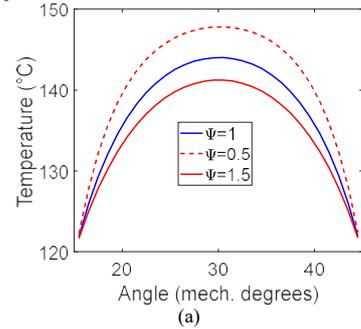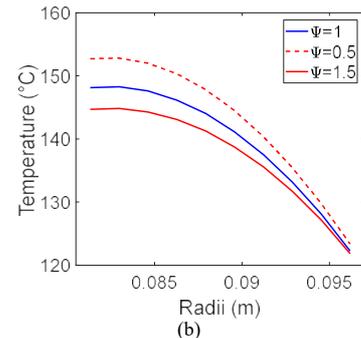


(b)

**Fig. 32:** Temperature in the middle of the first stator slot using HM.



(a)



(a)



(b)
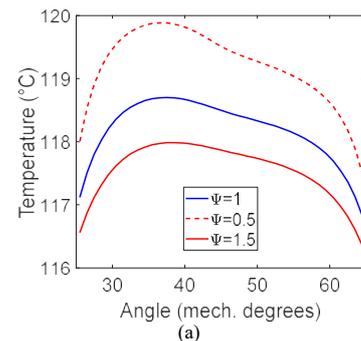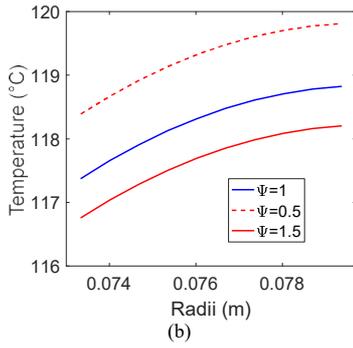
**Fig. 33:** Temperature in the middle of the first PM.

IV. CONCLUSION

The prediction of heat transfer in rotating electrical machines is usually performed using TEC and Fem. Recently, AM and TRNM have been introduced. In this paper, we have proposed an improved 2-D AM based on the exact SD technique, TRNM and HM for the prediction of steady-state temperature and heat flux components in an inset-PM machine with the materials anisotropy of thermal conductivity. The models are valid for most rotating electrical machines and allow the heat transfer in the electrical machine to be determined with very good accuracy. In the AM, the EDPs representing heat transfer in the electrical machine were solved using the separation of variables method with the thermal conductivity anisotropy in both directions (viz., $r$ and $\theta$). The implementation of TRNM in terms of mesh size, connectivity matrix, global matrix and introduction of BCs is presented where its equivalence with Fem is highlighted. As in Fem, the accuracy of TRNM depends on the adopted mesh and the number of elements.

For rotating electrical machines, AM can exhibit lower accuracy when the number of stator slots and poles is important. In this case, the number of ICs in the $r$- and $\theta$-direction is important; the dimensions of SDs and the harmonics number are small. TRNM is an alternative to AM in these regions with small dimensions and the other regions can be modeled using AM. The coupling between the two methods in HM based on the discrete Fourier series is presented and validated with Fem and TRNM. All results obtained by AM, TRNM and HM are in good agreement with each other and with those obtained by Fem.

APPENDIX A

We start with the internal elements of the conductive problem example as:

*For k from 1 to 32 do*

$$gm\left(k, isks(1,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(1,k)\theta}} ;$$

$$gm\left(k, isks(3,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(3,k)\theta}} ;$$

$$gm\left(k, isks(4,k)\right) = -\frac{1}{R_{kr} + R_{isks(4,k)r}} ;$$

$$gm\left(k, isks(5,k)\right) = -\frac{1}{R_{kr} + R_{isks(5,k)r}} ;$$

$$gm\left(k, isks(2,k)\right) = -\left(gm\left(k, isks(1,k)\right) + gm\left(k, isks(3,k)\right)..\right.$$
$$\left. .. + gm\left(k, isks(4,k)\right) + gm\left(k, isks(5,k)\right)\right);$$

*end do*

The equations of thermal convection at the rotor shaft are assembled in the global matrix $gm$ by

*For k from 33 to 40 do*

$$gm\left(k, isks(5,k)\right) = -\frac{1}{R_{kr} + R_{isks(5,k)r}} ;$$

$$gm\left(k, isks(2,k)\right) = -gm\left(k, isks(5,k)\right);$$

*end do*

The equations of thermal convection at the ambient air are assembled in the global matrix gm by

*For k from 41 to 48 do*

$$gm\left(k, isks(4,k)\right) = -\frac{1}{R_{kr} + R_{isks(4,k)r}} ;$$

$$gm\left(k, isks(2,k)\right) = -gm\left(k, isks(4,k)\right);$$

*end do*

The fixed temperature of 70 °C is introduced in the global matrix as in Fem by

*For k from 33 to 48 do*

$$gm\left(k,k\right) = gm\left(k,k\right) + 10E30 ;$$

$$f\left(k\right) = 10E30.\left(70 + 273.16\right);$$

*end do*

APPENDIX B

We start with the nodes of TRNM representing the rotor slots and teeth as

*For k from 1 to nsgr do*
  *If $k \in Bdrti$ then*

$$gm\left(k, isks(1,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(1,k)\theta}} ;$$

$$gm\left(k, isks(3,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(3,k)\theta}} ;$$

$$gm\left(k, isks(4,k)\right) = -\frac{1}{R_{kr}} ;$$

$$gm\left(k, isks(5,k)\right) = -\frac{1}{R_{kr} + R_{isks(5,k)r}} ;$$

$$gm\left(k, isks(2,k)\right) = -\left(gm\left(k, isks(1,k)\right) + gm\left(k, isks(3,k)\right)..\right.$$
$$\left. .. + gm\left(k, isks(4,k)\right) + gm\left(k, isks(5,k)\right)\right);$$

  *elif $k \in Bdrto$ then*

$$gm\left(k, isks(1,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(1,k)\theta}} ;$$

$$gm\left(k, isks(3,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(3,k)\theta}} ;$$

$$gm\left(k, isks(5,k)\right) = -\frac{1}{R_{kr}} ;$$

$$gm\left(k, isks(4,k)\right) = -\frac{1}{R_{kr} + R_{isks(4,k)r}} ;$$

$$gm\left(k, isks(2,k)\right) = -\left(gm\left(k, isks(1,k)\right) + gm\left(k, isks(3,k)\right)..\right.$$
$$\left. .. + gm\left(k, isks(4,k)\right) + gm\left(k, isks(5,k)\right)\right);$$

  *elif (k $\notin$ Bdrto) and (k $\notin$ Bdrti) then*

$$gm\left(k, isks(1,k)\right) = -\frac{1}{R_{k\theta} + R_{isks(1,k)\theta}} ;$$

$$gm(k, isks(3,k)) = -\frac{1}{R_{k\theta} + R_{isks(3,k)\theta}} \ ;$$

$$gm(k, isks(4,k)) = -\frac{1}{R_{kr} + R_{isks(4,k)r}} \ ;$$

$$gm(k, isks(5,k)) = -\frac{1}{R_{kr} + R_{isks(5,k)r}} \ ;$$

$$gm(k, isks(2,k)) = -\big(gm(k, isks(1,k)) + gm(k, isks(3,k)) ..$$
$$.. + gm(k, isks(4,k)) + gm(k, isks(5,k))\big) \ ;$$

*end if*
*end do*

The introduction of equations (62) to (64) in the global matrix gm is done as follow. (62) is introduced by

$$gm(nsgb+1, nsgb+6+9nn) = -1$$
$$gm(nsgb+1, nsgb+5+8nn) = -\ln(R_r)$$

*For j from 1 to ns do*
　　$k=nodb(1,1,j)$
　　$gm(nsgb+1,k) = 1/(2.nn)$
*end do*

$$f(nsgb+1) = -\frac{1}{4}\frac{q_r R_r^2}{\lambda_{rr}}$$

(63) gives

*For n from 1 to nn do*

$$gm(nsgb+1+n, nsgb+6+9.nn+n) = -\left(\frac{R_r}{R_i}\right)^{-n.\tau_r}$$

$$gm(nsgb+1+n, nsgb+5+8.nn+n) = -1$$

　　*For j from 1 to ns do*
　　　$k=nodb(1,1,j)$

$$gm(nsgb+1+n,k) = \frac{\cos(n.\theta_k)}{nn}$$

　　*end do*
*end do*

(64) is added to gm as
*For n from 1 to nn do*

$$gm(nsgb+1+nn+n, nsgb+6+11.nn+n) = -\left(\frac{R_r}{R_i}\right)^{-n.\tau_r}$$

$$gm(nsgb+1+nn+n, nsgb+6+10.nn+n) = -1$$

　　*For j from 1 to ns do*
　　　$k=nodb(1,1,j)$

$$gm(nsgb+1+nn+n,k) = \frac{\sin(n.\theta_k)}{nn}$$

　　*end do*
*end do*

There are 360 equations (66) to be introduced in the global matrix depending on the number of boundary nodes at $R_r$. They are given as

*For j from 1 to ns do*
　　$k=nodb(1,1,j)$

$$gm(k, isks(5,k)) = \frac{1}{R_{isks(5,k)r}}$$

$$gm(k, isks(2,k)) = -gm(k, isks(5,k))$$

$$gm(k, nsgb+5+8.nn) = -L_u \lambda_{rr}(2.d\theta)$$

$$f(k) = -\frac{1}{2}R_r^2 L_u q_r (2.d\theta)$$

*For n from 1 to nn do*

$$gm(k, nsgb+5+8.nn+n) =$$
$$-L_u \lambda_{rr}\tau_r \big(\sin(n.(d\theta-\theta_k)) + \sin(n.(d\theta+\theta_k))\big)$$

$$gm(k, nsgb+6+9.nn+n) =$$
$$L_u \lambda_{rr}\tau_r \left(\frac{R_r}{R_i}\right)^{-n.\tau_r} \big(\sin(n.(d\theta-\theta_k)) + \sin(n.(d\theta+\theta_k))\big)$$

$$gm(k, nsgb+6+10.nn+n) =$$
$$-L_u \lambda_{rr}\tau_r \big(\cos(n.(d\theta-\theta_k)) - \cos(n.(d\theta+\theta_k))\big)$$

$$gm(k, nsgb+6+11.nn+n) =$$
$$L_u \lambda_{rr}\tau_r \left(\frac{R_r}{R_i}\right)^{-n.\tau_r} \big(\cos(n.(d\theta-\theta_k)) - \cos(n.(d\theta+\theta_k))\big)$$

*end do*
*end do*

REFERENCES

[1] L. Siesing, A. Reinap, and M. Andersson, "Thermal Properties on High fill factor electrical windings: Infiltrated vs non infiltrated", *International Conference on Electrical Machines (ICEM)*, pp. 2218-2223, 2014. https://doi.org/10.1109/ICELMACH.2014.6960492).

[2] C.-C. Hwang, S.S. Wu, and Y.H. Jiang, "Novel Approach to the Solution of Temperature Distribution in the Stator of an Induction Motor", *IEEE Trans. Energy Conv.*, vol. 15, no. 4, pp. 401-406, Dec. 2000. https://doi.org/10.1109/60.900500

[3] X. Cai, M. Cheng, S. Zhu, and J. Zhang, "Thermal Modeling of Flux-Switching Permanent-Magnet Machines Considering Anisotropic Conductivity and Thermal Contact Resistance", *IEEE Trans. Ind. Elect.*, vol. 63, no. 6, pp. 3355-3365, Jun. 2016. https://doi.org/10.1109/TIE.2016.2522942

[4] Q. Chen, D. Liang, L. Gao, Q. Wang, and Y. Liu, "Hierarchical thermal network analysis of axial-flux permanent-magnet synchronous machine for electric motorcycle", *IET Electric Power Applications*, vol. 12, no. 12, pp. 859-866, 2018. https://doi.org/10.1049/iet-epa.2017.0719

[5] Y. Yang, B. Bilgin, M. Kasprzak, S. Nalakath, H. Sadek, M. Preindl, J. Cotton, N. Schofield, and A. Emadi, "Thermal Management of Electric Machines", *IET Electrical Systems in Transportation*, vol. 7, no. 2, pp. 104-116, 2017. https://doi.org/10.1049/iet-est.2015.0050

[6] R. Wrobel, P. Mellor, and D. Holliday, "Thermal Analysis of a Segmented Stator Winding Design", *IEEE Trans. Ind. Appl.*, vol. 47, no. 5, pp. 2023-2030, Sep./Oct. 2011. https://doi.org/10.1109/ECCE.2010.5617814

[7] K. Boughrara, F. Dubas, and R. Ibtiouen, "2-D Exact Analytical Method for Steady-State Heat Transfer Prediction in Rotating Electrical Machines", *IEEE Trans. Magn.*, vol. 54, no. 9, pp. 1-19, Sep. 2018, Art. ID 8104519. https://doi.org/10.1109/TMAG.2018.2851212

[8] F. Dubas, and K. Boughrara, "New scientific contribution on the 2-D subdomain technique in Cartesian coordinates: Taking into account of iron parts," *Math. Comput. Appl.*, vol. 22, no. 1, p. 17, Feb. 2017. https://doi.org/ 10.3390/mca22010017

[9] F. Dubas, and K. Boughrara, "New scientific contribution on the 2-D subdomain technique in polar coordinates: Taking into account of iron parts," *Math. Comput. Appl.*, vol. 22, no. 4, p. 42, Oct. 2017 https://doi.org/ 10.3390/mca22040042

[10] H. Ennassiri, G. Barakat, and Y. Amara, "Steady state hybrid thermal modeling of permanent magnet electrical machines," *International Conference on Ecological Vehicles and Renewable Energies (EVER)*, pp. 1-6, 2016. https://doi.org/10.1109/EVER.2016.7476420.

[11] S. Ouaged, M.A. Ben Hamida, Y. Amara, G. Barakat, and J.J.H. Paulides, "Thermal modelling of Tubular Linear Machines using a Hybrid Analytical Method," *International Conference on Sustainable Mobility Applications, Renewables and Technology (SMART)*, pp. 1-5, 2015. https://doi.org/10.1109/SMART.2015.7399242

[12] D. Jackson, "On the Order of Magnitude of the Coefficients in Trigonometric Interpolation", *Transactions of the American*

*Mathematical Society*, 21, 1920, pp. 321-332, https://doi.org/10.1090/S0002-9947-1920-1501147-2

[13] D.C. Meeker, "Rotating Losses in a Surface Mount Permanent Magnet Motor", October 2017, http://www.femm.info/wiki/SPMLoss.

[14] K-H. Shin, K. Hong, H-W. Cho, and J-Y. Choi, "Core Loss Calculation of Permanent Magnet Machines Using Analytical Method", *IEEE Trans. on Applied Superconductivity*, vol. 28, no. 3, pp. 1-5, April 2018, Art. ID 5205005. https://doi.org/10.1109/TASC.2018.2800706

[15] Sh. Utegenova, F. Dubas, M. Jamot, R. Glises, B. Truffart, D. Mariotto, P. Lagonotte, and Ph. Desevaux, "An Investigation Into the Coupling of Magnetic and Thermal Analysis for a Wound-Rotor Synchronous Machine," *IEEE Trans. Ind. Elect.*, vol. 65, no. 4, pp. 3406-3416, Apr. 2018. https://doi.org/10.1109/TIE.2017.2756597

[16] Z. Djelloul-khedda, K. Boughrara, F. Dubas, A. Kechroud, and A. Tikellaline, "Analytical Prediction of Iron Core Losses in Flux-Modulated Permanent-Magnet Synchronous Machines", *IEEE Trans. Magn.*, 2018. https://doi.org/ 10.1109/TMAG.2018.2877164

[17] D.C. Meeker. (Apr. 1, 2009). Finite Element Method Magnetics ver. 4.2. [Online]. Available: http://www.femm.info.

[18] E. Divo and A. Kassab, "On a Boundary-Only BEM for Heat Conduction in Non-Homogeneous Anisotropic Media", Transactions on Modeling and Simulation, vol. 15, pp. 279-288, WIT Press,1997.

[19] Flux2D, General Operating Instructions, Version 2017; Altair.

[20] G. Barakat, and Y. Amara, "A simple and effective way to couple analytical formal solution of magnetic potential and reluctance network models", *International Conference on Computation in Electromagnetics (CEM)*, pp. 1-2, 2014, https://doi.org/10.1049/cp.2014.0199

[21] K.J.W. Pluk, J.W. Jansen, and A. Lomonova, "Hybrid Analytical Modeling: Fourier Modeling Combined With Mesh-Based Magnetic Equivalent Circuits", *IEEE Trans. Magn.*, vol. 51, no. 8, pp. 1-12, Aug. 2015, Art. ID 8106812. https://doi.org/10.1109/TMAG.2015.2419197

[22] M. Johnson, M.C. Gardner, and H.A. Toliyat, "A Parameterized Linear Magnetic Equivalent Circuit for Analysis and Design of Radial Flux Magnetic Gears-Part I: Implementation", *IEEE Trans. Energy Conv.*, vol. 33, no. 2, pp. 784-791, June 2018. https://doi.org/10.1109/TEC.2017.2777875

[23] M. Johnson, M.C. Gardner, and H.A. Toliyat, "A Parameterized Linear Magnetic Equivalent Circuit for Analysis and Design of Radial Flux Magnetic Gears-Part II: Evaluation", *IEEE Trans. Energy Conv.*, vol. 33, no. 2, pp. 792-800, June 2018. https://doi.org/10.1109/TEC.2017.2777876

**Kamel Boughrara** was born in Algiers, Algeria, in 1969. He received the Engineer Diploma degree from Ecole Nationale Polytechnique (ENP), Algiers, in 1994, the Magister degree from the University of Sciences and Technology Houari Boumediene, Algiers, in 1997, and the Ph.D. degree from ENP, in 2008. He is currently a Professor with ENP, where he is also the head of department of Electrical Engineering.

His current research interests include the modeling and control of electrical machines.

**Frédéric Dubas** was born in Vesoul, France, in 1978. He received the M.Sc. degree and the Ph.D. degree from the "Univ. Bourgogne Franche-Comté" (Besançon, France) in 2002 and 2006, respectively, with a focus on the design and the optimization of high-speed surface-mounted permanent-magnet (PM) synchronous motor for the drive of a fuel cell air-compressor.

From 2014 to 2016, he has been the Head of "Unconventional Thermal and Electrical Machines" Team. He is the Head of the "Electrical Actuators" group in the "Hybrid & Fuel Cell Systems, Electrical Machines (SHARPAC)" Team. He works with ALSTOM Transports (Ornans, France), and RENAULT Technocenter (Guyancourt, France), where he is involved in the modelling, design and optimization of electrical systems and, in particular, induction and PM synchronous (radial and/or axial flux) machines, creative problem solving, and electrical propulsion/traction. He is currently an Associate Professor with the Dép. ENERGIE, FEMTO-ST Institute affiliated to the CNRS and jointly with the "Univ. Bourgogne Franche-Comté" (Besançon, France).

He has authored over 100 refereed publications and he holds a patent about the manufacturing of axial-flux PM machines with flux-focusing. Dr. Dubas received: i) the Prize Paper Awards in the IEEE Conference Vehicle Power and Propulsion (VPPC) in 2005, ii) the Prize Presentation Awards in the 19th International Conference on Electrical Machines and Systems (ICEMS) in 2017, iii) the RENAULT Internal Award (Direction Engineering Alliance – Innovation) in 2019.

# Fullers Earth Treatment for Esters Liquids used in Power Apparatuses: Inferences and Arguments

Issouf Fofana, Yohan Bergeron, Marie-Pier Gagnon, Jonathan Tremblay, Luc Loiselle, Kouba Marie Lucia Yapi,

and Ungarala Mohan Rao

**Abstract—Insulating Liquids are widely used for their electrical and thermal properties in power apparatuses, particularly at the level of liquid-filled transformers. With the shift in engineering aspects towards sustainable development, it is important to find a sustainable solution with ecofriendly nature. Therefore, alternative (biodegradable) liquids are of high importance in the global transformer communities. In the present study, the alternative dielectric fluids (ester-based) feasibility for potential regeneration with Fuller's earth is investigated. The experimental results are confined to the reclamation temperature as well as the ratio of Fuller's earth (the sorbent) and the liquid. A suitable laboratory treatment apparatus is designed and is adopted in this study. Promising measurements to comment on the effectiveness of the treatment have been performed at controlled treatment temperature and sorbent-liquid ratio with the ASTM 7150-13 as a reference norm. The results of this study allowed 80°C and 1 g/30 ml as affirmative conditions for the present experimental conditions. Diagnostic measurements include turbidity, particle counter, and UV spectrophotometry before and after treatments. It is inferred that fuller's earth is not a promising sorbent for the reclamation of ester liquids.**

**Keywords—Transformer, Dielectric Liquid, Esters, Regeneration.**

## I. INTRODUCTION

The electric power grid has evolved greatly as technologies have refined with advancements in digital technologies and high-power demands. Power transformers connected across the grid allow the voltage levels of power lines to be raised and lowered, which saves a lot in terms of loss of electrical power, thus saving a lot of money [1]. However, the power transformer is one of the extremely expensive equipment connected over the grid. For this reason, companies, factories, and owners of electrical networks do not hesitate to spend many thousands of dollars on transformer maintenance and devices to protect its health integrity. Given that insulation systems in the transformers play a critical role in deciding the health index and service life of a transformer, high emphasis is laid on the transformer's insulation system [2]. In addition to protecting these important machines, the utility engineers need to have the condition of transformers assessed in order to prevent malfunction and avoid premature aging [3]. Different stresses including electrical, thermal, and chemical, accelerate the degradation of liquid and solid insulation in the transformers [4]. Different stress has a different effect on the nature of degradation, and this is attributable to the type of insulating material. Since the beginning of transformer technology,

Issouf Fofana, Luc Loiselle, Ungarala Mohan Rao and Kouba Marie Lucia Yapi are with the Department of applied sciences, University of Quebec at Chicoutimi, CANADA.
(e-mail: ifofana@uqac.ca; luc.loiselle@uqac.ca; mohan.ungarala1@uqac.ca; KoubaMarieLucia1_Yapi@uqac.ca)
Yohan Bergeron, Marie-Pier Gagnon, and Jonathan Tremblay were with the Department of applied sciences, University of Quebec at Chicoutimi, CANADA. (e-mail: yohan.bergeron1@uqac.ca; marie-pier.gagnon3@uqac.ca; jonathan.tremblay14@uqac.ca )

mineral oils with solid cellulose insulations are widely accepted. However, mineral oils are facing critiques due to their toxicity, low flashpoints as well as limited dielectric properties. The details of the same are elaborated by major authors [5, 6].

Therefore, alternative and biodegradable liquids, both natural and synthetic esters, are reported affirmatively for possible use in liquid-filled transformers [6-9]. The service behavior [10], paper degradation [11, 12], pre-breakdown behavior [13], liquid degradation [11, 14, 15] are reported affirmatively by various authors. In addition, the gassing tendency and influence of various faults on the gassing and definite liquid degradations have been widely reported by authors [16]. Various authors in different works mentioned above also reported the physiochemical behavior of the ester liquids. It is inferred that esters are hydrophilic at the outset and therefore exhibit a higher water saturation limit, improve hydrolysis rates and thus keep the cellulose insulation dry. This in turn improves the rate of degradation of cellulose in esters and reduces decay concentration in the liquids. Despite few limitations, phenomenal service and technical benefits outweigh these new liquids for high-voltage devices. However, there is a need to understand these new liquids' end-life behavior in order to rejuvenate, recycle or reuse phenomena. The reclamation aspects of mineral insulating liquids used in transformers and liquid-filled electrical apparatus is reported in the IEEE standard C57. 637 [2]. There is no established standard for the reclamation of ester liquids [2]. Nevertheless, various authors reported the regeneration of esters with various adsorbents. The details of the adsorbents are not well discussed in the existing literature. The factory dehydration and degassing process remain the same; the treatment temperature should be around 70°C. High temperatures will influence the oxidation stability of the ester liquids and definitely have a notable impact on the weight of the available antioxidants in the bulk of the liquid [2]. To the best of the author's knowledge, this treatment temperature is not yet investigated by researchers or reported in the existing literature. Thus, this work aims to investigate the feasibility of fuller's earth for regeneration of esters before the

new adsorbents or absorbents are investigated.

In the authors' recent work [17], the ability and influence of ester to generate the decay particles viz. soluble particles and colloidal particles have been investigated for sludge monitoring in these new liquids. It is found that esters generate less sludge or colloids and have high scope for the generation of soluble decay particles. However, it is an argument that ester dissolves sludge after a generation or does not generate any sludge with degradation. Therefore, in this work, the potential of the widely accepted sorbent for transformer oils that is Fuller's earth has been identified as the subject. A large number of papers have reported the reclaiming of damaged insulating mineral oils. The processes used were re-refining with chemicals [18] activated alumina or Fuller's earth [19].The performance of Fuller's earth has been verified under selected conditions adopted for the treatment of the ester liquids in comparison to mineral oils. The laboratory-aged liquids are filtered using Fuller's earth at different treatment temperatures and different sorbent-liquid ratios to understand the effectiveness of the treatment. It is inferred that Fuller's earth is not a potential sorbent for treating esters and the filtration at 80°C is better than the other pre-treatment temperatures. Also, the 1 g/30 ml ratio is found to be effective for a single pass analysis. However, the effectiveness is claimed based on few important diagnostic measurements. Further, detailed analysis and more quantitative measurements need to be performed to open the door for research on alternative sorbents for ester liquids.

## II. EXPERIMENTAL

The present investigation is performed on three different insulating liquids with an aim to understand the potential of fuller's earth for possible regeneration of aged ester liquids. However, the emphasis is laid on the pretreatment temperature and ratio of the fuller's earth and liquid. Therefore, the conditions outlined in the ASTM D 7150 are considered as a reference for the present study. The insulating liquids are aged, and the diagnostic characteristics of these aged liquids are considered a benchmark for the treated liquids. Following the reclamation of the aged liquids at different aging factors, the diagnostic characteristics are compared at different reclamation conditions.

### A. Materials and Apparatus

The insulating liquid mineral oil (MO), natural ester (NE), and synthetic ester (SE) are aged as per ASTM D1934 under open beaker aging conditions in the presence of cellulose (1:20) and copper catalyst (3g/l). Thermal aging is performed at 115°C with an aging history at 500 hours, 1000 hours, and 1500 hours. These high-aging durations under open beaker conditions allowed the insulating materials to degrade to a greater extent and introduce the aging by-products into the liquids. As discussed earlier, treatment temperature and sorbent-liquid ratios are the prime focus for the present study. Thus, a laboratory-based setup is developed, which controls the treatment temperature and sorbent volume in the reclamation column. In order to have good temperature control, it is imperative to use a well-distributed heating system around our reclamation compartment. In this context, a heating element is wrapped around fuller's earth chamber to achieve the desired temperature where the liquid is reclaimed. The heating element is shown in Figure 1(a).This heating element is wrapped around the glass syringes in which a bed of cotton is placed and above this the Fuller's earth is hosted. Two parallel inlets for the liquid

are planned to save reclamation time. The view of these reclamation tubes is shown in Figure 1(b). The complete view of the experimental along with the supply and temperature control unit is presented in Figure 1.



(a)                              (b)



1. Oil Samples
2. Oil in reclaiming units
3. Heating element

4. Fullers earth
5. reclaimed oil
6. Variac or voltage regulator
7. Temperature controller unit

**(c)**
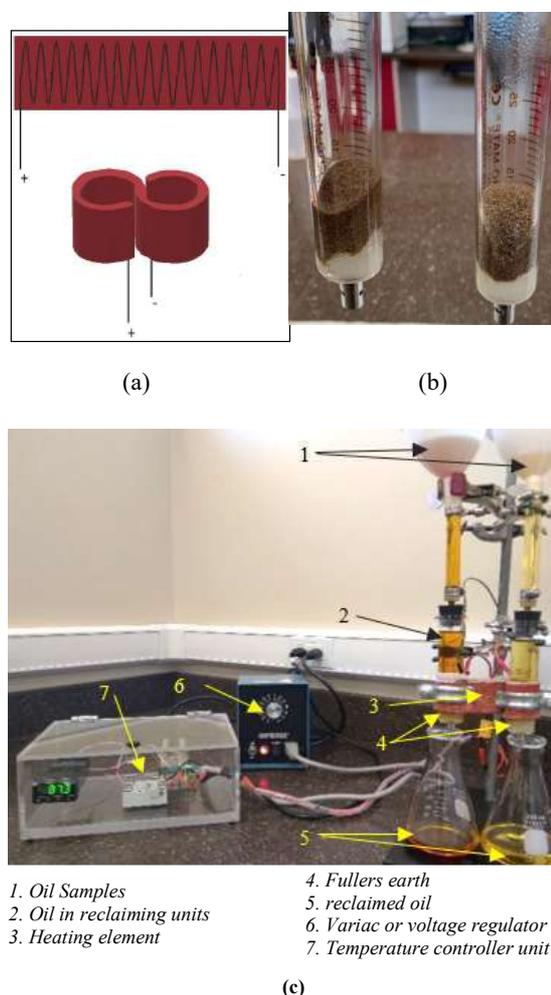
**Fig. 1:** View of the laboratory model for Fuller's earth reclamation.

### B. Measurements

The aged liquids are tested for turbidity, UV spectroscopy, and particles counter to understand the level of degradation and degree of contamination of the liquids due to aging. Besides, the aged liquids are reclaimed at sorbent-liquid ratios of 1 g/30 ml, 1 g/35 ml, and 1 g/40 ml to have control of the ratio. All the three ratios are tested at a treatment temperature of 70°C, 80°C, and 90°C to understand the influence of treatment temperature. The insulating liquids are heated to the set treatment temperature before passing through the heated reclaiming tubes which are maintained at the desired treatment temperature. This ensures that the reclamation is performed when the liquid and the fuller's earth are both held at the same treatment temperature. The reclaimed fluid is collected and tested for turbidity, UV spectroscopy, and particles counter to understand the level of reclamation. The results and discussion section will focus on the comparative discussion of the results to understand the influence of temperature and sorbent ratio at different aging conditions.

## III. RESULTS AND DISCUSSION

### A. Influence of Treatment Temperature

The treatment temperature is a vital parameter for the insulating liquid regeneration process. It is a major parameter that could decide the efficiency of the sorbent and the effectiveness of the treatment process. The sorption process, either adsorption or absorption that undergoes during the treatment, is attributable to the sorbent's surface properties. The surface properties of any sorbent are temperature-dependent and vary based on the nature of the decay products in the feed liquid. In particular, the sorption efficiency is high when the process is carried while the sorbent is activated. Thus highlighting, the temperature called activation temperature for the sorbent. However, this activation temperature should be in agreement with the oxidation stability of the feed liquid. Therefore, a careful evaluation of pretreatment temperatures is required for new insulating liquids and sorbents. In this section, the change in turbidity and dissolved decay products of the liquids at different aging factors treated at different treatment temperatures have been discussed.



**(a)** Change in the turbidity of various liquids reclaimed at 1 g/35 ml



**(b)** Change in the dissolved decay products of various liquids reclaimed at 1 g/35 ml
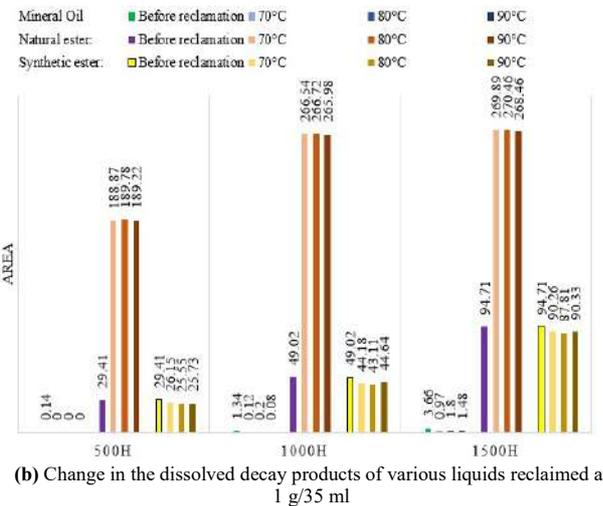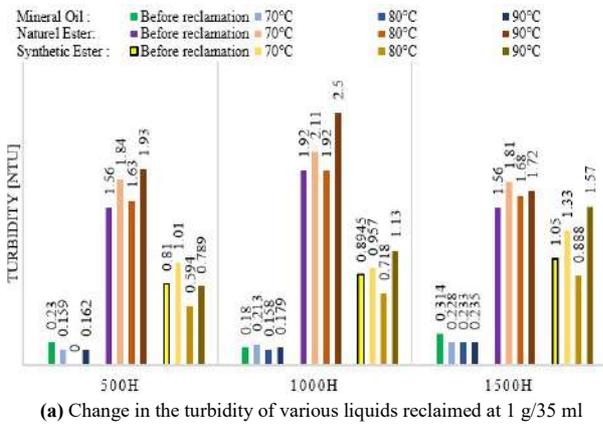
**Fig. 2.** Changes in the liquid properties before and after reclamation at different temperatures.

The change in turbidity of the liquids measured after reclamation at 70°C, 80°C, and 90°C are presented in Figure 2(a). It is seen that the turbidity values increase with normal aging. However, the turbidity is not noticed to be reduced with reclamation in liquids, especially for the ester-based liquids. It is noticed that the increase of turbidity is slightly less at 80°C of treatment temperature as compared to 70°C and 90°C. The dissolved decay contents are also noticed not to have any significant change with the fuller's earth reclamation in case of

ester liquids. Also, the selected range of treatment temperatures have the least influence on the reclamation of ester liquids. However, the reclamation of esters with Fuller's earth is found unsatisfactory for the present experimental conditions.

### B. Influence of Liquid-Sorbent Ratio

Similar to treatment temperature, the sorption ratio also has a significant influence on the treatment process. This is because the sorption process, either absorption or adsorption (starts with absorption phenomena) is controlled by the surface properties of the sorbent, as discussed in the previous sections. At this level, it is important to add the property called sorption saturation. This indicates that the sorbent loses its ability to absorb or adsorb any further particles from the feed after this saturation limit. Thus, an optimal sorbent-liquid ratio is essential to be obeyed for the reclamation of different liquids. Unfortunately, there is no published data on the knowledge of these ratios in the case of ester liquids. In other words, a sorbent will have different saturation limits with different insulating liquids. The reclamation or sorption process is efficient when the reclamation is performed below the sorption saturation limit. To comment on the ratios, liquids at different aging conditions have been reclaimed at different Fuller's earth-liquid ratios. The change in turbidity and dissolved decay products with reclamation at different ratios performed at 80°C are shown in Figure 3.



**(a)** Change in turbidity of various liquids reclaimed at 80°C



**(b)** Change in dissolved decay products of various liquids reclaimed at 80°C
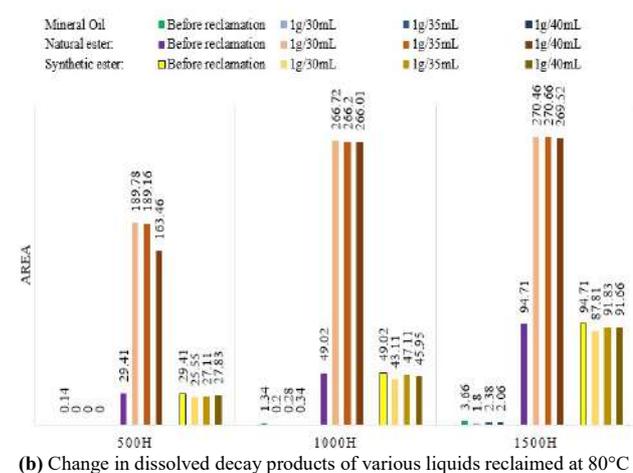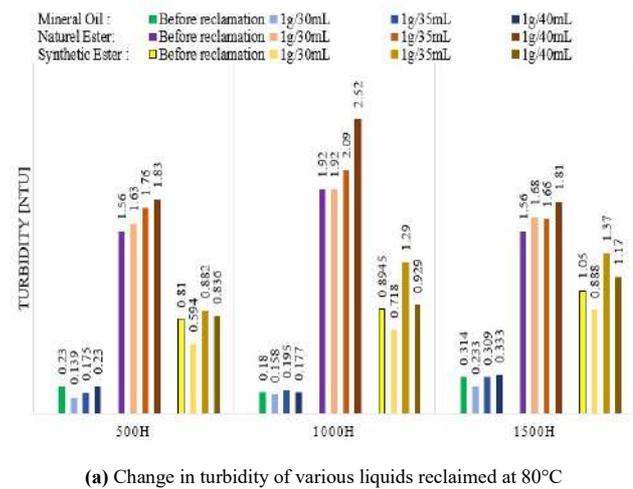
**Fig. 3.** Changes in the liquid properties before and after reclamation at different sorption-liquid ratios.

The change in turbidity with different sorbent-liquid ratios has been presented in Figure 3(a). It is observed that turbidity is

increasing with an increase in the feed liquid volume in the case of natural esters, and is randomly changing in the case of synthetic esters. However, 1 g/30 ml is found to be better in the present range of ratios in the case of synthetic esters. The change in dissolved decay products has been presented in Figure 3(b) for liquids reclaimed at different ratios were no significant changes are noticed with a change in the sorbent-liquid ratio. To further understand, the same particle count measurements are performed on all the liquids reclaimed at different ratios and are presented in Figure 4 for mineral oils, synthetic esters, and natural esters.
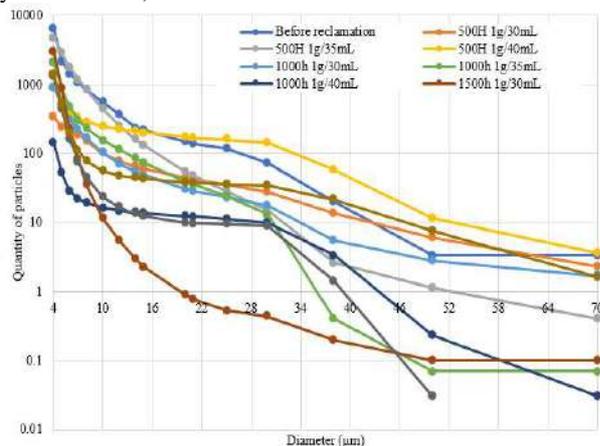


**(a)** Change in particle count of mineral oils reclaimed at 80°C
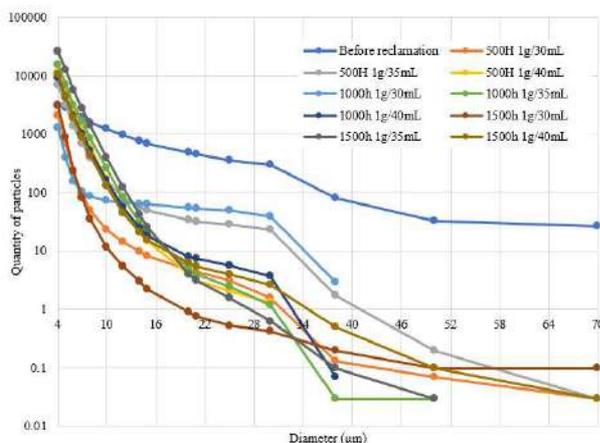


**(b)** Change in particle count of synthetic esters reclaimed at 80°C
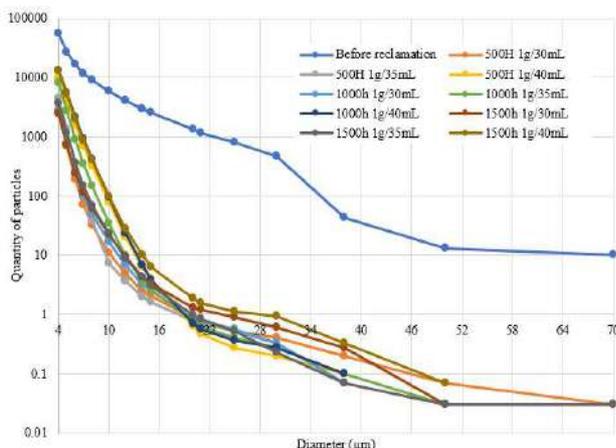


**(c)** Change in particle count of natural esters reclaimed at 80°C

**Fig. 4**. Changes in the particle count in the liquids before and after reclamation at different sorption-liquid ratios.

The change in total particles in mineral oil, synthetic esters, and natural esters with reclamation has been reduced. However, a reduction of higher diameter particles is noted with reclamation. This may be assumed that the decay particles having higher diameters have been reclaimed out. However, the increase in the smaller diameter particles with reclamation is questionable and needs further investigations.

## IV. CONCLUSION

In this work, mineral oil, synthetic ester, and natural ester reclamation have been performed using Fuller's earth at different temperatures and sorbent-liquid ratios. The aged liquids have been characterized for turbidity, UV spectroscopy (dissolved decay products), and particle count. Based on the present measurements, the reclamation performed at 1 g / 30 ml as well as heated to a temperature of 80°C is a compromising option. While Fuller's Earth is used worldwide for mineral reclamation, its application to ester liquids is questionable. There is a need to further analyze viscosity, acidity, FTIR, and density changes with filtration. In addition, the flow rate, the shape of the reclamation column (that is, a very narrow, less narrow, and a wide syringe), and the number of reclamation cycles are also influential factors of major importance.

### REFERENCES

[1] G. C. Montanari, R. Hebner, P. Morshuis, and P. Seri, "An Approach to Insulation Condition Monitoring and Life Assessment in Emerging Electrical Environments," IEEE Transactions on Power Delivery, vol. 34, no. 4, pp. 1357-1364, 2019, doi: 10.1109/tpwrd.2019.2897905.

[2] U. Mohan Rao, I. Fofana, and J. S. N'Cho, "On Some Imperative IEEE Standards for Usage of Natural Ester Liquids in Transformers," IEEE Access, vol. 8, pp. 145446-145456, 2020, doi: 10.1109/access.2020.3014600.

[3] V. M. Catterson et al., "The impact of smart grid technology on dielectrics and electrical insulation," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 22, no. 6, pp. 3505-3512, 2015, doi: 10.1109/tdei.2015.00518.

[4] I. Fofana, A. Bouaicha, M. Farzaneh and J. Sabau, "Ageing Behaviour of Mineral Oil and Ester Liquids: a Comparative Study," 2008 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, 2008, pp. 87-90.

[5] I. Fofana, "50 years in the development of insulating liquids," IEEE Electrical Insulation Magazine, vol. 29, no. 5, pp. 15-25, 2013.

[6] U. Mohan Rao, I. Fofana, T. Jaya, E. M. Rodriguez-Celis, J. Jalbert, and P. Picher, "Alternative Dielectric Fluids for Transformer Insulation System: Progress, Challenges, and Future Prospects," IEEE Access, vol. 7, pp. 184552-184571, 2019, doi: 10.1109/access.2019.2960020.

[7] M. Rafiq et al., "Use of vegetable oils as transformer oils – a review," Renewable and Sustainable Energy Reviews, vol. 52, pp. 308-324, 2015, doi: 10.1016/j.rser.2015.07.032.

[8] D. M. Mehta, P. Kundu, A. Chowdhury, V. K. Lakhiani, and A. S. Jhala, "A review on critical evaluation of natural ester vis-a-vis mineral oil insulating liquid for use in transformers: Part 1," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 2, pp. 873-880, 2016, doi: 10.1109/tdei.2015.005370.

[9] D. M. Mehta, P. Kundu, A. Chowdhury, V. K. Lakhiani, and A. S. Jhala, "A review of critical evaluation of natural ester vis-a-vis mineral oil insulating liquid for use in transformers: Part II," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 3, pp. 1705-1712, 2016, doi: 10.1109/tdei.2016.005371.

[10] D. Martin, T. Saha, and L. McPherson, "Condition monitoring of vegetable oil insulation in in-service power transformers: some data spanning 10 years," IEEE Electrical Insulation Magazine, vol. 33, no. 2, pp. 44-51, 2017.

[11] I. Fernández et al., "Thermal degradation assessment of Kraft paper in power transformers insulated with natural esters," Applied Thermal Engineering, vol. 104, pp. 129-138, 2016, doi: 10.1016/j.applthermaleng.2016.05.020.

[12] U. Mohan Rao and R. K. Jarial, "Measurement of transformer solid insulation degradation using dilatometry and X-ray diffraction analysis," Measurement, vol. 131, pp. 701-705, 2019, doi: 10.1016/j.measurement.2018.09.024.

[13] U. Mohan Rao et al., "A review on pre-breakdown phenomena in ester fluids: Prepared by the international study group of IEEE DEIS liquid

dielectrics technical committee," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 5, pp. 1546-1560, 2020, doi: 10.1109/tdei.2020.008765.

[14]  U. Mohan Rao, Y. N. Kumar, and R. K. Jarial, "Understanding the ageing behaviour of transformer oil–paper insulation with ester and mixed dielectric fluids," IET Science, Measurement & Technology, vol. 12, no. 7, pp. 851-857, 2018, doi: 10.1049/iet-smt.2018.0110.

[15]  I. Fernández, A. Ortiz, F. Delgado, C. Renedo, and S. Pérez, "Comparative evaluation of alternative fluids for power transformers," Electric Power Systems Research, vol. 98, pp. 58-69, 2013, doi: 10.1016/j.epsr.2013.01.007.

[16]  L. Loiselle, U. Mohan Rao, and I. Fofana, "Gassing Tendency of Fresh and Aged Mineral Oil and Ester Fluids under Electrical and Thermal Fault Conditions," Energies, vol. 13, no. 13, 2020, doi: 10.3390/en13133472.

[17]  L. Loiselle, U. Mohan Rao, I. Fofana, and T. Jaya, "Monitoring colloidal and dissolved decay particles in ester dielectric fluids," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 5, pp. 1516-1524, 2020. doi: 10.1109/TDEI.2020.008719

[18]  W.G. Barlow, "Reclaiming of Transil Oil by a Sulphuric Acid Process". Minutes of the Eleventh Annual International Conference of Doble Clients Boston. sec. 10-201, 1944.

[19]  J. S. N'cho, I. Fofana, A. Beroual, T. Aka-Ngnui and J. Sabau, ''Aged oils reclamation: Facts and arguments based on laboratory studies''. IEEE Transactions on Dielectrics and Electrical Insulation, vol. 19, 1583-1592, 2012. doi: 10.1109/TDEI.2012.6311504

**Issouf Fofana** obtained his electro-mechanical engineering degree in 1991 from the University of Abidjan (Côte d'Ivoire), and his master's and doctoral degrees from École Centrale de Lyon, France, in 1993 and 1996, respectively. He was a postdoctoral researcher in Lyon in 1997 and was at the Schering Institute of High Voltage Engineering Techniques at the University of Hanover, Germany from 1998 to 2000. He was a Fellow of the Alexander von Humboldt Stiftung from November 1997 to August 1999. He joined University of Quebec at Chicoutimi (UQAC), Quebec, Canada as an Associate Researcher in 2000, and he is now a professor there. Dr. Fofana has held the Canada Research Chair, tier 2, of insulating liquids and mixed dielectrics for electrotechnology (ISOLIME) from 2005 to 2015. Actually, he is leading the Research Chair on the Aging of Power Network Infrastructure (ViAHT). He is also director of the MODELE laboratory and the International Research Centre on Atmospheric Icing and Power Network Engineering (CenGivre) at UQAC. He is the chair of the IEEE DEIS Technical committee on Dielectric Liquids. He has authored/co-authored over 300 scientific publications, three book chapters, one textbook, edited three books and holds three patents. Dr Fofana is an IEEE Senior Member and IET Fellow.

**Yohan Bergeron** graduated in 2019 at the University of Quebec at Chicoutimi (UQAC), Quebec, Canada. So far, he is working with WSP Canada Saguenay, Québec, Canada as consultant. His main interests include renewable energy, energy management and power systems analysis.

**Marie-Pier Gagnon** graduated in 2019 at the University of Quebec at Chicoutimi (UQAC), Quebec, Canada. So far, he is working with Voltam Inc. Saguenay, Québec as professional engineer. Her main interests include power systems analysis and protection.

**Jonathan Tremblay** was a graduate student in the department of applied sciences at the University of Quebec at Chicoutimi (UQAC), Canada. During his finishing school project, he worked with the Research Chair on the Aging of Power Network Infrastructure (ViAHT) for four months.

**Luc Loiselle** received his Bachelor's degree in Electrical Engineering in 2006, his Master's degree in 2013 and his PhD in 2021 from the University of Quebec at Chicoutimi (UQAC), Canada. He is registered professional engineer in the province of Quebec (Ordre des Ingénieurs du Québec, Canada. From 2006-2016, he was a consultant in engineering at Centre des Services aux Enterprises (Saguenay, Canada). Since 2018, he is working as electrical engineer with Tetra Tech, Saguenay (Canada). His main research interests include high-voltage, electrical insulation, dielectric materials and safety. He has coauthored/authored dozen of scientific publications.

**Kouba Marie Lucia Yapi** graduated in Engineering in 2012 from the University of Quebec at Chicoutimi (UQAC), Canada. Since 2013, she is working as research professional at the UQAC (from 2013-2015: with the Canada Research Chair on insulating liquids and mixed dielectrics for electrotechnology: ISOLIME, and since 2013 with the Research Chair on the Aging of Power Network Infrastructure: ViAHT, UQAC). She is registered professional junior engineer in the province of Quebec (Ordre des Ingénieurs du Québec, Canada). Her main research interests focus on power transformer diagnostics.

**Ungarala Mohan Rao** obtained his bachelor's degree in electrical and electronics engineering from Jawaharlal Nehru Technological University, Kakinada, India in 2010. He obtained his master's and doctoral degrees from the National Institute of Technology (NIT), Hamirpur, India, in 2012 and 2017 respectively. At present, he is a lecturer in the Department of Applied Sciences (DSA) at University of Quebec at Chicoutimi (UQAC), Québec, Canada. Since 2018, he is also a postdoctoral researcher at UQAC with the Research Chair on the Aging of Power Network Infrastructure (ViAHT). Dr. Mohan is a Senior Member of IEEE and Member of the IEEE DEIS. He is also the Secretary for the IEEE Technical Committee on "Liquid Dielectrics". His main research interests include aging phenomena of high-voltage insulation, condition monitoring of electrical apparatus, alternative dielectric materials, transformer insulation in cold countries, and AIML applications.

# Vibration Signal Parameter Estimation in Variable Speed: Algorithms and performance bounds

Ali Abbadi, Cécile Capdessus, Karim Abed-Meraim, and Edgard Sekko

**Abstract—Vibration signal parameter estimation for rotating machinery diagnostics operating under variable speed conditions is considered. At first, we provide a brief survey of existing methods for Quadratic Phase Signal (QPS) parameter estimation. Then, we introduce improved solutions for the general QPS case and the Order QPS (O-QPS) case, respectively. For all considered cases (namely the QPS, O-QPS with tachometer and O-QPS without tachometer), we develop the Cramer Rao Bounds to assess and compare the estimation performance limits for each model. Finally, we compare the performance of all considered methods and highlight, in particular, the gain of the proposed solutions.**

*Keywords*—**Cramer-Rao Bound (CRB), Higher order Ambiguity Function (HAF), Quadratic Phase Transform (QPT), Variable Speed, Vibration Signal.**

## I. INTRODUCTION

In the last decades, vibration signal analysis techniques have attracted increasing attention and became a hot research topic in the rotating machinery diagnostics area thanks to its potential advantage over invasive diagnostics techniques.

It is based on the recording, via accelerometers or other modern technology sensors (e.g. high-speed laser sensors), of the vibration levels and frequencies of the machinery of interest and then using that signal to analyze how healthy the machine's components are. The vibration signal analysis is the realm of developing appropriate models, methods and algorithms which help determining the health of the machine and identify possible impending problems like unbalance, misalignment, looseness, lubrication issues, etc. Indeed, based on a physical model of the vibration signal, and taking into account shocks, friction, rotation speed, structural resonance and propagation, relevant indicators can be elaborated to monitor rotating machinery. Such a non invasive analysis has been shown to be very efficient to detect problems such as [1, 2]: Bearing failures, Imbalance, Resonance and natural frequencies, Gearbox failures, Electrical motor faults, etc.

Now, the introduction of artificial intelligence techniques for automating the diagnosis implies the development of appropriate methods that can be implemented in all operating conditions, even in the cases where classical approaches fail, such as in time varying operating conditions. In that case, unlike the conventional diagnostic approach based on stationary assumption, the non stationary assumptions open a way to the diagnosis of the rotating machinery under variable speed conditions.

Based on time varying signal analysis tools for rotating machinery diagnosis under variable speed operation, three classes of order tracking have been proposed and broadly discussed in the literature [3–6], which are the resampling based methods, the Kalman filter based methods (tracking methods) and the transform based methods. The resampling based methods transform the signal data from the time domain to the angular (order) domain to obtain a sampled signal at constant angular increment instead of a sampled signal at constant time increment. They result in limited order resolution related to the constant increment interval and a finite number of samples. In order to improve the accuracy of order tracking the Kalman filter based methods were introduced. They consist in recursively minimizing a cost function for parameter estimation of time-varying frequencies to overcome many of the limitations of order resolution and allow the optimal estimation of the amplitude and phase of an order. Unfortunately, they are computationally expensive and require prior knowledge of some information such as the number of orders (frequencies) to be extracted, needed for Kalman filter initialization and convergence. These disadvantages limit the use of these methods to experienced operators and in offline conditions contrary to the robust and rather easy-to use transform methods. The transform methods are based on the modification of the DFT (Discrete Fourier Transform) kernel to a time varying kernel. They are both the easiest to use, and the simplest to implement. These methods present some drawbacks related to certain resolution limits and for some of them the need to add a tachometer signal as reference regime.

In this paper, we will focus on this last class of methods where the data can be analyzed and processed with many techniques, which are briefly reviewed below. The basic principle of these transforms is the replacement of the classical kernel of Fourier transform by a kernel whose base functions have frequencies that are proportional to the rotation speed of the system to be monitored. A drawback of this modification is that the kernel functions do not constitute any more an orthonormal basis. The Speed Transform (ST), proposed by Capdessus et al. in [7, 8], is normalized with the integration interval, which ensures asymptotic orthogonality in the case of linear speed variations. To cope with a wider range of speed variations, the Time Variant Discrete Fourier Transform (TVDFT), proposed by Blough et

al. in [3, 4, 9], comprises an orthogonality compensation matrix. But the most general technique is the Velocity Synchronous Discrete Fourier Transform (VSDFT) proposed by Borghesani et al. in [5, 6, 10], which consists in applying the time-variant kernel Fourier transform to a corrected temporal signal. The correction, which is a mere product of the raw signal by the angular speed, ensures the orthogonality of the decomposition whatever the speed variations.

In the case of transform based methods like ST [7], prior knowledge of the rotation frequency variations is needed and this information is provided by a tachometer*; hence only the order of the signal harmonics has to be estimated. However, the tachometer information is not always available, in which case techniques that operate without such prior knowledge are required. Interestingly, these time-varying kernel Fourier transforms are linked with the Quadratic Phase Transform (QPT) in that they fit to Polynomial Phase Signals (QPS). The QPS has been already considered in many signal processing applications including power systems [11], radar [12], and geophysics [13].

In this paper, we will extend the works in [14, 15]. Several methods for polynomial phase signal analysis already exist in the literature. We first provide an overview of the existing algorithms for polynomial phase parameter estimation which includes the time-frequency based methods in [16, 17], the (higher order) ambiguity function based methods in [18, 19], the iterative methods in [20, 21], and the fast decomposition methods in [22, 23]. Note that this survey is mainly limited to methods introduced in the signal processing and communications community that deserve to be better known and recognized as they provide interesting extensions to the existing tools used in vibration signal analysis.

The proposed extended work consists of the previously mentioned survey of existing methods for QPS parameter estimation followed by two new methods that lead to improved estimation accuracy with reduced computational cost. Then, we derive the CRB for the considered data models with or without a tachometer information and provide a comparative performance analysis via simulation experiments.

## II. PROBLEM FORMULATION

Rotating machine with constant speed are characterized by harmonic vibration signal of the form [9]

$$x(n) = \sum_{p=1}^{P} A_p e^{j2\pi f_p n} + w(n) \qquad (1)$$

where $P$ is the number of modes (harmonics), while $A_p$ and $f_p$ are the complex amplitude and the frequency of the $pth$ mode. In this case, the signal is referred to as a linear phase one (i.e. the component's phase terms vary linearly with time $\phi_p(n) = 2\pi f_p n$). The additive noise $w(n)$ is assumed to be of zero-mean, white and gaussian distributed with variance $\sigma_w^2$.

In the case of variable speed, the signal phase would depend non-linearly with time and the model becomes [7, 8]

$$x(n) = \sum_{p=1}^{P} A_p e^{j\phi_p(n)} + w(n) \qquad (2)$$

$\phi_p(n)$ being the phase term of the $pth$ component.

A particular case of interest, widely considered in the literature, is the one with uniform angular acceleration that corresponds to multi-component chirp signal. This case may be encountered in two different situations. First, on test benches where part of the vibrations measurements are done during a run-up or a coast-down of the machine. One can mention as an example the run-up and run-down experiments performed on gears, turbofans, or thermal motors, and other systems in order to study their stability and their balance or determine their critical speeds [24–28]. These techniques are now classically provided in commercial test devices and are performed through a linearly varying speed scenario. Second, in real-life operation, though the speed variations generally cannot be assumed to be linear, yet during the start-up and the coast-down stages, there usually is a portion of the speed variations that can still be approximated linearly [29]. In that context, the signal model is given by

$$x(n) = \sum_{p=1}^{P} A_p e^{j2\pi(f_p n + \gamma_p n^2)} + w(n) \qquad (3)$$

As we can see, the study of this signal coincide with the well-known polynomial phase signal analysis [22, 30].

In this paper, we focus only on the second order case (i.e. Quadratic phase signal). Therefore, our first objective is to study the parameter estimation problem for the signal model in (3). More precisely, after a brief survey of the methods in the literature, we propose to refine one of them (namely the FQPT method in [22]) and to compare it to the other existing techniques to assess its performance.

Our second objective, consists of considering a particular case of QPT signals given by the model

$$x(n) = \sum_{p=1}^{P} A_p e^{j2\pi o_p(fn + \gamma n^2)} + w(n) \qquad (4)$$

where $o_p$ is referred to as the $pth$ harmonic order. Parameters $(f, \gamma)$ might be known if a tachometer is available or unknown otherwise. In the latter case, we have to assume that $o_1 = 1$ (corresponding to the fundamental frequency) and $o_p > 1$ for $p \geq 2$, otherwise the model is not uniquely identifiable*.

The model in (4) has been shown to well represent the vibration signal in presence of 'faults' or 'abnormal' behavior of the machine [7, 9]. In this case, the harmonic parameters are used to 'control' and diagnose the current status of the rotating machine. For this particular model, we propose to adapt the HAF (Higher order Ambiguity Function) method to extract the desired signal parameters.

Finally, we propose to compare the two models in (3) and (4) through the evaluation of their corresponding Cramer-Rao Bounds.

---

* Recall that a tachometer is an instrument measuring the rotation speed and provides prior knowledge of the instantaneous fundamental frequency.

*Indeed, the two sets of parameters $\{f, \gamma, o_p, \; p = 1, \cdots, P\}$ and $\{o_1 f, o_1 \gamma, o_p/o_1, \; p = 1, \cdots, P\}$ lead to the same observation $x(n)$ and hence, without this normalization, one cannot identify our parameters in a unique way.

## III. OVERVIEW OF MAJOR QPS PARAMETER ESTIMATION ALGORITHMS

Polynomial phase signals have been used in many signal processing applications to model different physical phenomena characterized by nonlinear phases. We consider here the case of quadratic phase signals corresponding to the second order polynomial phase model given in (3). The signal being gaussian distributed, the maximum likelihood (ML) estimation of the QPS parameters coincides with the nonlinear least squares problem

$$\min_{\Theta} \sum_{n=0}^{N-1} |x(n) - \sum_{p=1}^{P} A_p e^{j2\pi(f_p n + \gamma_p n^2)}|^2 \tag{5}$$

where $\Theta$ is the vector of unknown QPS model parameters. This optimization problem being highly nonlinear and complex, many suboptimal solutions have been developed in the literature and are briefly reviewed in this section.

### A. Fast Quadratic Phase Transform (FQPT) method

The FQPT that was developed by Ikram et al. in [22], is based on the quadratic phase transform (QPT) for joint phase parameter estimation of multi-component chirp signals. The QPT is a second order polynomial phase transform. It has the ability to analyze chirp signals in the same way as Fourier transform analyzes sinusoidal signals. The Discrete QPT (DQPT) is defined as

$$DQPT\{x(n)\} = X(k,l) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi(\frac{k}{N}n + \frac{l}{M}n^2)} \tag{6}$$

where $k = 0, \cdots, N-1$ and $l = 0, \cdots, M-1$.

The FQPT is a fast implementation of the DQPT that transforms the one-dimensional sequence $x(n)$ into a two-dimensional sequence $X(k,l)$. It proceeds to joint phase parameters estimation after signal dechirping search. It uses some symmetry and translation characteristics to perform a fast transform and provide a significant computational saving. The FQPT can be seen as an approximate maximum likelihood method for large data sample.

Note that once the phase parameters are estimated from the peaks location of the DQPT, the amplitude coefficients are obtained by a simple least squares optimization according to

$$\begin{bmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_P \end{bmatrix} = \mathbf{M}^{\#} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \tag{7}$$

where $\mathbf{M}^{\#}$ refers to the pseudo-inverse of matrix $\mathbf{M}$ which is given by

$$\begin{bmatrix} 1 & \cdots & 1 \\ e^{j2\pi(\hat{f}_1 + \hat{\gamma}_1)} & \cdots & e^{j2\pi(\hat{f}_P + \hat{\gamma}_P)} \\ \vdots & & \vdots \\ e^{j2\pi(\hat{f}_1(N-1) + \hat{\gamma}_1(N-1)^2)} & \cdots & e^{j2\pi(\hat{f}_P(N-1) + \hat{\gamma}_P(N-1)^2)} \end{bmatrix}$$

Typically, $M \gg N$ since the value of the second order coefficient is usually much smaller than the first order coefficient and hence requires a finer grid search.

### B. Higher Order Ambiguity Function (HAF) method and its variants

The conventional HAF method, first introduced by Peleg et al. in [31] as the so called Discrete Polynomial Phase Transform (DPT) and named later HAF by Porat in [32], is a recursive algorithm that finds the highest-order phase coefficient, removes its contribution from the signal and repeats the process until all required phase parameters have been estimated. For a monocomponent QPS, $x(n) = A e^{j2\pi(fn + \gamma n^2)}$, the basic HAF is defined as follows

$$QPT[x(n), \tau] = x(n+\tau)x(n-\tau)^* = |A|^2 e^{j2\pi 2f\tau} e^{j2\pi 4\gamma \tau n} \tag{8}$$

$$\begin{aligned} HAF(\omega) &= DTFT\{QPT[x(n), \tau]\} \\ &= \sum_{n=0}^{N-1} QPT[x(n), \tau] e^{-j\omega n} \end{aligned} \tag{9}$$

The highest order $\gamma$ is estimated as

$$\hat{\gamma} = \frac{1}{8\tau\pi} arg \max_{\omega} |HAF(\omega)| \tag{10}$$

Once the highest order $\hat{\gamma}$ is estimated, we proceed to the suppression of its contribution by demodulation as

$$z(n) = x(n) e^{j(-2\pi\hat{\gamma}n^2)} \approx A e^{j2\pi fn} \tag{11}$$

The lowest order $f$ is estimated as

$$\hat{f} = \frac{1}{2\pi} arg \max_{\omega} |DTFT\{z(n)\}| \tag{12}$$

For the multi-component case, the transform in (8) would provide $P$ sinusoids (plus non desired cross terms) that can be estimated using DTFT or high resolution methods like MUSIC or ESPRIT [20]. The dechirping in (11) is then achieved successively for $\hat{\gamma}_1, \cdots, \hat{\gamma}_P$ for the estimation of the first order parameters $f_p$, $p = 1, \cdots, P$.

Unfortunately, HAF-based methods are characterized by the use of matched filter that suffer from an identifiability problem when dealing with multiple component QPS's having the same highest order phase coefficients.

To overcome this problem a so called product higher-order ambiguity function (PHAF) was developed, by Barbarossa et al. in [18]. The PHAF algorithm combines HAF and HIM (High order Instantaneous Moment) techniques for the detection and parameter estimation of multi-component QPS impinged in white Gaussian noise. It is defined as the product of $L$ HAFs for $L$ sets of lag parameters, and formulated as follows

$$PHAF(\omega) = \prod_{l=1}^{L} HAF^{(l)}(\omega) \tag{13}$$

where $HAF^{(l)}(\omega) = \sum_{n=0}^{N-1} QPT[x(n), \tau^{(l)}] e^{-j\omega F^{(l)} n}$, $\tau^{(l)}$ refers to the $l$th lag term, and $F^{(l)} = \frac{\tau^{(l)}}{\tau^{(1)}}$ is a frequency scaling factor. The highest order $\gamma$ is estimated as

$$\hat{\gamma} = \frac{1}{8\tau^{(1)}\pi} arg \max_{\omega} |PHAF(\omega)| \tag{14}$$

The use of the PHAF offers a number of advantages over the HAF. It solves the identifiability problem, improves noise rejection capabilities, and leads to improved performance as compared to HAF method.

Another improved version introduced in the literature by Ikram et al. in [20, 21], is the iterative HAF method.

It is based on iterative calculation of the polynomial phase signal coefficients. First, the second-order phase parameter is iteratively estimated and updated using a quadratic transform with appropriate lag parameter in each iteration in order to improve the estimation accuracy. Then the first-order phase parameter is estimated based on the same transform technique of demodulated signal, as defined below

$$z_q(n) = z_{q-1}(n + \tau_{q+1})z_{q-1}^*(n), \quad n = 0, \cdots, N - \tau_{q-1} - 1$$
(15)

where $z_{q-1}(n)$ is the demodulated signal defined as $z_{q-1}(n) = x(n)e^{-j\hat{\gamma}_{q-1}n^2}$, $n = 0, \cdots, N-1$ with $q$ is the iteration index and $z_0(n) = x(n)$.

Each estimated phase parameter is then refined using MUSIC or other high resolution algorithms. It is shown in [20], that the iterative HAF leads to a significant performance gain, especially at low SNRs.

**Remark**: A main advantage of HAF-type methods is their ability to deal with polynomial phase signals of any order (see [23, 30–32] for details).

Therefore, in the case of non-linear variable speed with non-uniform angular acceleration, one can use polynomial phase models of order higher than two and estimate their parameters via the considered HAF-type methods. Note that any continuous function over a finite interval can be 'well approximated' by a polynomial function as shown in [33]. However, the estimation cost of the phase parameters increases significantly with the polynomial order and hence using a second order polynomial represents a good trade-off in many practical situations [24, 25, 27].

### C. Time-Frequency Distribution (TFD) based method

This algorithm was developed, by Barkat et al. in [16, 17], to select and extract separately all components of frequency modulated signal. It's based on joint time-frequency analysis of the signal. Firstly, a one dimensional signal is transformed to Time-Frequency (TF) domain using appropriate reduced interference TF kernel (in [16, 17], the B-distribution was used).

Then, the undesired low energy peaks are removed by noise thresholding. After that, the number of components is estimated as the number of peaks of the TFD slice by means of the maximum argument of the histogram of the number of peaks computed for each time instant. Finally, a components separation procedure is applied to extract separately all the components by finding the components frequencies as the peaks positions of the TFD slice and identifying the crossing components using a smallest distance criterion.

This algorithm proved its superiority over the HAF method for the estimation of a multi-component signal.

The algorithm does not use any prior information about the various components to be extracted and its performance depends on the ability to suppress the cross-terms and needs high time-frequency resolution. Unfortunately, the algorithm needs a distribution that can reveal the features of the signal as clearly as possible to have a clean TFD, and has expensive computational load due to TFD computation and 2D search.

Note that, chirp signals are represented by 'straight lines' in the TF domain. Hence the problem of chirp parameter estimation has been assimilated to a contour (straight lines) extraction problem in an image and dedicated image processing tools have been used in [34] for solving this particular estimation problem. Other works related to the use of TF signal analysis tools for vibration signals can be found in [35, 36].

### IV. PROPOSED ALGORITHMS

We introduce in this section an improved version of the FQPT referred to as the Refined FQPT (R-FQPT) as well as adapted versions of the HAF method to deal with the 'order' estimation problem related to the model in (4).

### A. R-FQPT Algorithm

The FQPT, developed by Ikram and al. in [22], is described by

$$QPT\{x(n)\} = X(k, l) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi(\frac{k}{N}n + \frac{l}{M}n^2)} \quad (16)$$

where $N$ is the number of samples and $M \geq N^2$. The fast computation of the FQPT is based on three properties (see [22] for more details)

- Decomposition of the QPT in multiple DFT:

$$z^{(l)}(n) = x(n)e^{-j2\pi\frac{l}{M}n^2} \quad (17)$$

$$X(k, l) = DFT_n(z^{(l)}(n)) \quad (18)$$

- Decimation : used by Fast Fourier transform FFT to compute with reduced computational cost the Discrete Fourier Transform (DFT):

$$DFT_m(z^{(l)}(2^r m + s)) = DFT_m(z^{(l)}(2^{r+1}m + s)) \\ + e^{-j2\pi(\frac{k}{N}2^r)}DFT_m(z^{(l)}(2^{r+1}m + s + 2^r)) \quad (19)$$

- Symmetry:

$$DFT_m(z^{(l+\frac{M}{2^{r+1}})}(2^r m + s)) = DFT_m(z^{(l)}(2^r m + s)) \\ .e^{-j2\pi(\frac{s^2}{2^{r+1}})} \quad (20)$$

By using the previous properties, it is shown in [22] that the computational cost can be reduced from $O(N^2M)$ to $O(NM)$ flops. This cost is still relatively high and increases with the 2D grid size $N$ and $M$. Therefore to preserve 'a moderate' numerical complexity, one should limit the number of grid points which unfortunately affects the estimation precision of the FQPT. To overcome this drawback, we propose in the following a new

version of FQPT that refines the parameter estimation through a two-step procedure.

For that, we will use a modified filtering/unwrapping algorithm to refine the initial parameter estimate $(\hat{f}^{(0)}, \hat{\gamma}^{(0)})$ given by the FQPT algorithm. First, we demodulate the received quadratic polynomial phase signal using the initial FQPT estimated parameters according to $y(n) = x(n)e^{-j2\pi(\hat{f}^{(0)}n + \hat{\gamma}^{(0)}n^2)}$. Then, we apply a low-pass filtering to enhance the SNR. After that, we create a vector $\mathbf{v} = [v(0), v(1), \cdots, v(Q-1)]^T$ of unwrapped phase which is modeled as a new polynomial phase signal (i.e. $v(k) \approx \varphi + \delta f(kw_s) + \delta\gamma(kw_s)^2$) in noise with $Q = \frac{N}{w_s}$, $\omega_s$ being a down-sampling factor used to reduce the cost and improve the resolution of the method (this is needed here since the residual parameters $\delta f$ and $\delta\gamma$ are close to zero). Then, we apply a standard least squares estimation to obtain the new unknown phase parameters vector, defined as $\mathbf{v} = [\varphi, w_s\delta f, w_s^2\delta\gamma]^T$, according to

$$
\arg\min_{\mathbf{v}} \sum_{k=0}^{Q-1} |v(k) - \varphi + \delta f(kw_s) + \delta\gamma(kw_s)^2|^2
$$

$$
= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & (Q-1) & (Q-1)^2 \end{bmatrix}^{\#} \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(Q-1) \end{bmatrix} \quad (21)
$$

where $^{\#}$ refers to the matrix pseudo-inverse,

Finally, we obtain refined phase parameters by accumulating each FQPT initial estimated phase parameters with corresponding estimated phase parameters deviations according to $\hat{f} = \hat{f}^{(0)} + \delta\hat{f}$ and $\hat{\gamma} = \hat{\gamma}^{(0)} + \delta\hat{\gamma}$.

*B. Adapted HAF methods for order QPS*

For the case of O-QPS model of (4), the two steps of the HAF method are not needed (since, for each signal component, we have only one parameter to estimate instead of two for the general QPS model). Therefore, if a tachometer is available, the HAF method consists of only one step according to:

- For a given lag time $\tau$, compute signal

$$
y(n) = x(n+\tau)x(n)^* = \sum_{p=1}^{P} |A_p|^2 e^{j\phi_p(\tau)} e^{j4\pi\tau\gamma o_p n}
$$
$$
+ \text{ cross terms + noise}
$$

with $\phi_p(\tau) = 2\pi o_p(f\tau + \gamma\tau^2)$. Since the orders $o_p$ are typically in the range $[1, 10]$, a proper choice for time lag $\tau$ corresponds to $2\tau|\gamma| \approx 0.1$ so that the frequency terms $|4\pi\tau\gamma o_p|$ are maximally separated while not exceeding the upper limit $2\pi$ to avoid estimation ambiguity problems.

- Apply ESPRIT algorithm to signal $y(n)$ for the high resolution estimation of the $P$ pulsation terms $\hat{\omega}_p \approx 4\pi\tau\gamma o_p$ or equivalently

$$
\hat{o}_p = \hat{\omega}_p/(4\pi\tau\gamma)
$$

- Eventually, apply the iterative procedure of [37] for the refining of the previous parameter estimate.

If the tachometer is not available, parameter $\gamma$ can be estimated by using the fact that $o_1 = 1$ so that the first (smallest) pulsation value would correspond to $\gamma$, i.e.

$$
\hat{\gamma} = \hat{\omega}_1/(4\pi\tau).
$$

Once $\gamma$ obtained, parameter $f$ can be estimated via (12) with $z(n) = x(n)e^{-j2\pi\hat{\gamma}n^2}$.

## V. CRAMER-RAO BOUNDS DERIVATION

We provide here the detailed derivation of the CRB for the statistical models given in section II. Recall that, for a given data model depending on an unknown parameter vector, the CRB provides the lower limit to the statistical error variance for the unbiased estimation of the unknown model parameters.

The CRB is computed as the inverse of the Fisher Information Matrix (FIM) which entries are given by

$$
[FIM]_{i,j} = -E[\frac{\partial^2 \ln(p(\mathbf{x}; \Theta))}{\partial\theta_i\partial\theta_j}] \quad (22)
$$

where $E[.]$ represents the statistical expectation operator, $p(\mathbf{x}; \Theta)$ denotes the likelihood function, $\mathbf{x}$ is the observation vector and $\Theta$ denotes the parameter vector to be estimated.

The error covariance matrix of any unbiased estimator $\hat{\xi}$ of parameter vector $\xi = \mathbf{g}(\Theta)$ (denoted $Cov(\hat{\xi})$), where $\mathbf{g}(\Theta)$ is a locally differentiable function, is lower bounded by the CRB according to (i.e. $Cov(\hat{\xi}) - CRB(\xi)$ is a positive semi-definite matrix)

$$
Cov(\hat{\xi}) \geq CRB(\xi) = \bigtriangledown\mathbf{g}(\Theta)CRB(\Theta)(\bigtriangledown\mathbf{g}(\Theta))^T \quad (23)
$$

where $CRB(\Theta) = [FIM]^{-1}$ and $\bigtriangledown\mathbf{g}(\Theta)$ is the gradient matrix which $(i,j)^{th}$ element is defined by $\frac{\partial g_i(\Theta)}{\partial\theta_j}$. In particular, this means that the error variance in the estimation of parameters $g_i(\Theta)$ is greater than the $i-th$ diagonal element of $CRB(\xi)$.

Our case of interest corresponds to the Gaussian CRB for which the FIM matrix can be expressed by the Slepian-Bang formula [32]

$$
[FIM]_{i,j} = -E[\frac{\partial^2 \ln(p(\mathbf{x}; \Theta))}{\partial\theta_i\partial\theta_j}] = \frac{2}{\sigma_w^2} Re\{\frac{\partial\mathbf{s}^H}{\partial\theta_i}\frac{\partial\mathbf{s}}{\partial\theta_j}\} \quad (24)
$$

where $\mathbf{s}$ denotes the noiseless QPS signal in (3) and $Re[.]$ refers to the real part of a complex number. Note that in (24), we do not include the unknown parameter $\sigma_w^2$ in the parameter vector $\Theta$ since the FIM of the desired QPS parameters is decoupled from the noise variance [32].

In the sequel, we provide the FIM expressions for the three parametric models given in section II.

(a) General QPS case: The observed signal model is given by

$$
s(n) = \sum_{p=1}^{P} \rho_p e^{i\varphi_p} e^{2\pi i(f_p n + \gamma_p n^2)} \quad (25)
$$

where the complex amplitudes are written as $A_p = \rho_p e^{i\varphi_p}$. In that case, the unknown parameter vector of size $4P \times 1$

is given by $\Theta = [\{\rho_p\}, \{\varphi_p\}, \{f_p\}, \{\gamma_p\}]^T$ and the FIM is expressed as

$$FIM = \frac{2}{\sigma_w^2} Re\{\tilde{\mathbf{S}}^H \tilde{\mathbf{S}}\} \qquad (26)$$

where the $N \times (4P)$ matrix $\tilde{\mathbf{S}}$ is given by

$$\tilde{\mathbf{S}} = [\{\frac{\partial \mathbf{s}}{\partial \rho_p}\}, \{\frac{\partial \mathbf{s}}{\partial \varphi_p}\}, \{\frac{\partial \mathbf{s}}{\partial f_p}\}, \{\frac{\partial \mathbf{s}}{\partial \gamma_p}\}] \qquad (27)$$

(b) Order QPS case without tachometer: In that case, the observed signal model is given by

$$s(n) = \sum_{p=1}^{P} \rho_p e^{i\varphi_p} e^{2\pi i o_p (fn + \gamma n^2)} \qquad (28)$$

with $o_1 = 1$. For this model, we have $\Theta = [\{\rho_p\}, \{\varphi_p\}, \{o_p\}_{p>1}, f, \gamma]$ and the FIM is as in (26) with

$$\tilde{\mathbf{S}} = [\{\frac{\partial \mathbf{s}}{\partial \rho_p}\}, \{\frac{\partial \mathbf{s}}{\partial \varphi_p}\}, \{\frac{\partial \mathbf{s}}{\partial o_p}\}_{p>1}, \{\frac{\partial \mathbf{s}}{\partial f}\}, \{\frac{\partial \mathbf{s}}{\partial \gamma}\}] \quad (29)$$

For parameters $f_p = o_p f = g_p(\Theta)$, and $\gamma_p = o_p \gamma = h_p(\Theta)$ the minimum error variances are computed as

$$CRB(f_p) = \bigtriangledown g_p(\Theta) CRB(\Theta) (\bigtriangledown g_p(\Theta))^T \qquad (30)$$

$$CRB(\gamma_p) = \bigtriangledown h_p(\Theta) CRB(\Theta) (\bigtriangledown h_p(\Theta))^T \qquad (31)$$

(c) Order QPS case with tachometer: The observed signal model is like in (28) but with parameters $f$ and $\gamma$ known. In that case, the unknown parameter vector becomes $\Theta = [\{\rho_p\}, \{\varphi_p\}, \{o_p\}]$ and the matrix $\tilde{\mathbf{S}} = [\{\frac{\partial \mathbf{s}}{\partial \rho_p}\}, \{\frac{\partial \mathbf{s}}{\partial \varphi_p}\}, \{\frac{\partial \mathbf{s}}{\partial o_p}\}]$.

## VI. PERFORMANCE ANALYSIS AND COMPARATIVE RESULTS

In this section, numerical investigations are conducted for the following objectives:

- For the general QPS case, we provide a comparative performance analysis of the different methods considered in the survey part. All results are compared to the CRB to assess their statistical efficiency.

- The previous comparative study is used to highlight the performance gain of the proposed R-FQPT method in the moderate and low SNR region.

- We compare the CRBs for the general QPS, O-QPS with tachometer, and O-QPS without tachometer cases. In particular, this comparison allows us to evaluate the potential performance gain brought by the tachometer use.

- We assess the performance of the proposed HAF method adapted to the O-QPS case and compare it to the corresponding CRB.

- Finally, the R-FQPT is compared to ST and VSDFT on a simulated vibration signal.

We assume that the observed vibration signal is a multi-component QPS in additive white Gaussian noise. The additive noise is randomly generated as circular complex white Gaussian process with zero mean and unknown variance $\sigma_w^2$. The signal to noise ratio (SNR) is defined as

$$SNR = 10 log_{10} \frac{1}{\sigma_w^2} \qquad (32)$$

As performance measures, we use the averaged mean squared error (MSE) defined as

$$MSE(\hat{\theta}_p) = \frac{1}{Mc} \sum_{i=1}^{Mc} (\hat{\theta}_p^{(i)} - \theta_p)^2 \qquad (33)$$

where $\hat{\theta}_p^{(i)}$ is the estimate of the $p$-th vibration signal parameter $\theta_p$ at the $i$-th run. The numerical values are obtained over $Mc = 100$ independent Monte Carlo trials where the SNR varies between $-20$ and $20 dB$ in all examples.

In the first example, we evaluate the performance of the proposed R-FQPT for the general QPS case. For that, we present a comparative study between the R-FQPT and the previously reviewed methods in section III. The simulated environment considers a single component and multiple component vibration signal with speed variation functions defined, in reduced frequency, as $f_1(t) = (0.02 + \frac{0.00025}{2}t)$ and $f_2(t) = (0.045 + \frac{0.00075}{2}t)$ and the complex amplitudes are $A_1 = 1e^{j\pi/10}$ and $A_2 = 1e^{j\pi/5}$. The signal is defined as

$$x(t) \approx \sum_{p=1}^{P} A_p e^{j2\pi \int_0^t f_p(u) du}$$

with $P = 1$ for single component and $P = 2$ for multiple components case. The signal duration is about 256 samples.

The accuracy of the proposed R-FQPT algorithm is compared to the HAF method introduced by Barbarossa et al. in [18, 19], the iterative HAF method developed by Ikram et al. in [20], and the TFD method proposed by Barkat et al. in [16] as well as with the CRB given in section V.

It can be seen in figure 1 that, the R-FQPT algorithm outperforms the HAF, the iterative HAF, and the TFD one for both single and multi-component signal cases, especially for low SNR. In the asymptotic region (higher SNR) the difference is very small but the proposed algorithm remains slightly better.

In the second example, we present a comparative study of the CRB expressions derived for general and order QPS cases. For that, we consider the signal models presented in 3 and 4, where the parameters are $N = 256$, $P = 3$, $\rho = [1, 1, 1]$, $\varphi = [\pi/10, \pi/5, \pi/2.5]$, $f = [0.1, 0.2, 0.3]$, $\gamma = [0.001, 0.002, 0.003]$ and $o = [1, 2, 3]$. From figure 2, it can be seen that there is almost no gain for the estimation of the amplitude parameters when considering the model in Eq. (4) instead of (3). However, for the estimation of the quadratic phase parameters, a gain of about 3dB is observed when considering the additional information given by the O-QPS model.
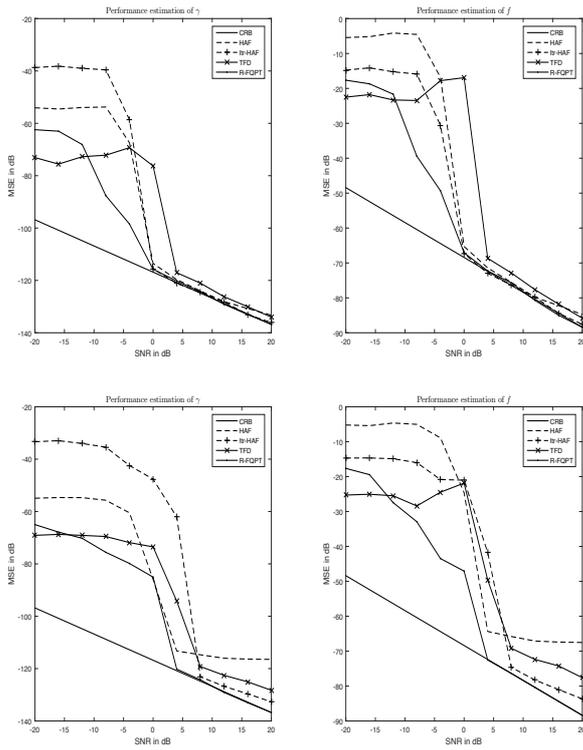
**Fig. 1**: MSE of $f$ and $\gamma$ versus SNR for a single component signal (top plots) and a two components signal (bottom plots).

**Fig. 2**: CRB comparison

Now, if a tachometer is used, the previous gain becomes much larger and exceeds 10 dB for the considered example. Consequently, when possible, the use of a tachometer for the vibration signal analysis is highly recommended.

In the third scenario, we evaluate the performance of the proposed Adapted-HAF method for the order QPS case with tachometer. The performance of the proposed method is compared to the corresponding derived CRBs in section V. For that, We consider the simulated multi-component signal of example 2. In figure 3, we compare the cumulated MSE of the 3 estimated orders with their corresponding CRB. As we can see, the MSE performance of the proposed adapted-HAF algorithm reaches the CRB for moderate SNR values (for $SNR > 0dB$ in the two components case and $SNR > 5dB$ for the three components case). Hence the Adapted-HAF presents a low cost alternative solution when the tachometer is available.

In the final experiment, we will apply the proposed methods to a time varying vibration of a Diesel engine. Results are compared to those of, Speed Transform (ST) proposed by Capdessus at al. in [7] and Velocity Synchronous-DFT developed by Borghesani et al. in [6].

R-FQPT is first compared to ST and VSDFT through a simulated scenario that considers a three component vibration signal with speed variation function defined, in reduced frequency, as $f_r(t) = (0.16 + 0.00003t)$ given by tachometer. The signal is defined as

$$x(t) \approx \sum_{p=1}^{3} A_p e^{j2\pi o_p \int_0^t f_r(u)du} \approx \sum_{p=1}^{3} A_p e^{j2\pi o_p (0.16t + \frac{0.00003}{2}t^2)}$$
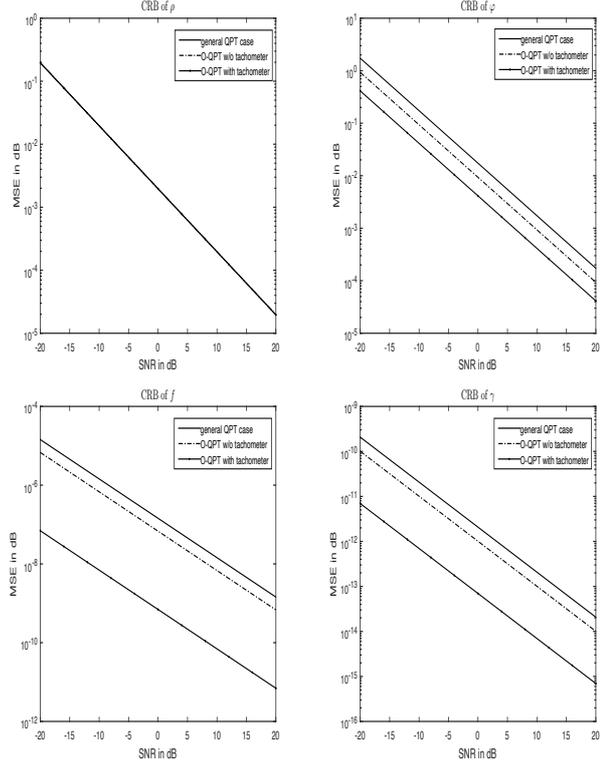
where $A_1 = A_2 = A_3 = 1$, $o_1 = 2$, $o_2 = 4$ and $o_3 = 6$. The signal comprises 512 samples.

It can be seen in figure 4 that the proposed approach allows estimating jointly the polynomial coefficient parameters of the rotation speed, that corresponds well to the considered harmonic's orders. Indeed, three corresponding peaks are located on a straight line whose direction vector corresponds to the reduced rotation frequency variations and the positions of the peaks allow estimating the corresponding order. R-FQPT thus allows estimating directly the parameters of the three chirp components whereas the ST and VSDFT give the order of the harmonic relatively to a known rotation speed.

## VII. CONCLUSION

The paper introduced a thorough analysis of the parameter estimation problem associated to variable speed vibration signal. The latter is shown to be a quadratic phase signal for which a plethora of estimation algorithms exist in the literature.

After a brief survey of the major class of methods, we proposed two new solutions, namely the R-FQPT for the general QPS case and the Adapted HAF for the O-QPS.

Simulation experiments have been conducted to illustrate the behavior and advantages of the proposed solutions and compare their performance to those of the major algorithms from the literature. We have also derived the CRB expressions for the different quadratic phase models and exploited them to investigate the potential gain due to the side information related to the Order QPS model and to the tachometer use, respectively.

As a perspective, we believe it would be highly beneficial to
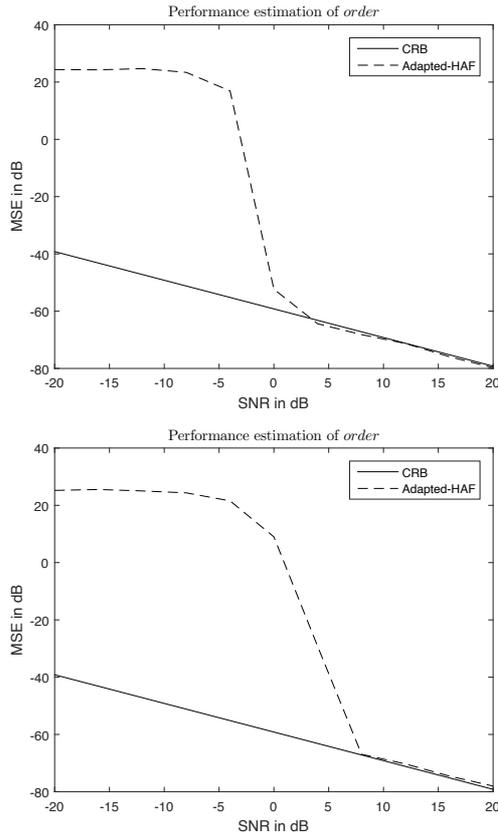
**Fig. 3**: MSE of the estimated order versus SNR for a two components signal (top plots) and a three components signal (bottom plots)

consider in future works an extension of our paper dedicated to an enlarged comparative study of all methods (QPT, HAF, ST, TVDFT, VSDFT) that can be used for the considered vibration signal analysis. In addition, an enriched validation and testing of these methods, based on real data measurements, would be of high importance for their practical use.

REFERENCES

[1] N. O. Myklestad, Fundamentals Of Vibration Analysis, Paperback edition, 2018.

[2] C. Scheffer, Practical Machinery Vibration Analysis and Predictive Maintenance, Newnes Publication, 2004.

[3] J. R. Blough, D. L. Brown, H. Vold, The time variant discrete Fourier transform as a tracking order method, Proceedings of Society of Automotive Engineers Noise and Vibration Conference, SAE Paper Number 972006, 3 (54) (1997) 823–837. https://doi.org/10.4271/972006

[4] J. R. Blough, Development and analysis of time variant discrete Fourier transform order tracking, Mechanical Systems and Signal Processing, 6 (17) (2003) 185—-1199. https://doi.org/10.1006/mssp.2002.1500

[5] P. Borghesani, P. Pennacchi, R. B. Randall, R. Ricci, Order tracking for discrete-random separation in variable speed conditions, Mechanical Systems and Signal Processing, (30) (2012) 1–22. https://doi.org/10.1016/j.ymssp.2012.01.015

[6] P. Borghesani, P. Pennacchi, S. Chatterton, R. Ricci, The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery, Mechanical Systems
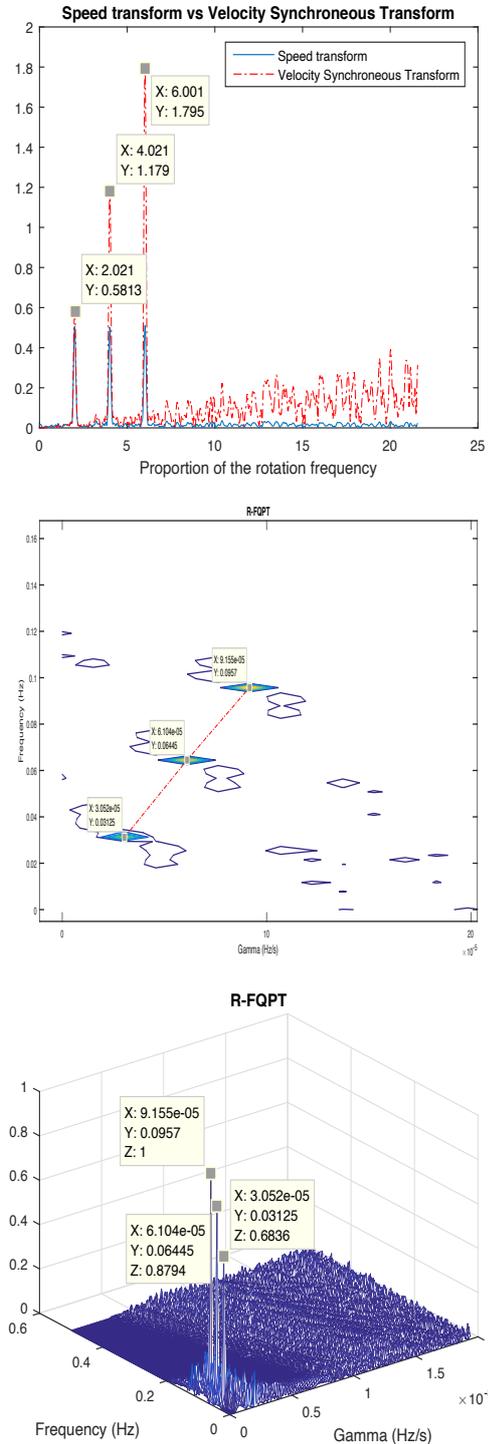


**Fig. 4**: ST, VSDFT and R-FQPT of a simulated vibration signal (function of reduced frequency)

and Signal Processing, 44 (2014) 118–133. https://doi.org/10.1016/j.ymssp.2013.03.026

[7] C. Capdessus, E. Sekko, J. Antoni, Speed transform, a new time-varying frequency analysis technique, Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Lecture Notes in Mechanical Engineering, (2014) 23–35. https://doi.org/10.1007/978-3-642-39348-8_2

[8] J. Roussel, M. Haritopoulos, E. Sekko, C. Capdessus, J. Antoni, Application of speed transform to the diagnosis of a roller bearing in variable speed, in Proc. of Surveillance 7, Chartres, France, Oct. 2013 (October 2013).

[9] J. R. Blough, A survey of DSP methods for rotating machinery analysis, what is needed, what is available, Journal of Sound and Vibration, (262) (2003) 707–720. http://doi.org/10.1016/S0022-460X(03)00118-4

[10] P. Borghesani, R. Ricci, S. Chatterton, P. Pennacchi, A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions, Mechanical Systems and Signal Processing, (38) (2013) 23–35. https://doi.org/10.1016/j.ymssp.2012.09.014

[11] P. O'Shea, M. Farquharson, G. Ledwich, Estimation of time-varying mains frequencies, in Proc. of the Australasian Universities Power engineering Conference (2003).

[12] V. Katkovnik, Nonparametric estimation of instantaneous frequency, IEEE Transactions on Information Theory, 43 (1) (1998) 183–189. https://doi.org/10.1109/18.567676

[13] J. Angeby, Estimating signal parameters using the nonlinear instantaneous least squares approach, IEEE Transactions on Signal Processing, 48 (10) (2000) 2721–2732. https://doi.org/10.1109/78.869022

[14] A. Abbadi, C. Capdessus, K. Abed-Meraim, E. Sekko, Vibration signal parameters estimation in variable speed using quadratic phase transform, in Proc. of Surveillance 8, Roanne, France, Oct. 2015.

[15] A. Abbadi, K. Abed-Meraim, C. Capdessus, A. Belouchrani, F. Soltani, Localisation en champ proche par formation de voies: Algorithme rapide et analyse de résolution, in Proc. GRETSI 2015, Lyon, France, Sep. 2015 (Sep. 2015).

[16] B. Barkat, K. Abed-Meraim, Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture, EURASIP Journal of Applied Signal Processing (JASP), (13) (2004) 2025–2033. https://doi.org/10.1155/S1110865704404193

[17] B. Barkat, K. Abed-Meraim, A blind components separation procedure for FM signal analysis, in Proc. of ICASSP, USA, (2) (2002) 1425–1428. https://doi.org/10.1109/ICASSP.2002.5745820

[18] S. Barbarossa, A. Scaglione, G. Giannakis, Product high-order ambiguity function for multicomponent polynomial-phase signal modelling, IEEE Transactions on Signal Processing, 46 (3) (1998) 691–708. https://doi.org/10.1109/78.661336

[19] S. Barbarossa, V. Petrone, Analysis of polynomial-phase signals by the integrated generalized ambiguity function, IEEE Transactions on Signal Processing, 45 (2) (1997) 316–327. https://doi.org/10.1109/78.554297

[20] M. Z. Ikram, K. Abed-Meraim, Y. Hua, Iterative parameter estimation of multiple chirp signals, Electronics Letters, 33 (8) (1997) 657–659. https://doi.org/10.1049/el:19970466

[21] M. Z. Ikram, K. Abed-Meraim, Y. Hua, Estimating the parameters of chirp signals: An iterative approach, IEEE Transaction on Signal Processing, 46 (12) (1998) 3436–3441. https://doi.org/10.1109/78.735320

[22] M. Z. Ikram, K. Abed-Meraim, Y. Hua, Fast quadratic phase transform for estimating the parameters of multicomponent chirp signals, Digital Signal Processing, 7 (2) (1997) 127–135. https://doi.org/10.1006/dspr.1997.0286

[23] Y. Ju, G. Bi, Generalized fast algorithms for the polynomial time-frequency transform, IEEE Transaction on Signal Processing, 55 (2007) 4907–4915. https://doi.org/10.1109/TSP.2007.896102

[24] M. Z. Khan, M. Suleman, M. Ashraf, A. Q. Khan, Some practical aspects of balancing an ultra-centrifuge rotor, Journal of Nuclear Science and Technology, 24 (11) (1987) 951–959. https://doi.org/10.1080/18811248.1987.9733526

[25] L. F. Villa, A. Reñones, J. R. Perán, L. J. de Miguel, Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load, Journal of Mechanical Systems and Signal Processing, 29 (May 2012) 436–446. https://doi.org/10.1016/j.ymssp.2011.12.013

[26] J. Drouet, Diagnostic qualité de boîtes de vitesse automatiques à partir de mesures acoustiques et vibratoires, VIbrations, SHocks and NOise (VISHNO), Le Mans, France, (2016) 2441–2446.

[27] M. Desbazeille, Diagnostic de groupes électrogènes diesel par analyse de la vitesse de rotation du vilebrequin, Thèse de doctorat de l'Université Jean Monnet, (2011, (HAL Id: tel-00563111 https://tel.archives-ouvertes.fr/tel-00563111)).

[28] A. C. Sall, Comportement dynamique d'un redresseur de turbomachine aéronautique: Effets du désaccordage, Thèse soutenue à l'Ecole Centrale de Lyon le 8 Décembre, (2011).

[29] E. Swanson, C. D. Powell, S. Weissman, A practical review of rotating machinery critical speeds and modes, Journal of Sound and Vibration, 39 (5) (2005) 10–17.

[30] S. Peleg, B. Porat, B. Friedlander, The discrete polynomial transform (DPT), its properties and applications, IEEE in Proc. Twenty-Fifth Asilomar Conf Signals, Systems and Computers, 1 (1991) 116–120. https://doi.org/10.1109/ACSSC.1991.186425

[31] S. Peleg, B. Porat, Estimation and classification of signals with polynomial phase, IEEE Trans. Inform. Theory, 37 (1991) 422—-430. https://doi.org/10.1109/18.75269

[32] B. Porat, Digital Processing of Random Signals, Theory and Methods, Vol. sec. 12, NJ: Prentice-Hall, Englewood Cliffs, 1994.

[33] A. Pinkus, Weierstrass and approximation theory, Journal of Approx. Theory, 107 (1) (2000) 1–66. https://doi.org/10.1006/jath.2000.3508

[34] K. Abed-Meraim, L. Nguyen, V. Sucic, F. Tupin, B. Boashash, An image processing approach for underdetermined blind separation of nonstationary sources, in Proc. of Int. Symp. on Sig. and Image Proc. and Analysis, Rome, (Sep. 2003) 185—-1199. https://doi.org/10.1109/ISPA.2003.1296856

[35] Z. Feng, M. Liang, F. Chu, Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples, Journal of Mech. Syst. Signal Process. 38 (1) (2013) 165—-205. https://doi.org/10.1016/j.ymssp.2013.01.017

[36] O. Cardona-Morales, L. Avendaño, G. Castellanos-Domínguez, Nonlinear model for condition monitoring of non-stationary vibration signals in ship driveline application, Journal of Mechanical Systems and Signal Processing, 44 (1–2) (2014) 134–148. https://doi.org/10.1016/j.ymssp.2013.08.029

[37] P. O'Shea, On refining polynomial phase signal parameter estimates, IEEE Transactions On Aerospace and Electronic Systems, 46 (3) (2010) 978–987. https://doi.org/10.1109/TAES.2010.5545168

**Ali Abbadi**  was born in 1979 in Rouiba, Algeria. He received the State Engineering Degree in Electrical Engineering in 2003 and the Magistère degree in 2008 in signal processing, both from Ecole Militaire Polytechnique, Algeria. He completed his PhD degree in 2018 with a specialization in radar detection and array signal processing which is his current field of research.


**Cécile Capdessus**  was born in France in 1965. She received an engineering degree and a Ph. D. in signal processing from the Institut National Polytechnique de Grenoble (INPG), in Grenoble, France, in 1989 and 1992 respectively. After a one year post-doctoral position at the Universitat Politècnica de Catalunya (UTC), in Barcelona, Spain, she spent two years as an assistant lecturer first at the University of Orléans and then at the INPG. Since 1995 she has worked as a lecturer at the University of Orléans, and is currently a member of PRISME Laboratory. Her main research field is the application of signal processing techniques to vibration analysis.


**Karim Abed-Meraim**  was born in 1967, Algiers, Algeria. He received the State Engineering Degree from Ecole Polytechnique, Palaiseau, France, in 1990, the State Engineering Degree from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1992, the M.Sc. degree from Paris XI University, Orsay, France, in 1992 and the Ph.D. degree from the ENST in 1995 (in the field of Signal Processing and communications). From 1995 to 1998, he took a position as a research staff at the Electrical Engineering Department of the University of Melbourne where he worked on several research project related to "Blind System Identification for Wireless Communications", "Blind Source Separation", and "Array Processing for Communications", respectively. From 1998 to 2012 he has been Assistant then Associate Professor at the Signal and Image Processing Department of Telecom-ParisTech. His research interests are in signal processing for communications, adaptive filtering and tracking, array processing and statistical performance analysis. In september 2012 he joined the University of Orléans (PRISME Lab.) as a full Professor. He is the author of over 480 scientific publications including book chapters, patents, and international journal and conference papers. Dr. Abed-Meraim is an IEEE Fellow, an IEEE SAM-TC member, a member of the TAC on Signal Processing for Multisensor Systems (EURASIP), and a senior area editor for the IEEE Transactions on Signal Processing.


**Edgard Sekko**  was born in Cotonou, Benin, in 1971. He received the M.Sc. and the doctoral degree in electrical engineering from Claude Bernard University, Lyon, France, in 1993 and 1998, respectively. Since 1999, he has been with the Technology Institute of Engineering Science at the University of Orleans, France, as an Assistant Professor. His main research activities interest vibration analysis applied to both the diagnosis of gearbox and rolling bearing in variable speed by using $H\alpha$ estimation theory.

# Non-conventional Band-notched UWB Antenna Design Based on Genetic Algorithm

Khelil Fertas, Soufiane Tebache, Fouad Fertas, and Rabia Aksas

**Abstract—In this paper, an improved applied approach based on Genetic Algorithm (GA) for antenna design is presented. It is basically useful to overcome electromagnetic interferences in UWB applications. First, an antenna in the range 2.8 - 20 GHz including UWB band was designed by using GA, then the undesired WiMAX band (3.5-3.8GHz) is rejected by optimizing appropriate Defected Ground Structure (DGS) filter. The proposed non-conventional filter consists of a matrix of rectangular cell shapes where each one is allocated by presence or absence of metal. The process is implemented in CST software using visual basic script. The method overcomes difficulties of conventional design. Furthermore, the optimized shape can be automatically obtained without appealing expert designers. Simulated and measured results demonstrate the effectiveness of the proposed design approach.**

*Keywords*—**Ultra Wide Band (UWB), Genetic Algorithm (GA), Non conventional DGS filter.**

## NOMENCLATURE

| | |
|---|---|
| GAO | Genetic Algorithm Optimisation. |
| UWB | Ultra Wide-Band. |
| DGS | Defected Ground Structure. |
| VBS | Visual Basic Script. |
| CST | Computer Simulation Technology. |

## I. INTRODUCTION

Ultra Wide Band (UWB) antennas have been of great interest in nowadays telecommunications systems. However, extensive usage of UWB technology provides remarkable high level of interferences between users. Hence, reducing interferences requires a special treatment which can be achieved by incorporating band-reject filters in UWB antenna systems [1–4].

Filter optimization performance becomes very challenging as long as it should satisfy various aspects such as high rejection level in single or multi-band configurations, accuracy and ease of integration in small devices. Consequently, various methods were investigated to overcome this challenge [5–7]. In [5], traditional DGS technique is used in filter design. Authors in [6] proposed a Tri-stopband filter using metamaterial cells.

It is worthwhile to point out that conventional methods based on adding resonators or etching slots are almost time- consuming, complex and hardly achieve desired requirements. To surmount the disadvantage of using such techniques, a genetic algorithm optimization (GAO) has been widely exploited in the microwave and antenna applications [8–11].

GAO is an automatic method based on genetic concept. It is commonly used to solve optimization and search problems by exploiting some biologically inspired operators known as mutation, crossover and selection.

In this paper, a novel non-conventional defected ground structure (DGS) filter is proposed. It is designed for WiMAX (3.5-3.8GHz) band rejection in wide-band antenna. The filter is optimized by GA method where the initial code is developed using visual basic script (VBS) which is integrated into CST environment for electromagnetic simulations.

## II. ANTENNA CONFIGURATION

The optimized antenna with non-conventional DGS filter is depicted in Fig. 1. It consists of two sides over a Teflon substrate with relative permittivity $\varepsilon_r = 2.4$ and height $h = 1.6mm$. The monopole antenna operating in the frequency band 2.8-20 GHz is illustrated in Fig. 1(a). It is designed using the code developed in [12], which is based on genetic algorithm optimization and implemented on CST Microwave studio as a simulation tool. The 50 $\Omega$ microstrip line is used to supply the antenna.
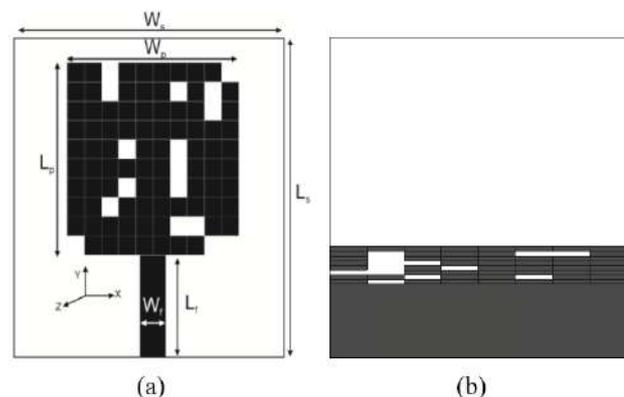


**Fig. 1**: Proposed antenna with non-conventional DGS filter. (a) Front view. (b) Back view. ($L_f = 18mm, W_f = 3.1mm, W_s = 33mm, L_s = 43mm, W_p = 20mm, L_p = 20mm$).

Further, the effectiveness of the method is studied and extended in order to reject some intentionally undesired frequency-

*Khelil Fertas, Soufiane Tebache, Rabia Aksas are with the Electronics Department, Ecole Nationale Polytechnique, Algiers, ALGERIA. (e-mail: khelil.fertas@g.enp.edu.dz, soufiane.tebache@g.enp.edu.dz, rabia.aksas@g.enp.edu.dz).*
*Fouad Fertas is with the Electrical Engineering Department, University of Boumerdes, ALGERIA. (e-mail: f.fertas@univ-boumerdes.dz)*

bands in the UWB. The process consists of optimizing a non-conventional DGS filter as seen in Fig. 1(b). The filter is made of a matrix of rectangular cell shapes where each one is allocated by presence or absence (slots) of metal , which are translated to binary code "1" and "0", respectively. The dimension of each rectangle is set to $0.5 \: x \: 4 \: mm^2$.

In order to find the best configuration, the GAO changes the pixel position in the area of DGS filter (situated at high current distribution location). Consequently, the current distribution is also modified according to the pixel distribution for suppressing undesired bands.

The code developed, based on different steps, is applied with optimization fitness value of multi-objective function, to achieve good rejection (good impedance mismatch) of $S_{11}$ in frequency band [3.5 - 3.8 GHz] that correspond to WiMAX standard. The first step consists to generating a random population of matrices corresponding to different shapes of the design where pixels are coded (either 1 or 0 value for presence/absence of the conductor respectively). In the second step, the generated populations are simulated so that the reflexion coefficient should satisfy the filter condition around the desired band. Then, the next generation is formed from the previous one with changes in pixel position in the area of DGS filter. This process is iteratively repeated until the satisfaction of the stop criterion (cost function) corresponding to the optimal solution.

The cost function of multi-objective problem is applied to attain good matching of the reflection coefficient ($|S_{11}| < -10dB$) in the desired bands (3 – 3.5 GHz & 3.8 – 20 GHz) and impedance mismatch ($|S_{11}| > -10dB$) in the rejected WiMAX frequency band (3.5 – 3.8 GHz). Moreover, the properties of the GA, applied in this case, are the following: initial population set consists of twenty (20) elements. The best cost functions are used to generate the novel population. The probability of crossover is 100% from the best selected and the single point crossover technique is used with mutation probability equal to 0.1%.

Compared to conventional methods, designers optimize all filter parameters with adding, in most cases, extra components in the cost of bulk and size [13–15].

## III. RESULTS AND DISCUSSION

### A. Simulation

The simulated reflexion coefficient $|S_{11}|$ of the wide-band antenna with and without non-conventional DGS filter is shown in Fig. 2. It is clearly noticeable that the DGS has effectively rejected the WiMAX band centred at 3.7 GHz. This result highlights the capability of the GAO to overcome optimization challenges.

To furthermore validate the proposed approach, the study of surface current distribution is carried out. The simulated current distribution is illustrated in Fig. 3, for resonant frequencies of 3.7 GHz (rejected frequency) and 6 GHz. One can clearly observe, from this figure, the effect of incorporating the non-conventional DGS filter where currents are concentrated around the filter part as seen in Fig. 3(a) corresponding to the rejected frequency (3.7GHz). In this case, the antenna does not radiate and the currents are trapped by the filter. When comparing to Fig. 3(b), we can conclude that the filter does not disturb the frequency 6 GHz, which is radiated by the antenna.

For more emphasis, the gain variation is shown in Fig. 4. It's obvious that the filtering effect is achieved at the rejected frequency of 3.7 GHz where the gain decreases drastically from
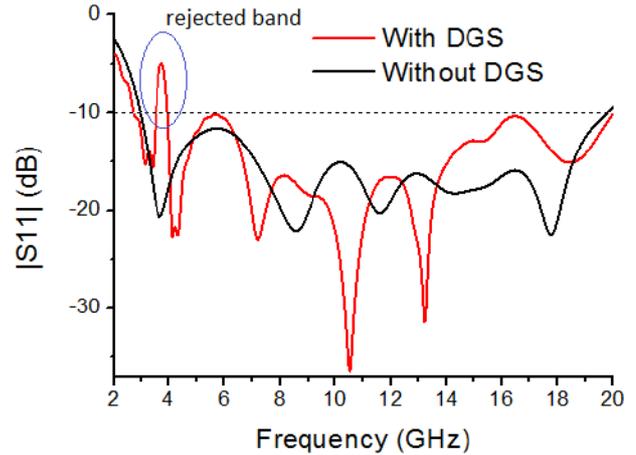


**Fig. 2**: Simulated $|S_{11}|$ with and without DGS filter.
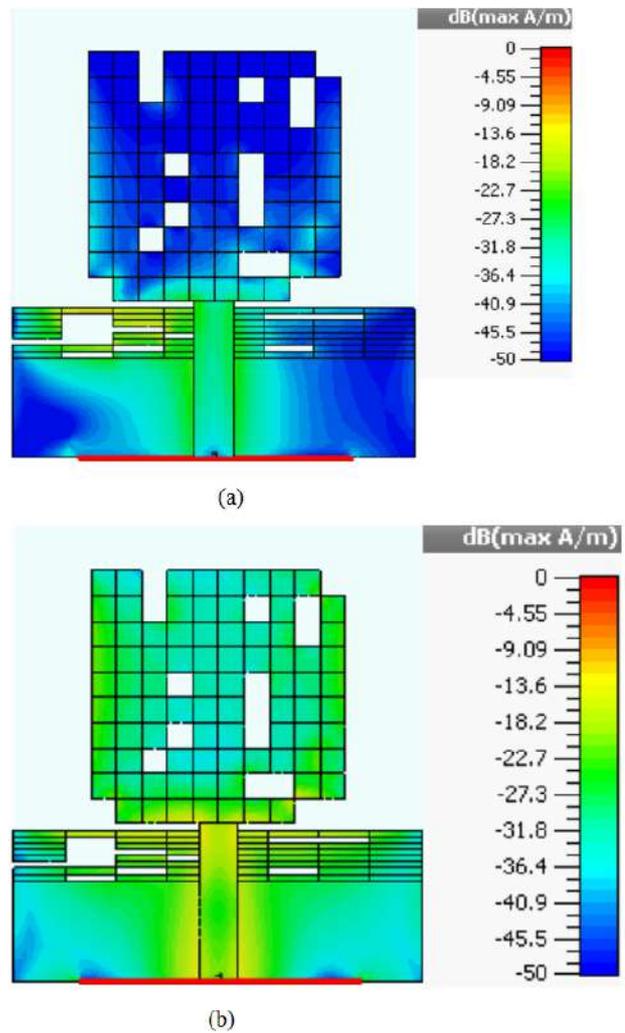


**Fig. 3**: Surface current distributions at (a) 3.7 GHz (b) 6 GHz.

4.4 dB to less than -14 dB.

Beside, the radiation pattern of the proposed design is investigated. Simulated radiation patterns before and after DGS filter optimization are shown in Fig. 5 and Fig. 6 respectively, at different frequencies 3GHz, 5GHz and 13GHz. It is seen that the
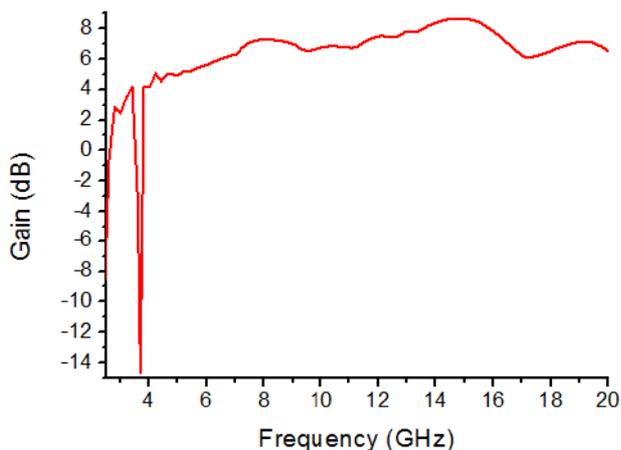
**Fig. 4**: Gain of the antenna with proposed DGS filter.

pattern is omnidirectional in H-plane ($Phi = 0°$) and takes the eight shape in the E-plane($Phi = 90°$) which is similar to that of a simple monopole in the lower band (3 and 5GHz). In the upper band (13GHz), the pattern is somehow distorted, this is may due to the excitation of higher modes. Moreover, it is observed that radiation patterns before and after DGS filter optimization are quasi-similar. This demonstrates that the optimized filter did not affect the radiation.
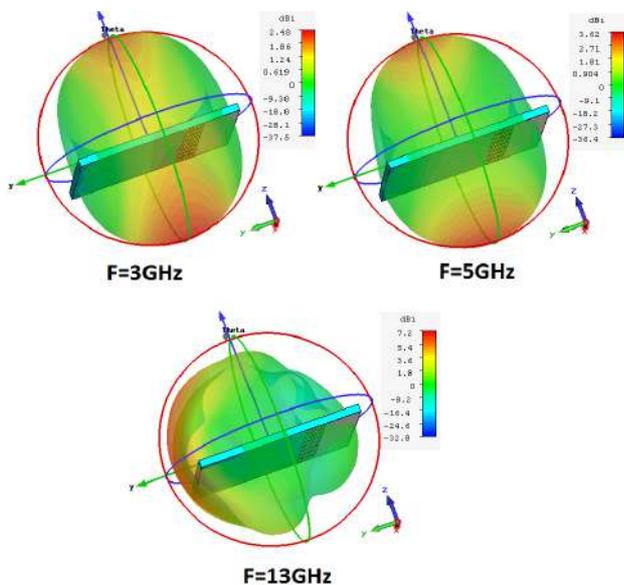


**Fig. 5**: Simulated 3D radiation pattern at 3GHz, 5GHz and 13GHz, before DGS filter optimization.

*B. Measurement*

In order to validate experimentally the proposed approach, a prototype of the antenna is fabricated and measured as seen in Fig. 7.

Fig. 8 depicts the simulated and measured reflexion coefficient $S_{11}$-Parameter of the proposed antenna with non-conventional DGS filter. It is well noticed that the WiMAX band centered at 3.7 GHz is effectively rejected. Moreover, simulation and measurement are in quite agreement with a small noticed discrepancy in high frequency (starting from 17GHz) where measured
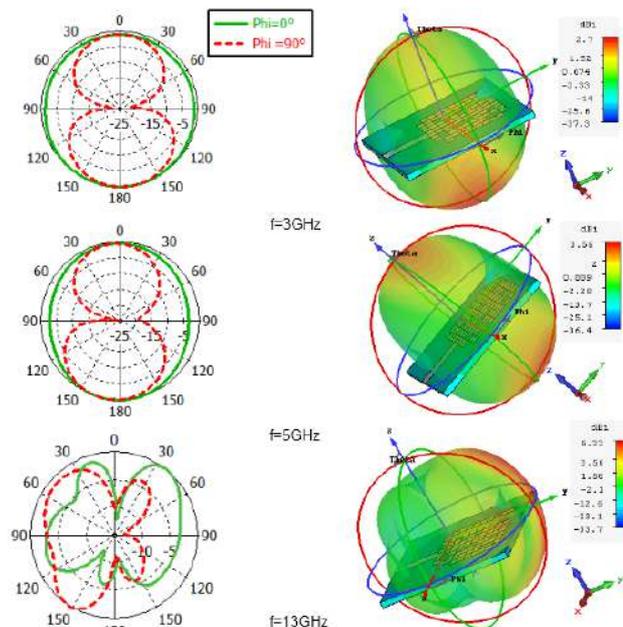


**Fig. 6**: Simulated 2D&3D radiation pattern at 3GHz, 5GHz and 13GHz after DGS filter optimization.



**Fig. 7**: $S_{11}$ and pattern measurement testbed (a) Network analyzer (b) Anechoic chamber.
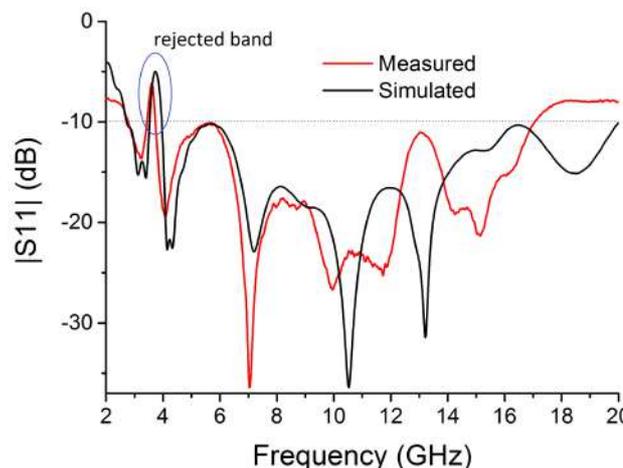


**Fig. 8**: Simulated and measured reflexion coefficient $S_{11}$-Parameter of the antenna with non-conventional DGS filter.

$S_{11}$ exceeds slightly -10dB, contrary to simulation. This discrepancy may due to fabrication tolerances, variation in relative permittivity of the substrate in such high frequencies, measurements conditions ($S_{11}$ measurements were carried out outdoor of anechoic chamber) and the use of SMA connector with maximum frequency of less than 18GHz. Although this small discrepancy, it can be confirmed that the proposed approach is
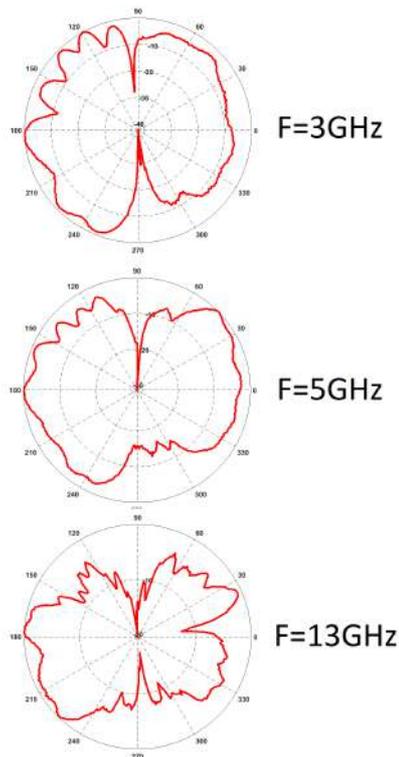
**Fig. 9**: Measured radiation pattern at 3GHz, 5GHz and 13GHz after DGS filter optimization.

practical and easily applicable. Further, Radiation pattern of the optimized antenna was measured in LEME Laboratory (Paris Nanterre University) anechoic chamber as depicted in Fig. 7(b). Obtained results at 3GHz, 5GHz and 13GHz are shown in Fig. 9. It is observed that patterns are almost similar to conventional monopole antenna as they are quasi-omnidirectional except for the higher frequency of 13GHz, where distortions, due to higher modes, are observed. It can be then stated that the DGS filter did not affect the radiation of the proposed antenna.

## IV. Conclusion

In this paper, a new approach to optimize a non-conventional DGS filter based on GA for wide-band antenna is presented. The main advantage of this method is the ability for automatically providing DGS filter structures to cancel some undesired bands. Hence, the WiMAX band (taken as example) is successfully rejected by using the proposed approach. The GAO code is developed under VBS and implemented in CST Microwave software. Measurement and simulation results confirm the effectiveness of the proposed approach which is also practical and easily applicable.

The future perspectives of our present work are the optimization of reconfigurable DGS filter in order to reject several bands. Consequently, the proposed antenna will demonstrate advanced flexibility in term of frequency functionalities.

## Acknowledgement

## References

[1] T. Abdurrahim, and M. Yerlikaya, "A compact reconfigurable ultra-wideband G-shaped printed antenna with band-notched characteristic," *Microwave and Optical Technology Letters*, 61.1, pp. 245-250, 2019. https://doi.org/10.1002/mop.31516

[2] R. Azim, M. T. Islam, and A. T. Mobashsher, "Dual Band-Notch UWB Antenna With Single Tri-Arm Resonator," *IEEE Antennas and Wireless Propagation Letters* 13, pp. 670-673, 2014. https://doi.org/10.1109/LAWP.2014.2314486

[3] L. C. Tsai, and W. J. Chen, "A UWB antenna with band-notched filters using slottype split ring resonators," *Microwave Opt. Technol. Lett.*, vol. 58, pp. 2595-2598, 2016. https://doi.org/10.1002/mop.30101

[4] F. Guichi, M. Challal and T. A. Denidni, "A Novel Dual Band-Notch Ultra-Wideband Monopole Antenna using Parasitic Stubs and Slot," *MOTL – Microw. and optic. Technol. Lett.*, vol. 60(7), pp. 1737-1744, Jul. 2018. https://doi.org/10.1002/mop.31231

[5] X.-K. Gao, H. M. Lee, and S.-P. Gao, "A Robust Parameter Design of Wide Band DGS Filter for Common-Mode Noise Mitigation in High-Speed Electronics," *IEEE Trans. on Electromagn. Compat.*, 59(6), pp. 1735-1740, 2017. https://doi.org/10.1109/TEMC.2017.2710202

[6] K. Fertas, F. Ghanem, M. Challal, and R. Aksas "Design and Development of Compact Reconfigurable Tri-Stopband Bandstop Filter Using Hexagonal Metamaterial Cells for Wireless Applications" *Progress In Electromagnetics Research M*, Vol. 80, pp. 93-102, 2019. https://doi.org/10.2528/PIERM1810230.

[7] J. B. Jadhav and P. J. Deore, "A compact planar ultra-wideband bandpass filter with multiple resonant and defected ground structure," *AEU-Intern. J. of Electron. and Commun.*, 81, pp. 31-36, 2017. https://doi.org/10.1109/APMC.2016.7931400

[8] K. Fertas, H. Kimouche, M. Challal, H. Aksas et al. "Design and optimization of a CPW-fed tri-band patch antenna using genetic algorithms," *ACES- Applied Computational Electromagnetics Society Journal*, vol. 30(7), 2015.

[9] S. Arianos, J. A., Quijano, G., Dassano, F. Vipiana et al. "Automated Design and Experimental Validation of a Reconfigurable, Board-Mounted Compact Antenna," *IEEE Antennas and Wireless Propag*. Lett. vol. 16, pp. 589-592, 2017. https://doi.org/10.1109/LAWP.2016.2591326

[10] K. Fertas, S. Tebache, F. Ghanem, S. Tedjini and R. Aksas, "Non-conventional Multiband Patch Antenna Design with Filtering Aspect Based on Genetic Algorithm," *IETE Journal of Research*. vol.66, n 6,pp. 815-822, 2020.https://doi.org/10.1080/03772063.2019.1634496.

[11] J. W. Jayasinghe, J. Anguera, D. N. Uduwawala and A. Andújar "A multipurpose genetically engineered microstrip patch antennas: Bandwidth, gain, and polarization," *Microw. and Opt. Technol. Lett.*, vol. 59(4), pp. 941-949, 2017. https://doi.org/10.1002/mop.30439

[12] K. Fertas, H. Kimouche, M. Challal, et al., "Development of a Novel UWB Planar Antenna using a Genetic Algorithm," *IEEE – 5th ICEE, Boumerdes, Algeria*, pp. 29-31 October 2017. https://doi.org/10.1109/ICEE-B.2017.8192102

[13] X.-L. Liu, Y.-Z. Yin, and J.-H. Wang, "A compact dual band-notched UWB antenna with meandering slot and C-shape strips," *Microw. Opt.Technol. Lett.*, vol. 55, no. 11, pp. 2631–2636, 2013. https://doi.org/10.1002/mop.27896

[14] X.-L. Liu, Y.-Z. Yin, P. A. Liu, J.-H. Wang, and B. Xu, "A CPW-fed dual band-notched UWB antenna with a pair of bended dual-L-shape parasitic branches," *Prog. Electromagn. Res*., vol. 136, pp. 623–634, 2013. https://doi.org/10.2528/PIER12122507

[15] C. Wang, Z.-H. Yand, B. Li, and P. Xu, "A dual band-notched UWB printed antenna with C-shaped and a U-shaped slot," *Microw. Opt.Technol. Lett*., vol. 54, no. 6, pp. 1450–1452, 2012. https://doi.org/10.1002/mop.26807

**Khelil Fertas** received the engineering degree in 2006 from University of M'sila, Algeria, and the MS degree in January 2014 from Ecole Militaire Polytechnique, Algiers, Algeria. In 2021, he received the PhD from Ecole Nationale Polytechnique (ENP), Algeria. His fields of interest include optimization, RF passive and active circuits and antennas.

**Soufiane Tebache** received the Engineering degree from Polytechnic School, Algeria, in 2009, with honors (first class), and the Magister degree in Signals & Communications from Biskra University, in 2013 with honors as well (first class). He defended recently his PhD thesis in telecommunications at Ecole Nationale Polytechnique (ENP), Algeria. His current research interests include: Antenna & RF system- design, Microwaves, Electromagnetic field theory and Propagation, MIMO systems, Wireless communications, IoT and 5G.

**Fouad Fertas** received the M.S. and Ph.D degrees from Boumerdes University, Algeria, respectively in 2016 and 2020. His fields of interest include Propagation and Antennas.

**Rabia Aksas** was born in 1950, Algeria. He received the MS and Doctorate degrees in electronics from Ecole Nationale Polytechnique, Algiers, Algeria, respectively in 1982 and 1995. Currently, he is a full Professor in the Electronics Department at Ecole Nationale Polytechnique. His research interests are in the areas of electromagnetic field theory, microwaves, propagation and antennas.

# Generalized Givens Rotations Applied to Complex Joint Eigenvalue Decomposition

Ammar MESLOUB

**Abstract—This paper shows the different ways of using generalized Givens rotations in complex joint eigenvalue decomposition (JEVD) problem. It presents the different schemes of generalized Givens rotation, justifies the introduced approximations and focuses on the process of extending an algorithm developed for real JEVD to the complex JEVD. Several Joint Diagonalization problem use generalized Givens rotations to achieve the solution, many algorithms developed in the real case exist in the literature and are not generalized to the complex case. Hence, we show herein a simple and not trivial way to get the complex case from the real one. Simulation results are provided to highlight the effectiveness and behaviour of the proposed techniques for different scenarios.**

*Keywords***—Complex Joint EigenValue Decomposition (JEVD), Shear and SHRT, generalized Givens rotations, exact JEVD, approximative JEVD.**

## I. INTRODUCTION

In this paper, we mainly try to expose the generalized Givens rotation used to solve the Joint EigenValue Decomposition (JEVD) problem of a set of complex square matrices. This problem of JEVD can be found in different applications such as multi-dimensional harmonic retrieval [1], Canonical Polyadic Decomposition (CPD) of tensors [2, 3], Direction of arrival estimation [4], joint angle-delay estimation [5] and Blind Sources Separation (BSS) [6].

The aforementioned problem is widely treated in the literature by using different schemes. The generalized Givens rotations have been used in [7, 8], the LU decomposition is used in [9] and first Taylor approximation is used in [10].

The JEVD is defined for $K$ complex square matrices of dimension $N \times N$ sharing the following structure :

$$\mathbf{M}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^{-1} \tag{1}$$

Where $k \in \{1, ..., K\}$, $K$ is the number of matrices, $N$ is the matrix dimension. $\mathbf{A}$ is unknown square non defective matrix (in the BSS context, this matrix is referred to as mixing matrix) and $\mathbf{D}_k$ is the $k^{th}$ diagonal matrix associated to the $k^{th}$ matrix $\mathbf{M}_k$.

The last problem consists of looking for $\{\mathbf{A}, \mathbf{D}_1, \cdots, \mathbf{D}_K\}$ by using only the set of the $K$ complex matrices $\{\mathbf{M}_1, \cdots, \mathbf{M}_K\}$. It can be also defined by finding a matrix $\mathbf{V}$ which makes the set of matrices $\{\mathbf{VM}_1\mathbf{V}^{-1}, \cdots, \mathbf{VM}_K\mathbf{V}^{-1}\}$ as diagonal as possible specially in the approximate JEVD case (see section V. for more details).

In this paper, we investigate generalized Givens rotations applied to the JEVD problem. The unknown matrix $\mathbf{A}$ is decomposed in

a product of generalized Givens rotations, according to:

$$\mathbf{A} = \prod_{\#\text{sweeps}} \prod_{1 \leq i < j \leq N} \mathbf{S}_{ij}\mathbf{G}_{ij} \tag{2}$$

where $\#$sweeps represents the number of iterations, $\mathbf{S}_{ij}$ and $\mathbf{G}_{ij}$ are the elementary Shear and Givens rotations [7, 11], respectively. The problem of JEVD reduces then in the estimation of these elementary rotations.

## II. GENERALIZED GIVENS ROTATION SCHEMES

The elementary rotations of equation (2) can be expressed according to two different schemes. The first one is based on sinus and hyperbolic sinus which needs some complicated developments leading to an efficient estimation.
This first scheme considers elementary rotations that are equal to the identity matrix except for $(i,i)^{th}$, $(i,j)^{th}$, $(j,i)^{th}$ and $(j,j)^{th}$ entries which are:

$$\begin{bmatrix} G_{ij}(i,i) & G_{ij}(i,j) \\ G_{ij}(j,i) & G_{ij}(j,j) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & e^{-j\varphi}\sin(\theta) \\ -e^{j\varphi}\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} S_{ij}(i,i) & S_{ij}(i,j) \\ S_{ij}(j,i) & S_{ij}(j,j) \end{bmatrix} = \begin{bmatrix} \cosh(y) & e^{j\alpha}\sinh(y) \\ e^{-j\alpha}\sinh(y) & \cosh(y) \end{bmatrix} \tag{4}$$

$\theta$ is the angle parameter and $\varphi$ is the phase parameter of considered Givens rotation. $y$ is the shear angle and $\alpha$ represents the phase parameter of the elementary Shear rotation. This scheme is the most used one as it can be found in [7, 8, 11–14].

The second scheme allows appropriate approximations that lead to a computationally simplified methods compared to the first scheme.This scheme allows elementary Shear and Givens rotations equal to the identity matrix except for some entries which are given by:

$$\begin{bmatrix} G_{ij}(i,i) & G_{ij}(i,j) \\ G_{ij}(j,i) & G_{ij}(j,j) \end{bmatrix} = \lambda_g \begin{bmatrix} 1 & g^* \\ -g & 1 \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} S_{ij}(i,i) & S_{ij}(i,j) \\ S_{ij}(j,i) & S_{ij}(j,j) \end{bmatrix} = \lambda_s \begin{bmatrix} 1 & s^* \\ s & 1 \end{bmatrix} \tag{6}$$

where $g$ and $s$ are complex numbers associated to Givens and Shear parameters, respectively. $g^*$ is the complex conjugate of $g$. $\lambda_g = \frac{1}{\sqrt{1+|g|^2}}$ and $\lambda_g = \frac{1}{\sqrt{1-|s|^2}}$. When this scheme is applied as in [14, 15], both $g$ and $s$ are considered with small norm which simplifies $\lambda_g$ and $\lambda_s$ to the unity ($\lambda_g = \lambda_s \approxeq 1$).

### III. JOINT EIGENVALUE DECOMPOSITION CRITERIA

The JEVD problem uses several criteria as shown in the literature, the difference of these criteria can be seen only in the shear rotation. Before discussing the existing criteria, let us note $\mathbf{M}'_k$ the transformed matrices by the shear rotation as

$$\mathbf{M}'_k = \mathbf{S}_{ij(y,\alpha)} \mathbf{M}_k \mathbf{S}_{ij(y,\alpha)} \qquad (7)$$

The different criteria can be discussed as

- In [12], the Shear rotation is chosen to minimize the frobenius norm of transformed matrices $\mathbf{M}'_h$, where the index $h$ is chosen by the matrix having the most departure from the normality (see [12] for more details). Then, some approximations and linearisation are introduced to simplify the explicit expression of the optimal value of '$y$'.

$$C_1(y) = \left\| \mathbf{M}'_h \right\|^2 \qquad (8)$$

This criterion $C_1(y)$ is used for real JEVD.

- In [7], the criterion of shear transformation takes into account the frobenius norm of off diagonal matrices, which is the frobenius norm of matrices $\mathbf{M}_k - \text{diag}(\mathbf{M}_k)$, $k = 1, \cdots, K$. the Shear parameter '$y$' is obtained by an analytical solution.

$$C_2(y, \alpha) = \sum_{k=1}^{K} \left\| \mathbf{M}'_k - \text{diag}(\mathbf{M}'_k) \right\|^2 \qquad (9)$$

This criterion $C_2(y, \alpha)$ is used in the complex case.

- In [8], the considered criterion is the sum of square norm of the $(i,j)^{th}$ and $(j,i)^{th}$ entries of matrices $\mathbf{M}_k$, $k = 1, \cdots, K$. The fixed couple $i$ and $j$ satisfies $1 \leq i < j \leq N$. This criterion is a simplified version of the one given in equation (9). However, JDTM (Joint Diagonalization algorithm based on Targeting hyperbolic Matrices) outperforms JUST (Joint Unitary Shear Transformation) in many scenarios (see [8] for details).

$$C_3(y) = \sum_{k=1}^{K} |m'_k(i,j)|^2 + |m'_k(j,i)|^2 \qquad (10)$$

- In [13], the considered criterion is the sum of square norm of difference between the $(i,j)^{th}$ and $(j,i)^{th}$ entries of matrices $\mathbf{M}_k$, $k = 1, \cdots, K$. This criterion is introduced to reduce the deviation of matrices $\mathbf{M}_k$ from normality. the Shear rotation is explicitly introduced to minimize the departure from symmetry of matrices $\mathbf{M}_k$. Hence, for each elementary rotation $\mathbf{S}_{ij}(y)$, the minimized criterion can be expressed as:

$$C_4(y) = \sum_{k=1}^{K} \sum_{1 \leq i < j \leq N} |m'_k(i,j) - m'_k(i,j)|^2 \qquad (11)$$

Now, let us generalize the SHRT (SHear RoTation algorithm) method developed in [12] to the complex case.

### IV. COMPLEX SHRT METHOD

The objective of this section is to generalize the method proposed in [12] to the complex case. Hence, only the Shear rotation is studied where the Givens rotation is obtained by applying the solution given in [6].

Let us study the transformed matrices by the Shear transform given in (7), where only $i^{th}$, $j^{th}$ rows and columns are affected by the Shear transformation. Modified entries can be written as:

$$\begin{array}{ll} m'_k(l,i) = & m_k(l,i)\cosh(y) + m_k(l,j)e^{-\jmath\alpha}\sinh(y) \\ m'_k(l,j) = & m_k(l,i)e^{\jmath\alpha}\sinh(y) + m_k(l,j)\cosh(y) \\ m'_k(i,l) = & m_k(i,l)\cosh(y) - m_k(j,l)e^{\jmath\alpha}\sinh(y) \\ m'_k(j,l) = & -m_k(i,l)e^{-\jmath\alpha}\sinh(y) + m_k(i,j)\cosh(y) \end{array} \qquad (12)$$

and the entries twice affected by the shear rotation can be expressed as:

$$\begin{array}{ll} m'_h(i,j)e^{-\jmath\alpha} & = m_h(i,j)e^{-\jmath\alpha} + \xi_h\sinh^2(y) + \frac{d_h}{2}\sinh(2y) \\ m'_h(j,i)e^{\jmath\alpha} & = m_h(j,i)e^{\jmath\alpha} - \xi_h\sinh^2(y) - \frac{d_h}{2}\sinh(2y) \\ m'_h(i,i) & = m_h(i,i) + d_h\sinh^2(y) + \frac{\xi_h}{2}\sinh(2y) \\ m'_h(j,j) & = m_h(j,j) - d_h\sinh^2(y) - \frac{\xi_h}{2}\sinh(2y) \end{array} \qquad (13)$$

where

$$\begin{cases} \xi_h = & (m_k(i,j)e^{-\jmath\alpha} - m_k(j,i)e^{\jmath\alpha}) \\ d_h = & (m_k(i,i) - m_k(j,j)) \end{cases} \qquad (14)$$

Note that we have introduced $m'_h(i,j)e^{-\jmath\alpha}, m'_h(j,i)e^{\jmath\alpha}$ instead of $m'_h(i,j), m'_h(j,i)$ in order to simplify the other workouts.

Let us find the optimal shear parameter $y$ and $\alpha$ minimizing the square frobinus norm of matrix $\mathbf{M}'_h$ as given in equation (8). We take into account only the modified entries given in (12) and (13), then the total criterion can be expressed as

$$C_T(y, \alpha) = C_{T1}(y, \alpha) + C_{T2}(y, \alpha) \qquad (15)$$

where the first term, $C_{T1}(y, \alpha)$ corresponds to the entries twice affected given in equation (13) as

$$\begin{array}{ll} C_{T1}(y, \alpha) = & \frac{|m'_h(i,i)+m'_h(j,j)|^2}{2} + \frac{|m'_h(i,i)-m'_h(j,j)|^2}{2} \\ & + \frac{|m'_h(i,j)e^{-\jmath\alpha}+m'_h(j,i)e^{\jmath\alpha}|^2}{2} \\ & + \frac{|m'_h(i,j)e^{-\jmath\alpha}-m'_h(j,i)e^{\jmath\alpha}|^2}{2} \end{array} \qquad (16)$$

The second term $C_{T2}$ contains the other terms given by

$$\begin{array}{ll} C_{T2}(y, \alpha) = & \sum_{l=1,l\neq i,j}^{N} |m'_h(i,l)|^2 + |m'_h(j,l)|^2 \\ & + |m'_h(l,i)|^2 + |m'_h(l,j)| \end{array} \qquad (17)$$

Let us analyse the different parts of $C_{T1}(y, \alpha)$ given equation (16), the part $\frac{|m'_h(i,i)+m'_h(j,j)|^2}{2}$ is independent from the shear parameter $y$ and $\alpha$, the part $\frac{|m'_h(i,j)e^{-\jmath\alpha}+m'_h(j,i)e^{\jmath\alpha}|^2}{2}$ depends only on the phase parameter $\alpha$ and the other parts depend on the two parameters. Then, we have to minimize the part $\frac{|m'_h(i,j)e^{-\jmath\alpha}+m'_h(j,i)e^{\jmath\alpha}|^2}{2}$ to get optimal phase parameter. The optimal value is

$$\alpha = \frac{1}{2}\left[\arg\{m_h(i,j)m^*_h(j,i)\} - \pi\right] \qquad (18)$$

Once this value is computed, we insert it in equation (15) in order to get the shear rotation parameter $y$. After some workouts,

the total criterion can be rearranged as

$$
\begin{aligned}
C_T(y) = \ & (|d_h|^2 + |\xi_h|^2)\sinh^2(2y) + \Re(\xi_h^* d_h)\sinh(4y) \\
& + G_h(\cosh(2y) - 1) - 2\Re(K_h e^{j\alpha})\sinh(2y) \\
& + \beta
\end{aligned}
\tag{19}
$$

where

$$
\begin{cases}
G_h = & \sum_{l \neq i,j} |m_h(i,l)|^2 + |m_h(j,l)|^2 \\
& + |m_h(l,i)|^2 + |m_h(l,i)|^2 \\
K_h = & \sum_{l \neq i,j} m_h(i,l)m_h^*(j,l) - m_h^*(l,i)m_h(l,j)
\end{cases}
\tag{20}
$$

and $\beta$ is a constant independent from $y$ and $\alpha$.

We differentiate equation (19) with respect to $y$ and make a linear approximation to the zeros of this derivative. We get

$$
y = \operatorname{arctanh}\left(\frac{\Re(K_h e^{j\alpha} - \xi_h^* d_h)}{2(|d_h|^2 + |\xi_h|^2) + G_h}\right)
\tag{21}
$$

Finally, the developed algorithm, referred to as Complex SHRT (CSHRT), is given in Table I.

**Table. I**
CSHRT ALGORITHM

---

**Require** : $\mathbf{M}_k$, $k = 1, \cdots, K$, fixed threshold $\tau$
   and maximum sweep number $M_{it}$.
**Initialization**:   $\mathbf{V} = \mathbf{I}_N$ and $\mathbf{A} = \mathbf{I}_N$.
 **while** $\max_{i,j}(|y|, |\theta|) > \tau$ and (#sweeps $< M_{it}$)
   **for** all $1 \leq i < j \leq N$
     Unitary transform
       Estimate $\mathbf{G}_{ij}(\theta, \varphi)$ using solution given in [6].
       Updates matrices $\mathbf{V}, \mathbf{M}_k$ using equation (22).
     Shear transform
       Estimate the matrix $\mathbf{M}_h$ which has the maximum
         departure to the normality.
       Compute the phase parameter $\alpha$ using (18).
       Compute Shear parameter $y$ using equation (21).
       Matrices updates using equation (23)
   **end for**
**end while**.

---

The proposed CSHRT algorithm can be summarized as follows. Given the complex matrices $\mathbf{M}_k$ $k \in \{1, ..., K\}$, an iterative scheme is applied to get the matrix $\mathbf{V}$ which is the left inverse of the mixing matrix $\mathbf{A}$ given in equation (1). One iteration is achieved by successive unitary and Shear transformations. Using the solution given in [6], the unitary transformation can be applied to get $\mathbf{G}(\theta, \varphi)$. The last transformation is used to update the matrices according to

$$
\begin{cases}
\mathbf{V} & \leftarrow & \mathbf{V}\mathbf{G}(\theta, \varphi) \\
\mathbf{M}_k & \leftarrow & \mathbf{G}(\theta, \varphi)^H \mathbf{M}_k \mathbf{G}(\theta, \varphi)
\end{cases}
\tag{22}
$$

where $\mathbf{V}$ is the estimated left inverse of the mixing matrix $\mathbf{A}$. The Shear transformation is achieved by computing the shear parameters $\alpha$ and $y$. The phase parameter $\alpha$ is obtained by using equation (18) and the angle parameter $y$ is computed by equation

(21). Once these parameters are obtained, the first scheme of shear rotation is applied to update the different matrices according:

$$
\begin{cases}
\mathbf{V} & \leftarrow & \mathbf{V}\mathbf{S}(y, \alpha) \\
\mathbf{M}_k & \leftarrow & \mathbf{S}(-y, \alpha)\mathbf{M}_k\mathbf{S}(y, \alpha)
\end{cases}
\tag{23}
$$

The overall proposed algorithm, named complex SHRT (CSHRT), is summarized in Table I. Note that the last algorithm is the first version of CSHRT, other modifications can be introduced to get the other versions.

The second version can be developed by taking into account only the entries twice affected given in equation (13) which leads to minimize the criterion $C_{T1}(y, \alpha)$ given in equation (16). The phase parameter $\alpha$ is the same as given in (18) but the shear angle $y$ is obtained by:

$$
y = \operatorname{arctanh}\left(-\frac{\Re(\xi_h^* d_h)}{2(|d_h|^2 + |\xi_h|^2)}\right)
\tag{24}
$$

The third version is realized by differentiating equation (19) with respect to $y$, $\alpha$ and making a linear approximations to the zeros of obtained derivatives. The formula of $y$ is kept unchanged as in (21) but the phase parameter is:

$$
\alpha = \arg\left(\sum_{l=1}^{N} m_h(i,l)m_h^*(j,l) - m_h(l,i)^* m_h(l,j)\right) - \pi
\tag{25}
$$

CSHRT$_1$ is the version given in Table I, the second version noted CSHRT$_2$ is the same as CSHRT$_1$ except for $y$ which is computed by (24) instead of (21). The last version CSHRT$_3$ is also the same as CSHRT$_1$ except for the phase parameter which is obtained by equation (25) instead of equation (18).

## V. SIMULATION, RESULTS AND DISCUSSIONS

In this section, we have tested the different proposed algorithm versions and compared with respect to JDTM and SJD (Simple Joint Diagonalization algorithm) given in [8, 14] respectively for different cases.

In the first case, the target matrices satisfy exactly the equation (1), this case is named the exact JEVD case. The aim of this part is to show the convergence rate of each algorithm.

In the second case, the target matrices did not satisfy the equation (1) which leads to the case of approximative JEVD. The objective of this case is to get the noise robustness of each algorithm.

The used Performance Index (PI) is the same as in [7, 16] evaluated over 100 Monte-Carlo realizations. The definition of PI is

$$
\begin{aligned}
\text{PI}(\mathbf{Q}) = \ & \frac{1}{2N(N-1)}\sum_{n=1}^{N}\left(\sum_{m=1}^{N}\frac{|q(n,m)|^2}{\max_k |q(n,k)|^2} - 1\right) \\
& + \\
& \frac{1}{2N(N-1)}\sum_{n=1}^{N}\left(\sum_{m=1}^{N}\frac{|q(m,n)|^2}{\max_k |q(k,n)|^2} - 1\right)
\end{aligned}
\tag{26}
$$

where $\mathbf{Q} = \hat{\mathbf{V}}\mathbf{A}$ is the global matrix. For a tested algorithm, when the obtained PI value is close to zero, it means that the JEVD quality reached is good.
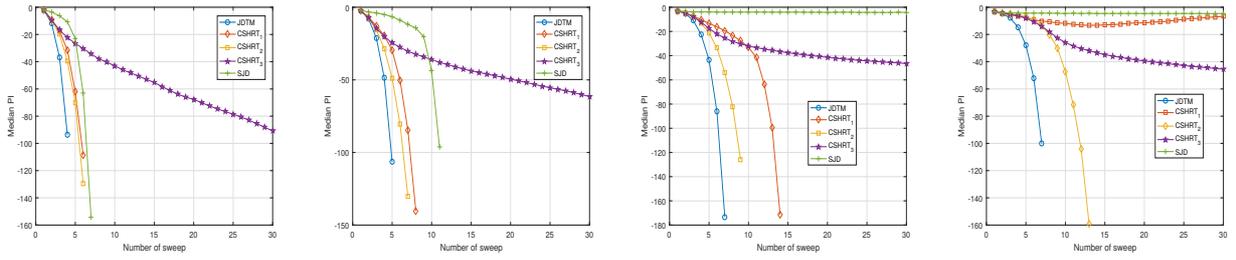
(a) Case of $K = 5$ and $N = 5$    (b) Case of $K = 5$ and $N = 10$ (c) Case of $K = 5$ and $N = 25$ (d) Case of $K = 5$ and $N = 35$

**Fig. 1**: Mean PI versus sweep number in exact JEVD for different matrix dimensions

*A. Exact JEVD case*

In this case, all matrices $\mathbf{A}$ and $\{\mathbf{D}_k\}_{k=1,...,K}$ are generated by considering normal distribution, independent and complex for all entries. Hence, the target matrices $\{\mathbf{M}_k\}_{k=1,...,K}$ are obtained by using equation (1). All considered algorithms are applied to these matrices in order to realize the JEVD.

We have realized the JEVD of five matrices ( $K = 5$) and varying the matrix dimension $N$. $N$ takes these values $\{5, 10, 25, 35\}$. The Figure 1 shows obtained results. Note that as the number of matrices $K$ is less important compared with the matrices dimension $N$, as the JEVD difficulty increases. However, the considered algorithms are differently affected by this difficulty. JDTM is the less affected one and presents the best convergence rate.

The SJD algorithm, as shown in Figures 1(c) and 1(d), diverges completely when the ratio $\frac{K}{N}$ is less than 20% due to the introduced approximations considered by this algorithm as explained in [14]. Note that this algorithm uses the second scheme of generalized Givens rotations by considering small changes ($\lambda_s = \lambda_g \approxeq 1$). It minimizes the simplified criterion given (9) which introduces this divergence for difficult JEVD.

The performance of the different proposed version of CSHRT are in-between those of SJD and JDTM. This is due to using only one target matrix to estimate shear parameters instead of considering all target matrices. Note that the second version is the most efficient when the CSHRT$_3$ is the worst version which implies that the phase parameter $\alpha$ is better estimated with equation (18) instead of formula (25). By analyzing curves of figure 1, CSHRT$_1$ and CSHRT$_2$ have quadratic convergence where CSHRT$_3$ has a linear convergence.

The second comparison can be done between CSHRT$_1$ and CSHRT$_2$. The difference between these versions is the computation of the shear angle $y$. The first version CSHRT$_1$ minimizes the total criteria given in (15) where the second version CSHRT$_2$ minimizes the simplified criterion given (16). Surprisingly, the second version is better than the first version, it can be seen in curves of figures 1(c) and 1(d).

*B. Approximative JEVD case*

In this case, the target matrices are generated using the equation given in (27) as:

$$\mathbf{M}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}^{-1} + \mathbf{\Xi}_k \qquad (27)$$
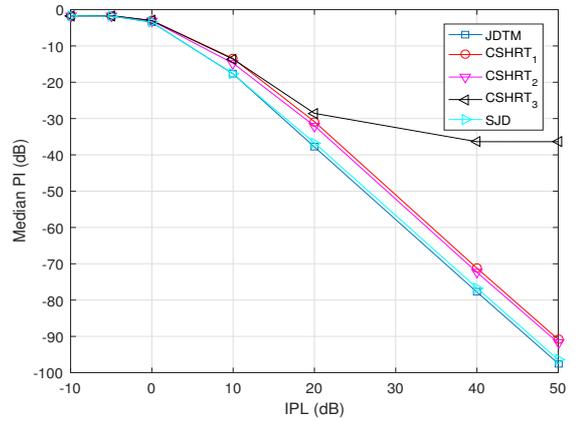


**Fig. 2**: Median PI versus perturbation level in approximate JEVD ($K = N = 5$)

where $\mathbf{\Xi}_k$ is a noise matrix. The inverse perturbation level (IPL) can be measured by:

$$PL(dB) = \frac{\left\| \mathbf{A}\mathbf{D}_k\mathbf{A}^{-1} \right\|_F}{\left\| \mathbf{\Xi}_k \right\|_F} \qquad (28)$$

The noise matrix $\mathbf{\Xi}_k$ is generated as $\mathbf{\Xi}_k = \varrho_k \mathbf{\Upsilon}_k$ where $\mathbf{\Upsilon}_k$ is a random matrix (generated at each Monte Carlo run) and $\varrho_k$ is a positive number allowing to tune the perturbation level, it can be seen as kind of matrix SNR.
Algorithms' performance are evaluated according to the IPL for $K = 5$ and $N = 5$. Obtained results are given in Figure 2 where PI versus IPL curves are shown. Note that all algorithms reach approximatively the same performance as it can be seen in Figures 2 except for CSHRT$_3$ where the phase is less estimated. The difference can be seen in the convergence rate as illustrated in subsection A..

VI. CONCLUSION

In this paper, the JEVD problem using generalized Givens rotation is considered in the complex case. We have seen the different JEVD criteria and rotation schemes. We have generalized SHRT algorithm to the complex case by considering three versions. Surprisingly, the second version is the most efficient compared to the other ones. By considering the results presented in the simulation section, one can deduce that minimizing the simplified criteria leads to most efficient estimation of shear

parameters ($\alpha$ and $y$). The presented algorithm is less efficient compared to JDTM, this is due to the introduced simplifications and approximations. These simplifications can be listed as: only one target matrix is considered for shear rotation and the shear parameter $y$ is obtained by using some linear approximations. However, the proposed algorithm seem to be efficient with relatively low computational cost.

## REFERENCES

[1] M. Haardt and J. A Nossek, "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE-Tr-SP*, vol. 46, no. 1, 1998. https://doi.org/10.1109/78.651206

[2] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM journal on Mat. Ana. and App.*, vol. 26, no. 2, 2004. https://doi.org/10.1137/S089547980139786X

[3] X. Luciani and L. Albera, "Canonical polyadic decomposition based on joint eigenvalue decomposition," *Chemometrics and Intelligent Laboratory Systems*, vol. 132, 2014.

[4] A-J Van Der Veen, P. B. Ober, and E. F. Deprettere, "Azimuth and elevation computation in high resolution DOA estimation," *IEEE-Tr-SP*, vol. 40, no. 7, 1992. https://doi.org/10.1109/78.143456

[5] A. N. Lemma, A-J Van Der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE-Tr-SP*, vol. 51, no. 5, 2003. https://doi.org/10.1109/TSP.2003.810306

[6] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE-Tr-SP*, 1997. https://doi.org/10.1109/78.554307

[7] R. Iferroudjene, K. Abed-Meraim, and A. Belouchrani, "A new Jacobi-like method for joint diagonalization of arbitrary non-defective matrices," *Appl. Math. and Comp.*, vol. 211, no. 2, May 2009. https://doi.org/https://doi.org/10.1016/j.amc.2009.01.045

[8] X. Luciani and L. Albera, "Joint eigenvalue decomposition using polar matrix factorization," in *LVA/ICA*. Springer, 2010.

[9] X. Luciani and L. Albera, "Joint eigenvalue decomposition of non-defective matrices based on the LU factorization with application to ICA," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, 2015. https://doi.org/10.1109/TSP.2015.2440219

[10] Rémi André, Xavier Luciani, and Eric Moreau, "Joint Eigen-Value Decomposition Algorithms Based on First-Order Taylor Expansion," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1716–1727, 2020. https://doi.org/10.1109/TSP.2020.2976580

[11] A. Souloumiac, "Nonorthogonal Joint Diagonalization by Combining Givens and Hyperbolic Rotations," *IEEE-Tr-SP.*, June 2009. https://doi.org/10.1109/TSP.2009.2016997

[12] T. Fu and X. Gao, "Simultaneous diagonalization with similarity transformation for non-defective matrices," in *IEEE ICASSP Proceedings*, May 2006, vol. 4. https://doi.org/

[13] A. Mesloub, A. Belouchrani, and K. Abed-Meraim, "Efficient and stable joint eigenvalue decomposition based on generalized Givens rotations," in *EUSIPCO*, Italy, Sep. 2018. https://doi.org/10.1109/ICASSP.2006.1661174

[14] Ammar Mesloub, Adel Belouchrani, and Karim Abed-Meraim, "New algorithms on complex joint eigenvalue decomposition based on generalized Givens rotations," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5. https://doi.org/10.23919/EUSIPCO.2018.8553362

[15] Jifei Miao, Guanghui Cheng, Yunfeng Cai, and Jing Xia, "Approximate joint singular value decomposition algorithm based on givens-like rotation," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 620–624, 2018. https://doi.org/10.1109/LSP.2018.2815584

[16] A. Mesloub, K. Abed-Meraim, and A. Belouchrani, "A new algorithm for complex non-orthogonal joint diagonalization based on Shear and Givens rotations," *IEEE Tr-SP*, vol. 62, no. 8, April 2014. https://doi.org/10.1109/TSP.2014.2303947

**Ammar MESLOUB** received the Engineer's and Master's degrees in telecommunications from EMP, Algiers, Algeria, in 2006 and 2009 respectively. He received a Ph.D. in signal processing from EMP. He is an associate professor in the signal processing laboratory, EMP. His research interests are in signal processing, time-frequency analysis, and time-frequency array signal processing with applications in radar and communications.

# Surface EMG signal segmentation based on HMM modelling: Application on Parkinson's disease

Hichem Bengacemi, Abdenour Hacine Gharbi, Philippe Ravier, Karim Abed-Meraim, and Olivier Buttelli

**Abstract**—The study of burst electromyographic (EMG) activity periods during muscles contraction and relaxation is an important and challenging problem. It can find several applications like movement patterns analysis, human locomotion analysis and neuromuscular pathologies diagnosis such as Parkinson disease. This paper proposes a new frame work for detecting the onset (start) / offset (end) of burst EMG activity by segmenting the EMG signal in regions of muscle activity (AC) and non activity (NAC) using Discrete Wavelet Transform (DWT) for feature extraction and the Hidden Markov Models (HMM) for regions classification in AC and NAC classes. The objective of this work is to design an efficient segmentation system of EMG signals recorded from Parkinsonian group and control group (healthy). The results evaluated on ECOTECH project database using principally the Accuracy (Acc) and the error rate (Re) criterion show highest performance by using HMM models of 2 states associated with GMM of 3 Gaussians, combined with LWE (Log Wavelet decomposition based Energy) descriptor based on Coiflet wavelet mother with decomposition level of 4. A comparative study with state of the art methods shows the efficiency of our approach that reduces the mean error rate by a factor close to 2 for healthy subjects and close to 1.3 for Parkinsonian subjects.

**Keywords**—surface EMG signal, EMG signal segmentation, muscle activity, wavelet analysis, HMM models, Parkinson disease.

## NOMENCLATURE

| | |
|---|---|
| sEMG | surface Electromyographic. |
| HMM | Hidden Markov Modelling. |
| GMM | Gaussian Mixture Models. |
| DWT | Discrete Wavelet Transform. |
| DWE | Discrete Wavelet decomposition based calculus Energy. |
| LWE | Log Wavelet decomposition based Energy. |
| WCC | Wavelet Cepstral Coefficients. |
| LPC | Linear Prediction Coefficients. |
| HTK | Hidden Markov Model Toolkit. |
| NAC | No-Activity. |
| AC | Activity. |
| MUAP | Motor Unit Action Potentials. |
| STD | Standard Deviation. |
| ALE | Absolute Latency Error. |
| Acc | classification accuracy. |
| Re | Error Rate. |

*Manuscript April 21, 2021; revised July 18, 2021.*

*Hichem Bengacemi is with the Signal Processing Lab, École Militaire Polytechnique, Algiers, Algeria (e-mail: bengacemi-hichem@yahoo.fr).*

*Abdenour Hacine Gharbi is with the LMSE lab, University of Bordj Bou Arreridj, Algeria. (e-mail: gharbi07@yahoo.fr).*

*Philippe Ravier, Karim Abed-Meraim and Olivier Buttelli are with the PRISME Laboratory, University of Orléans, 12 Rue de Blois, 45067, Orléans, France (e-mail: philippe.ravier@univ-orleans.fr, karim.abed-merim@univ-orleans.fr, olivier.buttelli@univ-orleans.fr)*

## I. INTRODUCTION

Muscle activity generates electric fields that can be picked up by means of electrodes located on the skin above the muscle under study. This signal and this technique are referred to as surface electromyographic activity (sEMG). Electromyography has been widely used to gain fundamental knowledge on neuromuscular control and muscle operating conditions (central and peripheral properties of this system, respectively) [1]. sEMG is also a relevant tool for the kinesiological analysis of movement disorders and for the evaluation of gait and posture such as for Parkinson's disease [2]. Accurate determination of the sEMG burst activation timing is an important consideration in motion analysis and remains a challenge and all the more so as sEMG recording are being done from patient with motor disorder. Timing features of the burst activation are the starting (onset) and the ending (offset) points which are relevant indicators to define motor coordination and its modifications during dynamic exercise such as during walking activity. This paper is dedicated to this key task, known in the literature under different narmes: signal segmentation or signal activity monitoring, or even burst detection.

### A. Related work

The automatic methods for EMG signal segmentation can be divided into two categories. The first category covers the unsupervised methods like single and double threshold based methods [3, 4], statistical optimal decision-based methods [5–8], probabilistic-based methods [9] where the authors characterize the muscle activity using EMG burst presence probability (EBPP) estimated from HMM parameters, wavelet transform-based methods [10], clustering based methods [11], nonlinear energy operator-based methods [12–14] and adaptive energy method [15]. Performance of these methods considerably vary as they have different properties, computational complexities and tuning parameters. Moreover, all these methods are user-

dependent since at least one of their tuning parameters must be set by the user for each considered sEMG signal.

The second category covers the supervised methods which learn from signals with segmentations (labels). In order to achieve the learning task, these methods need labelled dataset for the training phase. A testing dataset is used in a second phase for evaluating the performance of the method. The authors in [16] proposed to select the best one among a pool of onset estimators, given the measured features. The authors in [17] use Gaussian Mixture Models (GMM) clustering and Ant colony classifier (AntCC) as supervised method to automatize the processing of a large amount of sEMG signals. When applied to sEMG signals, these approaches clearly show improvement with respect to unsupervised methods. The proposed work falls within this category where we exploit the ECOTECH dataset [18] with Hidden Markov Modelling (HMM) to construct an automatic segmentation system for muscle activity monitoring.

### B. Our contribution

The proposed system is based on HMM modelling combined with feature extraction based on Discrete Wavelet Transform (DWT). Several features descriptors extracted from discrete wavelet analysis have been studied such as Discrete Wavelet decomposition based calculus Energy (DWE), Log Wavelet decomposition based Energy (LWE) and Wavelet Cepstral Coefficients (WCC). The main task consists of looking for optimal parameters of HMM and wavelets descriptors to achieve the best EMG activity monitoring performance.

The HMMs have been widely investigated and employed in the automatic speech recognition. Recently, it has been successfully used for both medical monitoring and diagnosis system applications such as ECG classification [19], EEG classification [20], electrical appliances identification [21]. Especially, this method is also used for the PD classification using the raw gait data [22]. The HMM has been also combined with support vector machine (SVM) classifier for natural gesture recognition using EMG signals for upper limb prostheses control [23]. In [24], the authors have combined the HMM and the multilayer perceptrons (MLP's) for classifying six motions based on EMG signals. A HMM based classifier is used for speech recognition using my-oelectric signals from the muscles of vocal articulation [25]. In [9], the authors have used the HMM on EMG signals to measure the EMG burst presence probability (EBPP). the study was limited to simulated signals and to one experimental signal just for illustration purpose. In this paper, we use HMM to classify EMG activity versus no EMG activity.

In brief, the main contributions of our work are: firstly, this work exploits the principal advantage of wavelet decomposition that is better adapted than Fourier decomposition for extracting the impulsive information of the action potentials (AP) of the motor units (MU). Secondly, this work adapts the HMM to automatic EMG signal segmentation . Note that HMM is one of the best tools to model signal state transitions and, to the best of our knowledge, this is the first time it is used for the considered supervised segmentation task. Finally, based on the ECOTECH data base, we provide a thorough performance analysis and comparative study with the state of the art methods.

The paper is organized as follows: section II. presents the problem formulation and the proposed method. Section III. is dedicated to the performance analysis and discussions while section IV. is dedicated to the concluding remarks.

## II. MATERIAL AND METHOD

In this section, we present in details the methodology of our proposed work. More precisely, after defining thoroughly the problem and work objectives, we introduce gradually the proposed method starting by the description of Hidden Markov Model (HMM) for EMG signal modelling. Then, the discrete wavelet transform and the proposed feature extraction technique are introduced. After that, we briefly describe the data base used in this study.

### A. Problem formulation of EMG signal segmentation

We consider $N$ measurements (*samples*) $\{y[n]\}_{n=1:N}$ of surface EMG signal. Given a chosen analysis window length, these observations are divided into overlapping frames. For each signal's frame, we are interested in determining whether it contains a signal $s[n]$ embedded in a random background noise $w[n]$ (EMG activity) or, on the contrary, it is just the confusing manifestation of the noise (no EMG activity). Hence, we have a binary decision problem expressed as:

$$\Gamma : \begin{cases} H_0 : no\ EMG\ activity \\ H_1 : EMG\ activity \end{cases} \quad (1)$$

This task is known as EMG signal segmentation. In this work, the HMM method has been adapted to resolve the muscle activity detection problem or equivalently the EMG signal segmentation.

In addition, the visual inspection for activation tags (bursts onset and offset) was done by taking into account biomechanical indicators of the gait cycle calculated from accelerometer measurement. We were thus able to recalibrate the EMG activity with respect to the gait cycle. This is why the burst of activity defined in our recordings corresponded to a functional activation for walking and not to a non-voluntary activation.

### B. Hidden Markov Model (HMM) for EMG signal modelling

Signal segmentation is the process of identifying the boundaries between segments of different classes. Signal segmentation system can be carried out using statistical modelling techniques such as Gaussian Mixture Models (GMM) and Hidden Markov models (HMM). Such system can be seen as pattern recognition system which requires a training phase for modelling the temporal pattern classes and recognition phase for identifying the different classes of segments of input signal. Hence, training and testing databases of signals are required to design the segmentation system. Both phases require feature extraction step to convert each signal in sequence of features vectors obtained by dividing the signal into overlapping windows and computing from each window a set of features that constitutes the feature vector (see Fig.1). This sequence of vectors can be considered as input sequence of observations in modelling or classification steps.

In this work, we used 50% overlapping windows.
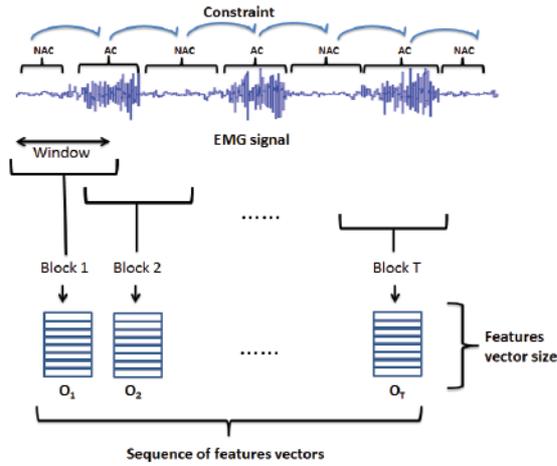
**Fig. 1**: Features extraction steps for HMM modelling. The blue arrows specify the authorized sequence of classes which defines constraints embedded in the language model for HMM design.

Particularly, in speech recognition task, the signal segmentation process has the aim to identifying the boundaries of segments of phonemes, word or sentences using respectively acoustic model, lexical model and language model. HMM models based system can be implemented using the HTK tools library (Hidden Markov Model Toolkit) which was designed firstly for speech recognition system including acoustic, lexical and language modelling [26]. This toolkit permits to model each phoneme by an HMM model of $N_{states}$ states each one associated to GMM model of $N_{GMM}$ Gaussian components for modelling the observation probability density function. The parameters of HMM models are estimated using the embedded Baum-Walsh algorithm which is implemented in HTK software using the HEREST command. The recognition phase can be carried out with the Viterbi algorithm using the HVITE command with respect to a language model and dictionary [26]. Nevertheless, the HMM models based system using HTK tools library has been used for others applications such as electrical appliance identification [27], speech recognition [28] and emotion recognition [29].

In this work, we use this toolkit to segment the EMG signal in regions (patterns) of activity (AC) class and regions of No-activity (NAC) class. The EMG signal is considered as a sequence of successive regions of AC and NAC classes, each one being represented by an observations sequence (sequence of features vectors) and modelled by HMM (see Fig.1)). The parameters of each model of AC class or NAC class are estimated from the training phase. By considering the same approach as for speech signal segmentation, the EMG signal segmentation uses the Viterbi algorithm to recognize the classes sequence using the trained HMM with respect to a language model that represents the constraint about the authorized sequence of classes.

*1— HMM models:* The HMM is a statistical model used to represent the evolution of temporal dynamics in time series or temporal observations sequence. The modelled phenomenon by HMM is a random and unobservable process that generates random observations. Thus, an HMM model results from the association of two stochastic processes: an unobservable process $Q(t)$ (states process) and an observable process $O(t)$. In general,

the $Q(t)$ is a Markov chain which is supposed to be at each instant $t$ in a state $q_t = i (1 \leq i \leq K)$ which emits an observation $o_t$ with probability distribution $b_i(o) = P(o\backslash q = i)$ [30, 31].

Let each region be represented by a sequence of feature vectors or observations sequence $O(t)$, defined as:

$$O = o_1, o_2, ...., o_T \qquad (2)$$

Where $o_t$ is the features vector observed at time t and let the states sequence $Q$ given by:

$$Q = q_1, q_2, ...., q_T \qquad (3)$$

HMM model can be considered as a probabilistic machine of finite states described by a set of nodes (or states) related together by arcs of transitions (see Fig.2). Each state $i (1 \leq i \leq K)$ at time instant t is associated with a probability distribution $b_i(o) = P(o\backslash q_t = i)$ of observation $o$ and with each transition from state $i$ to state $j$ is associated a transition probability $a_{ij} = P(q_t = j\backslash q_{t-1} = i)$ [30, 31]. If the observations are defined in a continuous space, then the probability density function of observation emission in state $j$ can be represented by a Gaussian mixture model(GMM) of M components, given by [26]:

$$b_j(o_t) = \sum_{i=1}^{M} \frac{c_i}{\sqrt{(2\pi)^d |\sum_i|}} \exp(-\frac{1}{2}(o_t - \mu_i)^*.\sum_i^{-1}.(o_t - \mu_i)) \qquad (4)$$

Where: $\mu_i$ and $\sum_i$ represent the mean and the covariance matrix of the $i^{th}$ Gaussian component, $c_i$ is the weight of the $i^{th}$ Gaussian satisfying the condition: $\sum_{i=1}^{M} c_i = 1$
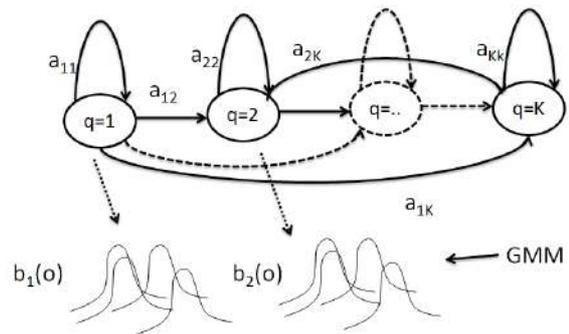


**Fig. 2**: Example of left- Right HMM Model [31].

The basic parameters of an HMM model are given as follows:

- The states transition matrix $A = \{a_{ij}\}$.

- The matrix of observation probability distribution $B = \{b_i(o)\}$.

- The initial state probability $\pi = \{\pi_i\backslash = P(q_1 = i)\}$

HMM can be written as $\lambda = (A, B, \pi)$

Given several HMM models $\lambda$ and given an input observations sequence $O$, the training phase consist to estimate the appropriate parameters of the HMM model that maximize the likelihood $P(O\backslash\lambda)$. Practically, the re-estimation of the HMM model parameters can be carried out using the Baum-Welch algorithm. The recognition phase has the objective to find the best HMM model that maximizes the probability $P(\lambda\backslash O)$. This phase can be carried out using the Viterbi algorithm.

*2— HMM models based segmentation system :* Fig.3 illustrates the diagram of our segmentation system. In the training phase, each region class is modelled by left-right HMM model of $N_{states}$, each one being represented by GMM model of $N_{GMM}$ Gaussian with diagonal covariance matrices. The parameters of HMM and GMM models are estimated using the HEREST command of the HTK tool, applied on the sequence of feature vectors extracted from the EMG signals of the training database. Furthermore, this estimation requires the reference text transcription that contents the class's sequence of each signal. The sequences of feature vectors are extracted using wavelet analysis applied on each EMG signal. This section will be described in the following section.

In the segmentation phase, the HVITE command of HTK tool uses the trained HMM models and constraints model (language model) for transcribing each input features vectors sequence in sequence of classes (AC and NAC) and detecting the boundaries of their segments [26, 32]. The constraints problem consists to accept only the classes's sequence in which each AC label (class) is followed by NAC label is followed by NAC label (see Fig.1).
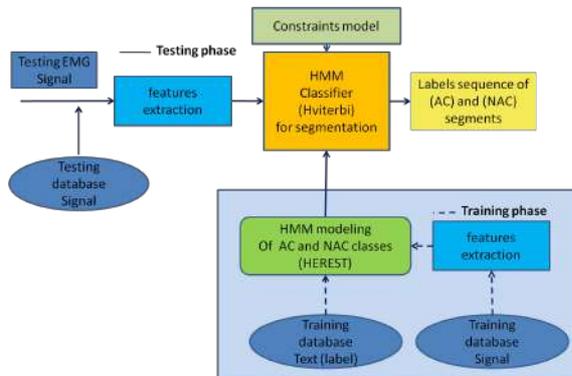


**Fig. 3**: HMM-models-based EMG signal segmentation.

Finally, the performance evaluation can be done using the HRE-SULTS command of HTK tool which compares each testing transcription of an EMG signal with its corresponding reference transcription [26]. The result of the segments identification is evaluated using the accuracy $Acc$ defined in subsection A..

### C. Discrete wavelet transform based feature extraction

Wavelet transform has largely been used as signal analysis technique in pattern recognition systems in order to extract a set of features from each analysis window (frame) of the signal of interest. The success of this technique essentially relies on the flexibility provided by the mother wavelet which choice depends on the properties of the data as well as on the targeted application.

For example, the orthogonality property between members of a wavelet family is desirable for compression purpose whereas the compactness property is desirable for the analysis and detection of impulsive data. Moreover, wavelet transform is particularly suitable for the analysis of data stemming from the real world because of its scale property that is compliant with many physical behaviors: long time events should be analyzed with slow long waveforms whereas short time or impulsive events should be analyzed with rapid short waveforms. The tool is thus suitable for correctly extracting the impulsive information of motor unit action potentials (MUAP) generated by EMG activation. Impulsive information can therefore be precisely localized and the information content precisely analyzed. To that aim, the authors of [10] proposed to use the first-order Hermite–Rodriguez function. The continuous wavelet transform has been used in [33] for the identification of the single action potentials in the time scale domain. The authors argue that the filtering effect of the volume conductor occurring between source signal taking its deep origin at the fiber level and MUAP detected signal at the skin surface produces different MUAP's that can be considered as dilated and attenuated versions of one single shape. This observation motivates the use of wavelet provided a mother wavelet with similar shape exists or can be created.

In the discrete form of the wavelet transform (DWT), decomposition coefficients are computed at different scales up to the maximum dyadic decomposition depth $p$ (to be *a priori* chosen or maximally fixed as $log_2(N)$ where $N$ is the number of data samples in the analysis window). At each scale or decomposition level $j$, a set $d_j[n]$ (indexed by $n$) $N_j = \frac{N}{2^j}$ of wavelet coefficients is computed (the name $d$ is for detail coefficients). The last decomposition level $p$ produces the set $a_p[n]$ (the name $a$ is for approximation coefficients). Therefore, a discrete wavelet energy (DWE) can be computed at each scale as:

$$
\begin{cases}
E(d_j) = \displaystyle\sum_{n=0}^{N_j-1} |d_j[n]|^2 & \text{for } j = 1, ..., p \\
E(a_p) = \displaystyle\sum_{n=0}^{N_p-1} |a_p[n]|^2
\end{cases}
\tag{5}
$$

Finally, the set of features extracted from each analysis window of a signal is composed of wavelet coefficient energies $E(d_1), E(d_2), ..., E(d_p), E(a_p)$, which makes an energy analysis on consecutive spectral bands of the data (from the highest to the lowest when $j$ goes from 1 to $p$). In order to extract other possible descriptors, the authors in [27] have used the DWE normalized on total energy of window analysis, the logarithm of wavelet energy (LWE) and the Wavelet Cepstral Coefficient (WCC) computed from the discrete cosine transform (DCT) of LWE (see Fig.4). In this work, we investigated the impact of these descriptors on the performance results of EMG signal segmentation. Hence different experiences have been carried out to search for the optimal configuration.

### D. Data base description

The HMM based method is applied to real sEMG signals. It has been recorded on different muscles which characterize the gait. These signals have several EMG activity bursts and are recorded in ecological conditions within the French National
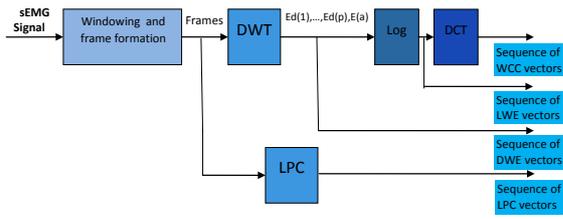
**Fig. 4**: Process for the extraction of LPC / DWE / LWE / WCC features [27].

Project ECOTECH [18]. The group was constituted of nine healthy subjects and eight Parkinsonian subjects recorded during the ECOTECH project. This study was approved by the local ethics committee and subjects provided written consent prior to commencement.

We have tested the proposed method for surface EMG signals recorded during the ECOTECH experiment. Patients were prepared for electrodes placement by shaving the skin and cleaning it with alcohol wipes. Differential EMG sensors were placed on the muscle belly parallel to the main direction of muscle fibres in accordance with study on the innervation zone [34]. Data were collected using an on board system of 16 wearable sensors (bandwidth 20-450 Hz, 16 bits per sample, 1926 Hz sampling rate, Delsys Trigno, Natick, USA) with SNR = 3.31 dB. Each sensor recorded the activity of a single muscle. In this work, we selected the data collected from the right *soleus* muscle ($RSol$) which is predominantly involved in gait and shows high energy. For this muscle, all EMG bursts were previously manually segmented by an expert. The data base description is reported in tables I and II.

The group of eight Parkinson's patients had a median age of 59.5 years [54-65.7, first and third quartiles] and was composed of two women and six men with a moderate bilateral deficit and a median UPDRS score of 7.5 [3.75-13.25], score obtained under anti-parkinsonian medication (ON state). The control group of nine healthy people had a median age of 38 years [38-49] and consisted of one woman and eight men. No motor disability or dysfunction were a prerequisite to be part of this latter. The electromyographic activities were collected during continuous walking sequences for which the conditions were the same between groups: free (natural gait), rectilinear and without obstacles or obstructions. More specifically for the patients, the gait tests were carried out in ON state.

### III. Results and discussion

The proposed HMM based method was applied to segment the real sEMG signals automatically. The following section describes the performance tools used for an objective comparison of the segmentation results. Then, an experimental setup is carried out in order to get the optimal configuration of the segmentation system. Statistical segmentation results are also provided and the obtained results are compared with those of the state-of-the-art methods reported in [15].

This expertise was conducted within the ECOTECH project [18] where the EMG signal segmentation has been achieved by biomedical researchers using visual inspection.

### A. Performance evaluation tools

In order to evaluate the performance of the segmentation task, we have considered two main criteria, namely the classification accuracy ($Acc$) and the error rate ($Re$). The $Acc$ is defined in Eq.(6) and is used to evaluate the number of correctly detected EMG burst activities using the HRESULTS command of HTK tool [26].

$$Acc = \frac{N - D - S - I}{N} \qquad (6)$$

$N$ represents the total number of segments labels in the reference transcriptions of EMG signals, $D$ is the number of deleted labels, $S$ is the number of substituted labels and $I$ is the number of inserted labels.

The $Re$ performance criterion is used for the fine detection of the beginning (onset) and the end (offset) of each EMG burst activity. In order to evaluate $Re$ (%), we have used a binary representation of our sEMG signal according to:

$$B[n] = \begin{cases} 0 & \text{if the decision made for the } n\text{-th sample is 'no activity'} \\ 1 & \text{if the decision made for the } n\text{-th sample is 'activity'} \end{cases}$$

$$Re = Prob(B[n] \neq B_{ex}[n]) \qquad (7)$$

where $B_{ex}[n]$ refers to the exact value of our binary signal representation given by the expert and Prob refers to the probability of a given event.

In this work, we have also achieved the performance's analysis by computing the mean and standard deviation (STD) values of the burst duration values as well as the corresponding absolute latency error (ALE) between the mean of the experimental duration and the mean of the expert duration as given in Eq.(8).

$$ALE = |\text{Mean of real bursts durations } - \text{ Mean of estimated bursts durations}|$$
$$(8)$$

### B. Optimal configuration of the segmentation system

The purpose of this section is to describe the experimental setup that will give the best configuration of our system by evaluating the proposed HMM method's performance for two types of signals, namely real EMG signals of healthy subjects and real EMG signals of Parkinsonian subjects.
We present in this section a number of experiments we carried out to find the optimal configuration of our framework that gives the best performance To find the optimal parameters of the HMM (state number and gaussian components number).

Three experiments are processed in order:

(1) to compare the performance of the DWE, LWE and WCC descriptors to other features commonly used in the literature such as prediction analysis (LPC: Linear Prediction Coefficients);

| Subjects | | Number of EMG bursts | Mean of real EMG burst duration (ms) | STD of real EMG burst duration (ms) | Signal length (samples) | Signal duration (s) |
|---|---|---|---|---|---|---|
| Training phase | $Control_1$ | 22 | 1037.7000 | 98.5539 | 50206 | 26.0685 |
| | $Control_2$ | 10 | 1059.7000 | 288.3023 | 21595 | 11.2128 |
| | $Control_3$ | 11 | 1068.1000 | 85.4207 | 27733 | 14.3998 |
| | $Control_4$ | 11 | 1069.3000 | 103.7267 | 28396 | 14.7441 |
| | $Control_5$ | 11 | 809.9091 | 33.3090 | 21500 | 11.1635 |
| Testing phase | $Control_6$ | 6 | 1031.7000 | 73.5193 | 14853 | 7.7121 |
| | $Control_7$ | 6 | 1013.3000 | 321.9687 | 12576 | 6.5298 |
| | $Control_8$ | 12 | 898.0833 | 48.5264 | 27629 | 14.3458 |
| | $Control_9$ | 26 | 969.2308 | 104.8490 | 55024 | 28.5702 |

**Table. I**
DESCRIPTION OF sEMG SIGNALS FOR HEALTHY SUBJECTS.

| Subjects | | Number of EMG bursts | Mean of real EMG burst duration (ms) | STD of real EMG burst duration (ms) | Signal length (samples) | Signal duration (s) |
|---|---|---|---|---|---|---|
| Training phase | $Park_1$ | 10 | 1225.3000 | 65.0898 | 55024 | 28.5702 |
| | $Park_2$ | 10 | 959.2000 | 121.2910 | 17587 | 9.1317 |
| | $Park_3$ | 5 | 1362 | 233.4309 | 9371 | 4.8657 |
| | $Park_4$ | 37 | 933.8108 | 128.3433 | 76296 | 39.6152 |
| Testing phase | $Park_5$ | 10 | 766.2000 | 53.1012 | 21739 | 11.2876 |
| | $Park_6$ | 9 | 990.5656 | 55.1750 | 17170 | 8.9152 |
| | $Park_7$ | 5 | 773.8000 | 157.9421 | 8617 | 4.4742 |
| | $Park_8$ | 5 | 652.2000 | 80.6021 | 8953 | 4.6487 |

**Table. II**
DESCRIPTION OF sEMG SIGNALS FOR PARKINSONIAN SUBJECTS.

(2) to find the optimal analyzing frame duration;

(3) to search for the optimal combination of decomposition level and mother wavelet;

(4) to analyze the performance results of our framework and to compare them with the results obtained using other state-of-the-art existing methods.

In order to deeply evaluate the performance of our method, we carried out the first experiment for different numbers of HMM states $N_{states}$, different numbers of Gaussians for GMM modelling $N_{GMM}$ and different levels of wavelet decomposition $L_{decomp}$ or $P$ order for LPC descriptor. The mother wavelet 'Sym4' was chosen with $66.45ms$ window durations as recommended in [35] for EMG pattern recognition. The best descriptors with the optimal parameters were chosen for the second experiment of window duration study then for the optimal wavelet configuration. Finally a comparison with state of the art methods is carried out.

*1— Performance comparison of the descriptors:* This section shows evaluation results of the DWE descriptor for the task of surface EMG signal segmentation. The performance analysis is compared to LPC descriptor, LWE (Log Wavelet decomposition based Energy) descriptor and WCC descriptor. In this experiment, we search for the optimal parameters configuration that gives the best $Acc$ and $Re$. For each descriptor, we have varied the state number for each experiment $N_{states} = (2,3,4,5,6,7,8,9,10)$, the number of components in GMM modelling $N_{GMM} = (1,2,3,6,12,24,48)$ and the order $P = (2,3,4,5,6,7,8,9,10)$ of LPC descriptor. The analysis frame duration is fixed to $66.45\ ms$ (which represents 128 samples) and the mother wavelet is chosen to 'Sym4' and the wavelet decomposition level $L_{decomp} = 4$. The obtained results are given in Table III. The latter shows the optimal configurations in terms of number of Gaussians $N_{GMM}$ and number of states $N_{states}$ for each wavelet descriptor DWE, LWE and WCC as well as the optimal order $P$ added for the LPC descriptor. The

$Acc$ and $Re$ values are given for these optimal configurations. In order to evaluate the robustness of each descriptor regarding the variations of $N_{GMM}$ and $N_{states}$, the mean and STD values of $Re$ are reported by computing the $Re$ values when considering all the values of the tested parameter sets. From all these results, we adopted the following rules in order to identify the best configuration of the system. First, we examine the best accuracy results. Second, when multiple best configurations exist, the best Re values are considered also taking into account the robustness results. Results shown in Table III highlight the performance gain of the wavelet descriptors that all show 100% of $Acc$ compared to the LPC descriptor showing the bad score of 87.34%. Now, the best descriptor is WCC with 4.88% for the $Re$ value. However, the WCC descriptor requires more calculus and is more influenced by the variations of the state number and the gaussian number (mean of 7.41 and standard deviation of 0.51). The next candidate with a weak $Re$ value is LWE also showing better $Re$ statistics than WCC (in particular with an STD value which is about 4 times lower). Hence, in the following sections, we choose the LWE descriptor by taking a 2-states number and a 3-gaussians number.

*2— Influence of the window duration:* After choosing the feature's descriptor, we investigate the appropriate analysis frame duration. This section aims at investigating performance improvements by taking into account the advantages of wavelet analysis appropriate for the non stationary case of EMG signals. We have varied the analysis frame duration with 'Sym4' wavelet, number of GMM $N_{GMM} = 3$, number of states $N_{states} = 2$ and decomposition level equal to $L_{decomp} = 4$ for LWE (Log Wavelet decomposition based Energy) descriptor. Table IV shows the $Acc$ and $Re$ values for different values of analysis frame duration. The best performance is obtained for analysis frame duration equal to $66.45\ ms$ which gives $Acc$ and $Re$ values equal to 100 % and the lowest $Re$ equal to 5.37 %.

The following subsection focuses on the performance analysis for the optimal choice of mother wavelet.

| Descriptors | $LPC^*$ | $DWE$ | $LWE$ | $WCC$ |
|---|---|---|---|---|
| Optimal parameters | $N_{GMM} = 24$ $N_{states} = 2$ and $P = 2$ | $N_{GMM} = 3$ $N_{states} = 3$ | $N_{GMM} = 3$ $N_{states} = 2$ | $N_{GMM} = 48$ $N_{states} = 2$ |
| $Acc$ % | 87.34 | 100 | 100 | 100 |
| $Re$ % | 43.48 | 6.16 | 5.37 | 4.88 |
| Mean of $Re$ % | 49.14 | 6.24 | 6.23 | 7.41 |
| STD of $Re$ % | 0.731 | 0.083 | 0.138 | 0.510 |

**Table. III**

PERFORMANCE COMPARISON OF THE $Acc$ % FOR DWE, LPC, LWE AND WCC FEATURES USING 'Sym4' AND $L_{decomp} = 4$ WITH ANALYSIS FRAME DURATION EQUAL TO 66.45 $ms$ FOR THE HMM OPTIMAL PARAMETERS.THE MEAN AND STD VALUES ARE STATISTICAL RESULTS EVALUTED BY CONSIDERING $N_{GMM}$ AND $N_{states}$ RANGES PLUS $P$ RANGE FOR LPC DESCRIPTOR. THE STAR * INDICATES THAT $Re$ WAS CALCULATED ONLY TAKING INTO ACCOUNT CORRECTLY CLASSIFIED SEQUENCES

.

| Analysis frame duration ($ms$) | 16.61 | 33.22 | 49.84 | 66.45 | 83.07 | 99.68 | 116.30 | 132.91 |
|---|---|---|---|---|---|---|---|---|
| $Acc$ % | 80.38 | 91.14 | 97.47 | 100 | 100 | 99.37 | 100 | 100 |
| $Re$ % | 6.41 | 5.39 | 5.33 | 5.37 | 5.505 | 5.61 | 5.81 | 8.46 |

**Table. IV**

PERFORMANCE COMPARISON OF THE $Acc$ (%) AND $Re$ (%) FOR DWE FEATURES USING 'Sym4' FOR DIFFERENT ANALYSIS FRAME DURATIONS FOR THE HMM OPTIMAL PARAMETERS.

*3— Choice of the mother wavelet:* Many studies of surface EMG analysis have concluded that the Daubechies (Db) wavelet family is the most suitable wavelet for sEMG signal analysis [36–38]. In [35], the authors have concluded that the 'Sym4' is the most appropriate one for EMG pattern recognition. This experiment aims at selecting the optimal mother wavelet order within its family for an analysis window durations equal to 66.45 $ms$, a number of GMM $N_{GMM} = 3$, a number of states $N_{states} = 2$ with a decomposition level varying between 1 to log2 of samples number of analysis window (max level=7). In this work, we consider the following wavelet families:

- the Daubechies family with orders 1 to 8: Db1, Db2, ... , Db8;

- the Symlets family with orders 1 to 8: Sym1, Sym2, ..., Sym8;

- the Coiflets family with orders 1 to 5: Coif1, Coif2,..., Coif5.

The $Acc$, $Re$ and $L_{decomp}$ values are reported in Tables V, VI and VII for each of the three wavelet families, respectively. For all the families, the optimal decomposition level decreases with the augmentation of the mother wavelet order. The results highlight the performance robustness of the proposed approach in terms of $Acc$ value where we notice that the mean values of $Acc$ are greater than 97.47 % and the mean of error rate $Re$ are lower than 5.70 %. In particular, we also notice that the 'Coif5' wavelet for decomposition level equal to $L_{decomp} = 4$ leads to the minimum of $Re = 4.68$ % and the maximum of $Acc = 100$ %. This latter result shows improved performance in terms of error rate compared to WCC performance as reported in Table III. The improvement of the results is consistent with the step by step experimental strategy we follow, which purpose

is to propose a tuned set for the configuration of the system, without requiring too much exhaustive empirical calculus. We thus retain this last result as a new starting point for further comparison studies that will be presented in section C.

### C. Statistics of the bursts duration estimation

In this part of our work, we conduct a performance analysis of the segmentation task using statistics on the segmented bursts duration values (mean, STD and ALE). The results are obtained with an analysis frame duration equal to 66.45 $ms$, a number of GMM $N_{GMM} = 3$, a number of states $N_{states} = 2$ and a decomposition level equal to $L_{decomp} = 4$ with the mother wavelet 'Coif5'.

Table VIII presents the mean and the STD of the estimated bursts duration values, the estimated number of EMG activity bursts and ALE for the healthy's and Parkinson's subjects from the sEMG records of right *soleus* muscle. The table shows that the estimated numbers of bursts are compliant with the expert numbers. The results also show low ALE values (less than 46 ms for all subjects) except for one Parkinson's subject (showing 80 ms).

Fig.5 and Fig.6 show illustrative examples of a real sEMG signal for right *Soleus* muscle of control and Parkinson's subjects respectively. The signal is segmented in AC and NAC parts using HMM method for $N_{GMM} = 3$, $N_{states} = 2$, $L_{decomp} = 4$, wavelet = $Coif5$ and frame length = $66.45ms$ for an error rate equal to $Re = 2.71\%$ for control subject and for Parkinson's subject $Re = 6.06\%$.

### D. Performance comparison with some state-of-the-art segmentation methods

In order to compare our method with state-of-the-art segmentation methods, we have considered the five methods presented

| Daubechies | | $Db1$ | $Db2$ | $Db3$ | $Db4$ | $Db5$ | $Db6$ | $Db7$ | $Db8$ | Mean | STD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L_{decomp}$ | 7 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | // | // |
| 66.45 $ms$ | $Acc\%$ | 97.47 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | // | // |
| | $Re\%$ | 5.20 | 5.08 | 5.28 | 5.47 | 5.53 | 4.98 | 5.23 | 5.02 | 5.22 | 0.199 |

**Table. V**

PERFORMANCE RESULTS IN TERMS OF $Acc$ (%), $Re$ (%) AND OPTIMAL $L_{decomp}$ FOR LWE FEATURES USING **DAUBECHIES** WAVELET FAMILY.

| Symlets | | $Sym1$ | $Sym2$ | $Sym3$ | $Sym4$ | $Sym5$ | $Sym6$ | $Sym7$ | $Sym8$ | Mean | STD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L_{decomp}$ | 7 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | // | // |
| 66.45 $ms$ | $Acc\%$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | // | // |
| | $Re\%$ | 5.20 | 5.08 | 5.28 | 5.21 | 5.18 | 5.70 | 5.20 | 5.30 | 5.268 | 0.186 |

**Table. VI**

PERFORMANCE RESULTS IN TERMS OF $Acc$ (%), $Re$ (%) AND OPTIMAL $L_{decomp}$ FOR LWE FEATURES USING **SYMLETS** WAVELET FAMILY.

| Coiflets | | $Coif1$ | $Coif2$ | $Coif3$ | $Coif4$ | $Coif5$ | Mean | STD |
|---|---|---|---|---|---|---|---|---|
| | $L_{decomp}$ | 6 | 4 | 4 | 4 | 4 | // | // |
| 66.45 $ms$ | $Acc\%$ | 100 | 100 | 100 | 100 | 100 | // | // |
| | $Re\%$ | 4.91 | 4.99 | 5.26 | 5.23 | 4.68 | 5.014 | 0.2399 |

**Table. VII**

PERFORMANCE RESULTS IN TERMS OF $Acc$ (%), $Re$ (%) AND OPTIMAL $L_{decomp}$ FOR LWE FEATURES USING **COIFLETS** WAVELET FAMILY.

| Subject | | $Control_6$ | $Control_7$ | $Control_8$ | $Control_9$ | $Park_5$ | $Park_6$ | $Park_7$ | $Park_8$ |
|---|---|---|---|---|---|---|---|---|---|
| | Mean(ms) | 1031.7000 | 1013.3000 | 898.0833 | 969.2308 | 766.2000 | 990.5556 | 773.8000 | 652.2000 |
| Real burst duration | STD (ms) | 73.5192 | 321.9687 | 48.5264 | 104.8490 | 53.1012 | 55.1750 | 157.9421 | 80.6021 |
| | Number | 6 | 6 | 12 | 26 | 10 | 9 | 5 | 5 |
| | Mean (ms) | 1003.7020 | 1044.0991 | 925.2054 | 1003.3472 | 719.77 | 959.2504 | 693.7117 | 616.0029 |
| Estimated burst duration | STD (ms) | 57.2459 | 350.6888 | 48.3647 | 121.5427 | 125.0333 | 49.9449 | 124.7589 | 72.9712 |
| | Number | 6 | 6 | 12 | 26 | 10 | 9 | 5 | 5 |
| Performance | ALE (ms) | 27.95 | 31.09 | 27.12 | 34.11 | 46.43 | 31.30 | 80.0883 | 36.1971 |

**Table. VIII**

MEAN, STANDARD DEVIATION (STD) AND ABSOLUTE LATENCY ERROR (ALE) OF EMG BURST ACTIVITY DURATION FOR **RSOL** RIGHT *soleus* MUSCLE FOR HEALTHY AND PARKINSON'S SUBJECTS.

in [15]. Table IX presents $Re$ values for the healthy and Parkinson's subjects of sEMG records from the right *Soleus* muscle. The results highlight the performance gain of the proposed HMM approach compared to the other existing methods. In order to facilitate the reading of the results, the Mean and STD values are graphically displayed in Fig.7. In order to highlight the differences between the two classes, the graphical statistical results are also presented in Fig.8 for both categories. From these graphs, we notice that, the HMM method seems to be the most appropriate for an automatic EMG activity monitoring for all the subjects (Fig.7). More specifically, by comparison with the other state-of-the-art methods, the segmentation is far better for the control subjects and also shows a slight improvement for the Parkinsonian's subjects (Fig.8).

Table X presents the comparative ALE values and estimated numbers of EMG activity bursts for the $Control_6$ and $Park_5$ subjects of sEMG records from the right *Soleus* muscle. The results highlight the performance gain of the proposed HMM approach compared to the other existing methods, in particular w.r.t. the number of EMG activity bursts. For the Parkinson case, all methods except our failed in estimating the right number of bursts. The ALE values for our method are also always the lowest among the concurrent methods.

### E. Recommendation and discussion

Several techniques already exist for EMG signal activity segmentation, but the detection performance is far from being satisfactory in difficult cases like in neurodegenerative EMG activity. This work represents a further step towards developing efficient

| Subject | DoubleTh | Wavelet | TKEO | RMS | SampEn | FM-ALED | HMM |
|---|---|---|---|---|---|---|---|
| $Park_5$ | 20.11 | 22.04 | 18.96 | 18.09 | 17.65 | 9.78 | 6.06 |
| $Park_6$ | 22.00 | 20.63 | 22.54 | 12.90 | 14.54 | 5.83 | 3.16 |
| $Park_7$ | 19.82 | 18.65 | 15.27 | 9.40 | 14.50 | 7.92 | 10.35 |
| $Park_8$ | 17.08 | 21.94 | 17.59 | 16.06 | 13.72 | 10.77 | 5.55 |
| $Control_6$ | 22.96 | 14.42 | 10.81 | 8.03 | 7.53 | 4.38 | 2.71 |
| $Control_7$ | 18.04 | 17.77 | 12.27 | 11.14 | 8.83 | 5.07 | 3.07 |
| $Control_8$ | 19.74 | 20.43 | 15.76 | 15.27 | 23.46 | 7.62 | 3.02 |
| $Control_9$ | 17.08 | 22.75 | 15.47 | 18.92 | 9.55 | 7.60 | 3.52 |

**Table. IX**

COMPARISON OF ERROR RATE $Re$ WITH THEIR MEAN AND STANDARD DEVIATION VALUES OF EMG ACTIVITY BURST DETECTION FOR **RSOL** RIGHT *soleus* FOR HEALTHY AND PARKINSON'S SUBJECTS.

| Subject | | Real burst duration | DoubleTh | Wavelet | TKEO | RMS | SampEn | FM-ALED | HMM |
|---|---|---|---|---|---|---|---|---|---|
| Estimated burst duration for $Control_6$ | Mean (ms) | 1031.70 | 209.47 | 604.67 | 549.00 | 904.29 | 891.00 | 986.47 | 1003.70 |
| | STD (ms) | 73.52 | 157.85 | 434.41 | 604.15 | 394.28 | 520.51 | 128.60 | 57.24 |
| | Number | 6 | 19 | 9 | 14 | 7 | 8 | 6 | 6 |
| Performance for $Control_6$ | ALE (ms) | // | 822.22 | 427.03 | 482.70 | 127.41 | 140.07 | 45.18 | 27.95 |
| Estimated burst duration for $Park_5$ | Mean (ms) | 766.20 | 281.18 | 327.46 | 284.20 | 300.88 | 327.55 | 667.56 | 719.77 |
| | STD (ms) | 53.10 | 201.96 | 180.38 | 177.59 | 281.33 | 105.15 | 16.33 | 125.03 |
| | Number | 10 | 17 | 13 | 15 | 34 | 11 | 9 | 10 |
| Performance for $Park_5$ | ALE (ms) | // | 485.02 | 438.73 | 482.00 | 465.31 | 438.65 | 98.64 | 46.43 |

**Table. X**

MEAN, STANDARD DEVIATION (STD) AND ABSOLUTE LATENCY ERROR (ALE) OF EMG BURST ACTIVITY DURATION FOR **RSOL** RIGHT *soleus* MUSCLE OF $Control_6$ AND $Park_5$ SUBJECTS USING DIFFERENT SEGMENTATION METHODS.
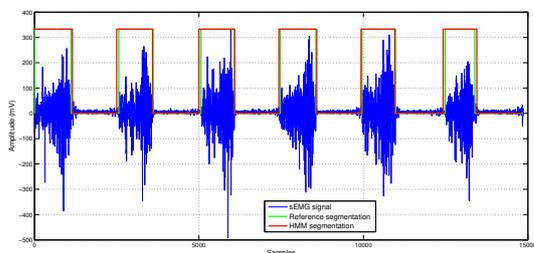


**Fig. 5**: Automatic segmentation of real sEMG signal of the control subject ($Control_6$) with HMM method for right *Soleus* muscle with $Re$ = 2.71 % ($N_{GMM}$ = 3, $N_{states}$ = 2, $L_{decomp}$= 4, wavelet = $Coif5$ and frame length = $66.45ms$.)
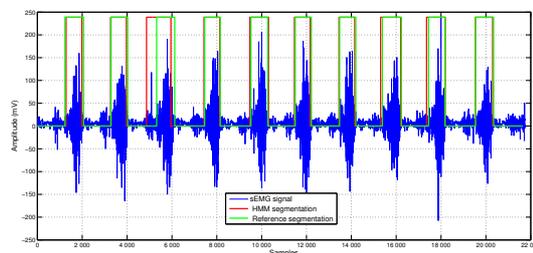


**Fig. 6**: Automatic segmentation of real sEMG signal of the Parkinson's subject ($Park_5$) with HMM method for right *Soleus* muscle with $Re$ = 6.06 % ($N_{GMM}$ = 3, $N_{states}$ = 2, $L_{decomp}$= 4, wavelet = $Coif5$ and frame length = $66.45ms$.)

automatic segmentation solution which is an important task in raw data processing in order to be able to exploit massive EMG datasets necessary for the development of systems that can be used as genuine medical diagnosis supports in future, especially in Parkinson's disease diagnostic. The proposed approach can be seen as an improvement of the EMG activity detection problem in particular because the study has been carried out on the ECOTECH dataset gathering ecological difficult situations close to conditions that can be encountered in the targeted applications. In this context, our proposed HMM method has been applied to real EMG signals recorded for healthy and Parkinson's subjects. Different experiments have been carried out to find the optimal configuration of the segmentation system that gives the best performance in terms of accuracy and error rate. The results have
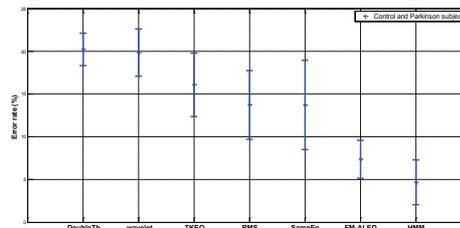


**Fig. 7**: Comparison of error rate $Re$ showing mean and standard deviation of burst EMG activity detection for **RSol** right *soleus* for healthy and Parkinson's subjects.
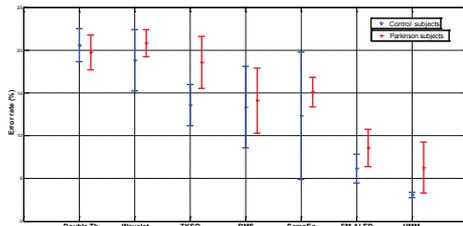
**Fig. 8**: Comparison of error rate $Re$ showing mean and standard deviation of burst EMG activity detection for **RSol** right *soleus* for healthy and Parkinson's subjects.

shown clear improvements of the system performance gained step by step from consecutive experiments. We summarize and discuss here the main improvements made out of this study and highlight some perspectives for this research work.

- The DWT analysis presents the first improvements compared to the LPC analysis whatever the wavelet descriptor type (DWE, LWE and WCC). This analysis gives the accuracy of 100%. Particularly, the LWE descriptor is shown to be more appropriate for the segmentation task. These results have been obtained by taking the best configuration of HMM models of 2 states associated each one with a GMM model of 3 gaussian components, analysis window of $66.45ms$ duration and giving error rates around $5\%$ whatever the family of mother functions.

  It can be noticed that the optimal decomposition level decreases by increasing the order of the Daubechies wavelet family (this is also true for the other families). This can be explained by the fact that the length of the associated filter used for wavelet coefficients computation increases with the Daubechies order. At the first Daubechies orders, filter lengths are rather small which remains applicable for filtering purpose with a small number of data samples that are considered at the deepest decomposition levels. The order increasing, the filter length increasing thus requiring data with a higher number of samples for a correct filtering application. Therefore limiting the optimal decomposition level values to 4 seems to be a appropriate choice in this context. Experimental results show that the Coiflet wavelet function with level decomposition of 4 presents the best error rate of $4.68\%$. Also the Sym2 and Db6 give a good performance too.

- The obtained segmentation performance in terms of error rate, absolute latency error and the number of burst EMG activity out performs those of the existing methods from the literature used in our comparative study (see Tables IX and X). A key point stemming from these results is the ability of the proposed method to estimate the correct number of bursts whatever the subjects under study. Having access to the number of bursts with their durations is essential for the monotoring of pakinsonian subjects and particularly in the frame of the following of a rehabilitation protocol. To that aim, we also recommend the use of supervised techniques even if labelling data can be controversial and difficult to obtain from experts. These methods are far

less common than current state-of-the-art unsupervised methods.

- The proposed HMM technique is most appropriate for EMG signal segmentation of healthy subjects where we have observed a significant performance gain in terms of error rate (see Fig.8).

- In this study, we have used the discrete wavelet transform as feature descriptor because this decomposition is more adapted than the Fourier decomposition to extract the impulsive information of the action potentials of the motor units (provided that the mother wavelet is well selected).

- Our proposed HMM method is robust to the choice of the mother wavelet type since we have concluded that all the tested mother wavelets family gave an error rate close to 5 % and classification accuracy close to 98 %.

## IV. CONCLUSION

Muscle activity detection of human skeletal muscle has important clinical applications. In this paper, we have proposed a novel framework for muscle activity detection based on the EMG signal segmentation using Hidden Markov models combined with DWT analysis. Different descriptors extracted from this analysis have been used and the results have shown that the HMM models combined with LWE descriptor extracted using the Coiflet 5 with decomposition level 4 give the best performance in terms of accuracy and error rate. A comparative study with state-of-the-art methods confirms the effectiveness of our approach. The proposed approach offers a good detection performance suitable for clinical applications. It may be a useful tool for analysing the EMG signals recorded during the assessment of postural adjustment of patients suffering from Parkinson's disease. As perspectives, we believe the use of an enlarged EMG dataset including all the EMG channels of ECOTECH distribution would be necessary to reinforce these first results and move a step towards the real-life application of the method. Another important issue (not treated in this work) would be to analyze the method's performance for EMG signals relative to other types of neurodegenerative diseases, e.g. Huntington and ALS (Amyotrophic lateral sclerosis).

### REFERENCES

[1] R Merletti, D Farina, M Gazzoni, A Merlo, et al. Surface electromyography: A window on the muscle, a glimpse on the central nervous system. *European Journal of Physical and Rehabilitation Medicine*, 37(1):57, 2001.

[2] Seth L Pullman, Douglas S Goodin, Anthony I Marquinez, Samer Tabbal, and Michael Rubin. Clinical utility of surface EMG: report of the therapeutics and technology assessment subcommittee of the american academy of neurology. *Neurology*, 55(2):171–177, 2000. https://doi.org/10.1212/wnl.55.2.171

[3] Paolo Bonato, Tommaso D'Alessio, and Marco Knaflitz. A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. *IEEE*

*Transactions on Biomedical Engineering*, 45(3):287–299, 1998. https://doi.org/10.1109/10.661154

[4] Gerhard H Staude, Werner M Wolf, Ulrich Appel, and Reinhard Dengler. Methods for onset detection of voluntary motor responses in tremor patients. *IEEE Transactions on Biomedical Engineering*, 43(2):177–188, 1996. https://doi.org/10.1109/10.481987

[5] Silvestro Micera, Angelo M Sabatini, and Paolo Dario. An algorithm for detecting the onset of muscle contraction by EMG signal processing. *Medical Eng. & Physics*, 20(3):211–215, 1998. https://doi.org/10.1016/S1350-4533(98)00017-4

[6] Lev Vaisman, José Zariffa, and Milos R Popovic. Application of singular spectrum-based change-point analysis to EMG-onset detection. *Journal of Electromyography and Kinesiology*, 20(4):750–760, https://doi.org/2010. 10.1016/j.jelekin.2010.02.010

[7] Qi Xu, Yazhi Quan, Lei Yang, and Jiping He. An adaptive algorithm for the determination of the onset and offset of muscle contraction by EMG signal processing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(1):65–73, 2012. https://doi.org/10.1109/TNSRE.2012.2226916

[8] Usman Rashid, Imran Khan Niazi, Nada Signal, Dario Farina, and Denise Taylor. Optimal automatic detection of muscle activation intervals. *Journal of Electromyography and Kinesiology*, 48:103–111, 2019. https://doi.org/10.1016/j.jelekin.2019.06.010

[9] Jie Liu, Dongwen Ying, and William Zev Rymer. EMG burst presence probability: A joint time–frequency representation of muscle activity and its application to onset detection. *Journal of Biomechanics*, 48(6):1193–1197, 2015. https://doi.org/10.1016/j.jbiomech.2015.02.017

[10] Andrea Merlo, Dario Farina, and Roberto Merletti. A fast and reliable technique for muscle activity detection from surface EMG signals. *IEEE Transactions on Biomedical Engineering*, 50(3):316–323, 2003. https://doi.org/10.1109/TBME.2003.808829

[11] Hsiu-Jen Liu, Kuu-Young Young, et al. An adaptive upper-arm EMG-based robot control system. *International Journal of Fuzzy Systems*, 12(3):181–189, 2010. https://doi.org/10.30000/IJFS.201009.0001

[12] Xiaoyan Li, Ping Zhou, and Alexander S Aruin. Teager-Kaiserenergy operation of surface EMG improves muscle activity onset detection. *Annals of biomedical engineering*, 35(9):1532–1538, 2007. https://doi.org/10.1007/s10439-007-9320-z

[13] Richard T Lauer and Laura A Prosser. Use of the Teager-Kaiser energy operator for muscle activity detection in children. *Annals of biomedical engineering*, 37(8):1584–1593, 2009. https://doi.org/10.1007/s10439-009-9727-9

[14] Stanislaw Solnik, Patrick Rider, Ken Steinweg, Paul DeVita, and Tibor Hortobágyi. Teager–Kaiser energy operator signal conditioning improves EMG onset detection. *European Journal of Applied Physiology*, 110(3):489–498, 2010. https://doi.org/10.1007/s00421-010-1521-8

[15] Hichem Bengacemi, Karim Abed-Meraim, Olivier Buttelli, Abdelaziz Ouldali, and Ammar Mesloub. A new detection method for EMG activity monitoring. *Medical & Biological Engineering & Computing*, 58(2):319–334, 2020. https://doi.org/10.1007/s11517-019-02048-0

[16] Paolo Soda, Stefano Mazzoleni, Giuseppe Cavallo, Eugenio Guglielmelli, and Giulio Iannello. Human movement onset detection from isometric force and torque measurements: A supervised pattern recognition approach. *Artificial intelligence in medicine*, 50(1):55–61, 2010. https://doi.org/10.1016/j.artmed.2010.04.008

[17] Amal Naseem, Meryem Jabloun, Olivier Buttelli, and Phillippe Ravier. Detection of sEMG muscle activation intervals using gaussian mixture model and ant colony classifier. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 1713–1717. IEEE, 2016. https://doi.org/10.1109/EUSIPCO.2016.7760541

[18] Olivier Buttelli. Agence nationale de la recherche. http://www.agence-nationale-recherche.fr/Projet-ANR-12-TECS-0020, 2012.

[19] Dinesh D Patil, RP Singh, Vilas M Thakare, and Avinash K Gulve. Automatic classification of ECG arrhythmia using morphological parameters with HMM and SVM. *International Journal of Applied Engineering Research*, 12(20):10376–10384, 2017.

[20] Dihong Jiang, Ya-nan Lu, MA Yu, and WANG Yuanyuan. Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and hmm-based refinement. *Expert Systems with Applications*, 121:188–203, 2019. https://doi.org/10.1016/j.eswa.2018.12.023

[21] Mohamed Nait-Meziane, Abdenour Hacine-Gharbi, Philippe Ravier, Guy Lamarque, Jean-Charles Le Bunetel, and Yves Raingeaud. HMM-based transient and steady-state current signals modeling for electrical appliances identification. In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*, pages 670–677. SCITEPRESS-Science and Technology Publications, Lda, 2016.

[22] Abed Khorasani and Mohammad Reza Daliri. HMM for classification of parkinson's disease based on the raw gait data. *Journal of medical systems*, 38(12):147, 2014. https://doi.org/10.1007/s10916-014-0147-5

[23] Matteo Rossi, Simone Benatti, Elisabetta Farella, and Luca Benini. Hybrid EMG classifier based on HMM and SVM for hand gesture recognition in prosthetics. In *2015 IEEE International Conference on Industrial Technology (ICIT)*, pages 1700–1705. IEEE, 2015. https://doi.org/10.1109/ICIT.2015.7125342

[24] Jang-Woo Kwon, Jung-Ho Kim, and Heung-Ho Choi. Classification of the EMG signal using cascaded classifier. In *World Congress on Medical Physics and Biomedical Engineering 2006*, pages 1222–1225. Springer, 2007. https://doi.org/10.1007/978-3-540-36841-0_295

[25] Adrian DC Chan, Kevin Englehart, Bernard Hudgins, and Dennis Fenton Lovely. Hidden markov model classification of myoelectric signals in speech. *IEEE Engineering in Medicine and Biology Magazine*, 21(5):143–146, 2002. https://doi.org/10.1109/MEMB.2002.1044184

[26] Steve Young, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, et al. The HTK book. *Cambridge university engineering department*, 3:175, 2002.

[27] Abdenour Hacine-Gharbi and Philippe Ravier. Wavelet cepstral coefficients for electrical appliances identification using hidden markov models. In *ICPRAM*, pages 541–549, https://doi.org/2018. 10.5220/0006662305410549

[28] V Shah, R Anstotz, I Obeid, and J Picone. Adapting an automatic speech recognition system to event classification of electroencephalograms. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5. IEEE, 2018. https://doi.org/10.1109/SPMB.2018.8615625

[29] Abdenour Hacine-Gharbi and Philippe Ravier. On the optimal number estimation of selected features using joint histogram based mutual information for speech emotion recognition. *Journal of King Saud University-Computer and Information Sciences*, 2019. https://doi.org/10.1016/j.jksuci.2019.07.008

[30] Jean-Paul Haton, Christophe Cerisara, Dominique Fohr, Yves Laprie, and Kamel Smaïli. *Reconnaissance automatique de la parole: Du Signal à son Interprétation*. Dunod, 2006.

[31] Abdenour Hacine-Gharbi. *Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole*. PhD thesis, 2012.

[32] Claude C Chibelushi, Farzin Deravi, and John SD Mason. A review of speech-based bimodal recognition. *IEEE transactions on multimedia*, 4(1):23–37, 2002. https://doi.org/10.1109/6046.985551

[33] Dario Farina, Elena Fortunato, and Roberto Merletti. Noninvasive estimation of motor unit conduction velocity distribution using linear electrode arrays. *IEEE Transactions on Biomedical Engineering*, 47(3):380–388, 2000. https://doi.org/10.1109/10.827303

[34] Marco Barbero, Roberto Merletti, and Alberto Rainoldi. *Atlas of muscle innervation zones: understanding surface electromyography and its applications*. Springer Science & Business Media, 2012.

[35] Jingwei Too, AR Abdullah, Norhashimah Mohd Saad, N Mohd Ali, and H Musa. A detail study of wavelet families for EMG pattern recognition. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(6):4221–4229, 2018.

[36] MS Hussain, Mamun Bin Ibne Reaz, Faisal Mohd-Yasin, and Muhammad I Ibrahimy. Electromyography signal analysis using wavelet transform and higher order statistics to determine muscle contraction. *Expert Systems*, 26(1):35–48, 2009. https://doi.org/10.1111/j.1468-0394.2008.00483.x

[37] K Mahaphonchaikul, D Sueaseenak, C Pintavirooj, M Sangworasil, and S Tungjitkusolmun. EMG signal feature extraction based on wavelet transform. In *ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pages 327–331. IEEE, 2010.

[38] Angkoon Phinyomark, Chusak Limsakul, and Pornchai Phukpattaranont. An optimal wavelet function based on wavelet denoising for multifunction myoelectric control. In *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 2, pages 1098–1101. IEEE, 2009. https://doi.org/10.1109/ECTICON.2009.5137236

**Hichem Bengacemi** received the engineering degree in electrical engineering and Magister degree in advanced signal processing techniques from the Ecole Militaire Polytechnique (EMP) of Algiers, Algeria, in 2005 and 2013 respectively. He received his Cotutelle PhD degree from the EMP, Algeria and the university of Orlèans, France in June 2021. His research interests are pbiomedical signal processing and neuromuscular physiology.

**Abdenour Hacine-Gharbi** received his Engineering and Magister degrees from University of Setif, Algeria in 1995 and 2002, respectively. He received his PhD from the University of Setif, Algeria and the University of Orlèans, France in 2012. He is currently an Assistant Professor at the Electronics Department of Bordj Bou Arreridj University, Algeria and member of LMSE Laboratory. His main research interests are pattern recognition, classification and feature selection.

**Philippe Ravier** received the Ph. D. degree in engineering science speciality signal, image and speech, from the Polytechnic National Institute of Grenoble (France) in 1998. Since 1999, he is assistant professor at Polytech Orléans, the engineering school of Orléans University, with habilitation to direct research obtained in 2011. His main researches deal with wavelet and time-frequency analysis with application to bioelectrical signals.

**Karim Abed-Meraim** was born in 1967. He received the State Engineering Degree from the Ecole Polytechnique, Palaiseau, France, in 1990, the State Engineering Degree from the Ecole Nationale Supèrieure des Tèlècommunications (ENST), Paris, France, in 1992, the M.Sc. degree from Paris XI University, Orsay, France, in 1992, and the Ph.D. degree in the field of signal processing and communications from ENST, in 1995. From 1995 to 1998, he was a Research Staff with the Electrical Engineering Department, The University of Melbourne, where he worked on several research project related to Blind System Identification for Wireless Communications, Blind Source Separation, and Array Processing for Communications. From 1998 to 2012, he has been an Assistant Professor, then an Associate Professor with the Signal and Image Processing Department, Tèlècom ParisTech. In September 2012, he joined the University of Orlèans, France (PRISME Laboratory), as a Full Professor. His research interests include signal processing for communications, adaptive filtering and tracking, array processing, and statistical performance analysis. He is the author of about 500 scientific publications, including book chapters, international journal and conference papers, and patents. He is an IEEE Fellow. Dr. Abed-Meraim is currently a member of the IEEE SAM-TC and a Senior Area Editor of the IEEE Transactions on Signal Processing.

**Olivier Buttelli** was born in 1966. He receivedthe MS and Ph.D. degrees in biomechanics andmotor physiology in 1991 and 1996 both from the University of Paris 11 Orsay, respectively. He is associated professor at University of Orléans, specialized in biomechanics, physiology of muscle and ergonomics. His main activities relate to fatigue processes and physical exercise and development of signal processing for electrophysiological signals.

# HPLC Method Development for the Fast Separation of a Complex Explosive Mixture

Benmalek Boulesnam, Fahima Hami, Djalal Trache, and Toudert Ahmed Zaid

**Abstract−The growing threat of terrorism in many parts of the world has called for the urgent need to find rapid and reliable means of analyzing explosives. This is in view to help forensic scientists to identify different swabs from post-blast debris. The present study aims to achieve an efficient separation and identification of a mixture of sixteen explosive compounds (including nitroaromatics, nitramines, and nitrate esters) by high performance liquid chromatography using a diode array detection (HPLC/DAD) and an Agilent Poroshell 120 EC-120 C$_{18}$ column at two wavelengths (235 and 214 nm). Relevant chromatographic parameters such as capacity factors, resolution, selectivity and number of theoretical plates have been optimized in order to achieve the best separation of the different components. In this respect, the effects of various parameters such as gradient time, column temperature, flow rate of mobile phase and initial percentage of organic mobile phase on the separation of these compounds were investigated. It was revealed that the method allowed a fairly acceptable separation of all the compounds in less than 15 minutes except for two isomers, namely 4-A-2,6-DNT, 2-A-4,6-DNT and 2,6- DNT which could not be resolved by the used C$_{18}$ column. These shortcoming notwithstanding, the authors believe the developed method produced satisfactory results and demonstrated sensitive and robust separation, furthermore indicating that the HPLC developed method can be both fast and efficient for the analysis of complex mixtures of explosive compounds.**

***Keywords−*HPLC method development, UV detection, explosives, optimization.**

## NOMENCLATURE

| | |
|---|---|
| HPLC | High Performance Liquid Chromatography |
| DAD | Diode Array Detection |
| FPD | Flame Photometric Detector |
| UV | Ultraviolet |
| US EPA | United States Environmental Protection Agency |
| F | flow rate (mL/min) |
| HRMS | High resolution mass spectrometry |
| MeOH | methanol |
| k | capacity factor |
| α | selectivity factor |
| N | number of theoretical plates |
| R | resolution |
| $t_R$ | retention time (min) |
| $t_G$ | gradient time (min) |
| XRD | X-ray diffractometry |
| XRF | X-ray fluorescence spectroscopy |
| $\Delta t_R$ | difference in retention times for two peaks (min) |

| | |
|---|---|
| $\Delta\varnothing$ | gradient range, equal to the final value of $\varnothing$ in the gradient ($\varnothing_f$) minus the initial value ($\varnothing_0$) |
| $\varnothing$ | volume fraction of B solvent in the mobile phase |
| $\varnothing_f$ | value of $\varnothing$ for mobile phase at end of gradient |
| $\varnothing_0$ | value of $\varnothing$ for mobile phase at start of gradient |

## I. INTRODUCTION

Conventional munitions constituents such as nitroaromatics, aminoaromatics, nitramines, and nitrate esters are the most common used organic high explosives by either armed forces or terrorist groups around the world. There is a current need to improve security/screening methods for explosive detection or recognition. Such methods can be applied for either environmental considerations or forensics. The stat-of-the-art of the analysis of such explosives during the last few decades has shown that the application of high performance liquid chromatography (HPLC) allowed obtaining a high degree of accuracy and precision. Such analytical tool presents an obvious advantage over gas chromatography, because it is carried out at room temperature and the above-mentioned explosives are known to present a low vapor pressure [1]. Being nondestructive, HPLC can be utilized for the combined analysis of both volatile and nonvolatile materials.

The analysis of explosives mixtures by liquid chromatography equipped by UV [2-6] or FPD [7] detectors has already been carried out. However, such combinations often required extending analysis time. Recent chromatography methods based on mass spectrometric detection have been revealed to be efficient [8-10, 11-13], due to high level of confirmation and accuracy. Nevertheless, UV absorbance detection remains one of the universal methods used in micro separations due to its simplicity, ease-of-use and low cost [5]. Furthermore, most of organic compounds can be analyzed by HPLC equipped by UV detectors. This latter displays further advantages such as rapidity, accessibility, durability, low toxicity and cost efficiency. However, it is demonstrated that the detection of

explosive mixtures is challenging because of poor mass transfer efficiencies and long analysis times.

At present, there is no simple method, which efficiently separates and quantifies munitions constituents or mixtures of explosives [6,14].

On the other hand, the identification of trace explosives can be extremely difficult because of the complexity of the different matrices that can be investigated due to their low content in explosive compounds. Thus, sophisticated analytical techniques that are sensitive, robust, fast and cost-effective are often required. A comprehensive review dealing with the high performance liquid chromatography methods for the analysis of explosives was reported by Gaurav et al. [5]. Mohamad Afiq Mohamed Huri et al. published an exhaustive review concerning the analysis of explosive residue from the forensic point of view [15]. This latter reported the approaches to track traces of explosives and the respective extraction methods. These authors provided a deep insight on the methods used to analyze the explosive residue as well. However, this research area remains an ongoing subject that needs more investigations to find new efficient approaches. Other advanced techniques such as nuclear magnetic resonance (NMR) have made it possible to identify the structural composition of explosives from post-blast debris [16] while combined techniques including HPLC-HRMS, XRD and XRF were used to gain fingerprints of various brands of explosives when including the analysis of additives and by-products [17]. These authors claim that these combined methods of analysis can be useful for the creation of a database on explosives that enables to assign specific formulations to certain manufacturers and countries of origin.

The objective of this work is to implement an analytical technique using HPLC equipped with photodiode array detector (HPLC/DAD) for the rapid separation of a mixture of 16 explosive substances by optimizing the separation through the variation of relevant chromatographic parameters such as capacity factors, resolution, selectivity and number of theoretical plates.

## II. Materials and Methods

### II.1. Chemicals and Materials

Explosive standards solutions used in this study were purchased from AccuStandard[TM] and supplied in a solvent in 1 mL size glass ampoules dissolved in acetonitrile (AcN) or methanol (MeOH) (or a mixture of both (AcN: MeOH (1:1)) at a concentration of 1000 μg/mL concentration. The list of the sixteen studied compounds is given in Table I. All solutions were stored in amber glass vials at 4°C to avert degradation.

Methanol, of HPLC-grade, purchased from VWR (Fontenay Sous Bois, France), was degassed prior to use. Organic-free reagent water was used as mobile phase in linear gradient elution mode.

**TABLE I**
EXPLOSIVE COMPOUNDS UNDER STUDY

| Compounds | Abbreviations |
| --- | --- |
| *Nitroaromatics and nitramines* | |
| Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine | HMX (C1) |
| Hexahydro-1,3,5-trinitro-1,3,5-triazine | RDX (C3) |
| 1,3,5-Trinitrobenzene | 1,3,5-TNB (C4) |
| 1,3-Dinitrobenzene | 1,3-DNB (C5) |
| Nitrobenzene | NB (C6) |
| 3,5-Dinitroaniline | 3,5-DNA (C7) |
| 2,4,6-Trinitrophenylméthylnitramine | Tetryl (C8) |
| 2,4,6-Trinitrotoluene | 2,4,6-TNT (C9) |
| 4-Amino-2,6-Dinitrotoluene | 4-A-2,6-DNT (C10) |
| 2-Amino-4,6-Dinitrotoluene | 2-A-4,6-DNT (C11) |
| 2,6-Dinitrotoluene | 2,6-DNT (C12) |
| 2-Nitrotoluene | 2-NT (C13) |
| 4-Nitrotoluene | 4-NT (C14) |
| 3-Nitotoluene | 3-NT (C15) |
| *Nitrate esters* | |
| Ethylene glycol dinitrate | EGDN (C2) |
| Pentaerythritoltetranitrate | PETN (C16) |

### II.2. Analytical Instrumentation

An Agilent Model 1200 HPLC, coupled with a diode array detector (DAD), was used to separate the mixture of explosives. Experimental monitoring and data acquisition are performed by using HPLC ChemStation for LC 3D systems, Rev. B.04.03 Software. The employed analytical column was an Agilent Poroshell 120 EC-120 $C_{18}$ (4.6 x 150 mm, 4 μm). The mobile phase was a MeOH – $H_2O$ mixture. The elution mode is a linear gradient from 5% to 100% of MeOH during 15 min, with a mobile phase flow rate of 1.2 mL/min. The column temperature was 25°C and two wavelengths were used for the detection (214 nm and 235 nm).

### II.3. Sample Preparation

Using explosive standard solutions (AccuStandard, purity > 99%), a mixture solution containing 10 ppm of each of the following compounds: NB, 2-NT, 3-NT, 4-NT, EGDN, 1,3-DNB, 2,6-DNT, 3,5-DNA, 2-A-4,6-DNT, 4-A-2,6-DNT, 1,3,5-TNB, RDX, 2,4,6-TNT, Tetryl, HMX, PETN was prepared by dissolving explosive standards solutions in methanol HPLC grade.

## III. Separation Optimization

The goal of the present study is to separate a mixture of sixteen (16) explosive substances with an adequate peak resolution higher than 1.5, and through a fast and complete separation. These requirements can be achievable by optimizing chromatographic parameters such as the capacity factor (k), factor of selectivity α, and the number of theoretical plate N. Once ''best'' values of k and α have been established (optimization of selectivity), the resolution and the run time will depend only on N. The experimental conditions that favor a fast separation include small particles and short columns of the stationary phase, in addition to high flow rate of the mobile phase. [18]

Based on literature and availability reasons, a $C_{18}$ packed column described in the above section 2 was selected. Poroshell 120 columns are based on superficially porous particle technology, which features a solid silica core and a porous outer layer providing higher chromatographic efficiencies, fast and high-resolution separations.

### III.1. Initial separation test

Starting with the initial separation conditions mentioned in section 2, the chromatogram given in Fig. 1 was obtained.

The elution mode (gradient) was justified by a $\Delta t_R / t_G$ Ratio > 4 (calculated value = 4.6). Based on their polarity, the studied compounds will leave the column, where the most polar one will be the first to be eluted. As can be seen, the majority of the compounds are well separated, except the compounds (C10, C11 and C12). The UV spectra of these compounds match those of 4-A-2,6-DNT, 2-A-4,6-DNT, and 2,6-DNT

respectively. These co-eluted compounds are indeed difficult to separate with most of commercial $C_{18}$ columns and are baseline resolved by specific columns such as Acclaim E1 Explosives Analytical Columns, designed for US EPA Method 8330 [19]. So, a separation optimization needs to be performed.
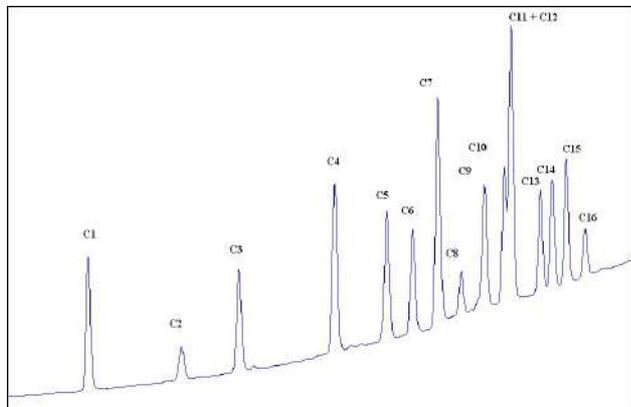


**Fig. 1:** Chromatogram of the mixture of explosives at 214 nm with concentration of 10 ppm. HPLC conditions: Poroshell 120 EC-120 $C_{18}$ (4.6 x 150 mm, 4 μm), MeOH-$H_2O$ mixture at flow rate of 1.2 mL/min, injection volume = 5 μL, elution mode: linear gradient from 5% to 100% of MeOH in $H_2O$ during 15 min, column temperature: 25°C.

### III.2. EFFECT OF CAPACITY FACTOR (k)

The capacity factor k in isocratic elution is usually controlled by varying the mobile-phase composition. In elution gradient mode, the variation of the applied gradient duration ($t_G$), affects the capacity factor (k). The usual separation goal is to reach $k \leq 10$ for all peaks because this corresponds to narrower and taller peaks, which improves the detection at short run times. Fig. 2 shows the effect of the increase of $t_G$ on the peak separation whereas Table II, Table III and Table IV display the effect of $t_G$ on the capacity factors (k), the number of theoretical plates (N), the resolutions (R) and selectivity (α) of each detected peak.
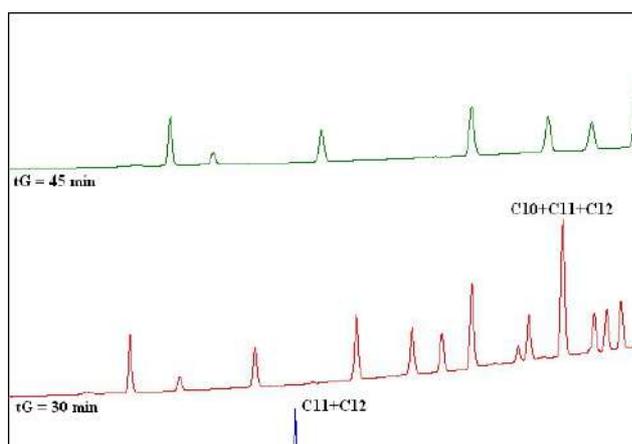


**Fig. 2:** Effect of $t_G$ increase on the separation profile. HPLC conditions: Poroshell 120 EC-120 $C_{18}$ (4.6 x 150 mm, 4 μm), MeOH-$H_2O$ mixture at flow rate of 1.2 mL/min, injection volume = 5 μL, elution mode: linear gradient from 5% to 100% of MeOH in $H_2O$ for 15, 30 and 45 min, column temperature: 25°C.

As can be seen from Fig. 2, the application of an elution gradient of 15 min allowed a good separation of the sixteen explosive compounds. However, the application of higher elution gradient durations of 30 min and 45 min resulted in an overlap "coelution" of the peaks (2,6-DNT, 2-A-4,6-DNT)

and (2,6-DNT, 2-A-4,6-DNT, 4-A-2,6-DNT) respectively. This indicates that $t_G = 15$ min is the most appropriate gradient duration with k values ranging between 4.0 and 9.2.

With the employment of Poroshell 120 EC-120 $C_{18}$ column, the retention of compounds increases in the following order: PETN > 3-NT > 4-NT > 2-NT > 2,6-DNT > 2-A-4,6-DNT > 4-A-2,6-DNT > 2,4,6-TNT > Tetryl > 3,5-DNA > NB > 1,3-DNB > 1,3,5-TNB > RDX > EGDN > HMX. It should be noted that no reversal elution occurred in the separation runs what can cause changes of the relevant parameters during the optimization steps, except when changing the initial percentage of the organic phase (MeOH). However, as was already mentioned, 4-A-2,6-DNT and 2-A-4,6-DNT isomers could not be separated and are co-eluted with 2,6- DNT.

**TABLE II**
INFLUENCE OF $t_G$ ON k

| $t_G$ (min) | | 15 | 30 | 45 |
|---|---|---|---|---|
| Compound | Peak order | | k | |
| HMX | 1 | 4.0 | 5.2 | 6.0 |
| EGDN | 2 | 5.0 | 6.2 | 6.8 |
| RDX | 3 | 5.6 | 7.6 | 9.0 |
| 1,3,5-TNB | 4 | 6.6 | 9.6 | 11.9 |
| 1,3-DNB | 5 | 7.1 | 10.7 | 13.4 |
| NB | 6 | 7.4 | 11.3 | 14.2 |
| 3,5-DNA | 7 | 7.7 | 11.9 | 15.1 |
| Tetryl | 8 | 7.9 | 9.5 | 16.9 |
| 2,4,6-TNT | 9 | 8.2 | 13.0 | - |
| 4-A-2,6-DNT | 10 | 8.4 | 13.7 | 18.1 |
| 2-A4,6-DNT | 11 | 8.4 | - | - |
| 2,6-DNT | 12 | - | - | - |
| 2-NT | 13 | 8.7 | 14.3 | 19.0 |
| 4-NT | 14 | 8.9 | 14.5 | 19.3 |
| 3-NT | 15 | 9.0 | 14.8 | 19.7 |
| PETN | 16 | 9.2 | 15.5 | 20.9 |

**TABLE III**
INFLUENCE OF $t_G$ ON N

| $t_G$ (min) | | 15 | 30 | 45 |
|---|---|---|---|---|
| Compound | Peak order | | N | |
| HMX | 1 | 50316 | 47341 | 42390 |
| EGDN | 2 | 49494 | 38099 | 33400 |
| RDX | 3 | 74861 | 58838 | 49587 |
| 1,3,5-TNB | 4 | 96123 | 92537 | 80411 |
| 1,3-DNB | 5 | 114878 | 106316 | 90758 |
| NB | 6 | 129999 | 120123 | 100272 |
| 3,5-DNA | 7 | 135268 | 128584 | 111341 |
| Tetryl | 8 | 158320 | 180379 | 113503 |
| 2,4,6-TNT | 9 | 157171 | 163642 | - |
| 4-A-2,6-DNT | 10 | 177625 | 151348 | 160783 |
| 2-A4,6-DNT | 11 | 164951 | - | - |
| 2,6-DNT | 12 | - | - | - |
| 2-NT | 13 | 206239 | 230409 | 217680 |

| | | | | |
|---|---|---|---|---|
| 4-NT | 14 | 214348 | 242032 | 203886 |
| 3-NT | 15 | 219374 | 248145 | 226031 |
| PETN | 16 | 228438 | 276265 | 256827 |

**TABLE IV**
INFLUENCE OF $t_G$ ON R AND α

| $t_G$ (min) | | **15** | | **30** | | **45** | |
|---|---|---|---|---|---|---|---|
| Compound | Peak order | R | α | R | α | R | α |
| HMX | 1 | - | - | - | - | - | - |
| EGDN | 2 | 10.0 | 1.25 | 7.4 | 1.19 | 5.5 | 1.14 |
| RDX | 3 | 5.9 | 1.12 | 10.1 | 1.24 | 12.1 | 1.31 |
| 1,3,5-TNB | 4 | 10.4 | 1.18 | 14.0 | 1.26 | 16.3 | 1.33 |
| 1,3-DNB | 5 | 5.6 | 1.08 | 7.6 | 1.11 | 8.0 | 1.13 |
| NB | 6 | 2.9 | 1.04 | 4.1 | 1.05 | 4.4 | 1.06 |
| 3,5-DNA | 7 | 2.8 | 1.04 | 4.1 | 1.05 | 4.6 | 1.06 |
| Tetryl | 8 | 2.7 | 1.03 | 6.6 | 1.08 | 9.0 | 1.12 |
| 2,4,6-TNT | 9 | 2.7 | 1.03 | 1.6 | 1.02 | - | - |
| 4-A-2,6-DNT | 10 | 2.3 | 1.03 | 4.6 | 1.05 | 5.5 | 1.07 |
| 2-A4,6-DNT | 11 | 0.8 | 1.01 | - | - | - | - |
| 2,6-DNT | 12 | - | - | - | - | - | - |
| 2-NT | 13 | 3.5 | 1.04 | 4.4 | 1.04 | 5.0 | 1.05 |
| 4-NT | 14 | 1.4 | 1.01 | 1.9 | 1.02 | 2.0 | 1.02 |
| 3-NT | 15 | 1.7 | 1.02 | 2.2 | 1.02 | 2.3 | 1.02 |
| PETN | 16 | 2.3 | 1.02 | 5.2 | 1.04 | 6.5 | 1.06 |

## III.3. SELECTIVITY (α) OPTIMIZATION

For a further improvement of the separation, relative retention (peak spacing, selectivity, or separation factor α) is then adjusted by varying the organic solvent, temperature or type of column [17]. In order to improve the resolution of co-eluting isomers, the column temperature was carried over the range of 25 – 38°C. The separation runs were performed under the following conditions: a MeOH-$H_2O$ mixture with a flow rate of 1.2 mL/min is used as a mobile phase and the linear gradient as elution mode is varied from 5% to 100% of MeOH in $H_2O$. The duration of the analysis is  around 15 min, whereas the column temperatures used are, respectively, 25, 28, 30, 32, 35 and 38°C. The corresponding chromatograms are shown in Fig. 3.

The effect of temperature on the capacity factors (k), the number of theoretical plates (N), the resolutions (R) and selectivity (α) of recorded peaks is shown in Table V, Table VI, Table VII and Table VIII respectively.

As can be seen from the chromatograms (Fig. 3), the two compounds (2,6-DNT and 2-A-4,6-DNT) overlap at temperatures of 25°C and 28°C while the peaks (4-A-2,6-DNT and 2-A-4,6-DNT) overlap at temperatures of 35°C and 38°C, respectively, also the peaks (NB and 3,5-DNA) overlap at temperatures of 38°C.

Regarding the capacity factors (k), they are all within the optimal domain, whatever the temperature is (Table V).

During the optimisation of column temperature, the co-elution of peaks of 4-A-2,6-DNT and 2-A-4,6-DNT, especially at 30 and 32°C, was the major issue as it had limited the resolution so far (R < 1.5). From Fig. 3 obtained, when running the samples at different temperatures, it was clearly seen that the

increase of the column temperature substantially affects the resolution of the NB and 3,5-DNA compounds, where the highest resolution of different compounds is obtained at T = 32°C.
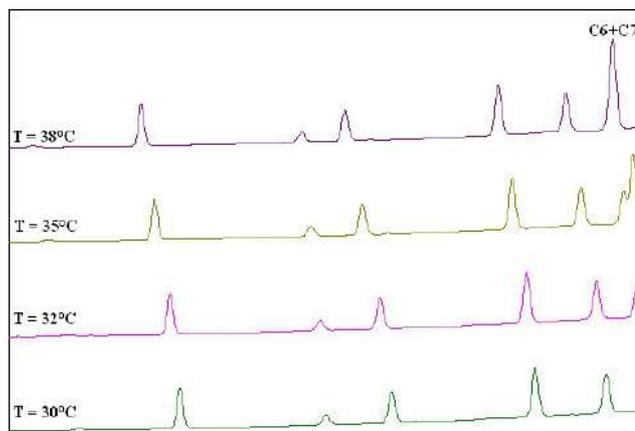


**Fig.3:** Effect of column temperature on separation profile. HPLC conditions: Poroshell 120 EC-120 C18 (4.6 x 150 mm, 5 μm), MeOH-$H_2O$ mixture at flow rate of 1.2 mL/min, injection volume = 5 μL, elution mode: linear gradient from 5% to 100% of MeOH in $H_2O$ for 15 min.

**TABLE V**
INFLUENCE OF COLUMN TEMPERATURE ON k.

| T (°C) | | **25** | **28** | **30** | **32** | **35** | **38** |
|---|---|---|---|---|---|---|---|
| Compound | Peak order | | | **k** | | | |
| C1 | 1 | 4.0 | 3.9 | 3.8 | 3.7 | 3.6 | 3.5 |
| C2 | 2 | 5.0 | 4.9 | 4.9 | 4.8 | 4.8 | 4.7 |
| C3 | 3 | 5.6 | 5.5 | 5.4 | 5.3 | 5.1 | 5.0 |
| C4 | 4 | 6.6 | 6.5 | 6.4 | 6.4 | 6.3 | 6.1 |
| C5 | 5 | 7.1 | 7.0 | 7 | 6.9 | 6.8 | 6.6 |
| C6 | 6 | 7.4 | 7.3 | 7.2 | 7.2 | 7.1 | 7.0 |
| C7 | 7 | 7.7 | 7.5 | 7.4 | 7.3 | 7.2 | - |
| C8 | 8 | 7.9 | 7.8 | 7.7 | 7.6 | 7.5 | 7.4 |
| C9 | 9 | 8.2 | 8.1 | 8 | 7.9 | 7.8 | 7.7 |
| C10 | 10 | 8.4 | 8.2 | 8.1 | 8.1 | 8.0 | 7.8 |
| C11 | 11 | 8.4 | 8.3 | 8.2 | 8.1 | - | - |
| C12 | 12 | - | - | 8.3 | 8.2 | 8.1 | 8.0 |
| C13 | 13 | 8.7 | 8.7 | 8.6 | 8.5 | 8.4 | 8.3 |
| C14 | 14 | 8.9 | 8.8 | 8.7 | 8.6 | 8.6 | 8.5 |
| C15 | 15 | 9.0 | 8.9 | 8.9 | 8.8 | 8.7 | 8.6 |
| C16 | 16 | 9.2 | 9.2 | 9.1 | 9.0 | 9.0 | 8.9 |

**TABLE VI**
INFLUENCE OF COLUMN TEMPERATURE ON N

| T (°C) | | **25** | **28** |
|---|---|---|---|
| Compound | Peak order | | **N** |
| HMX | 1 | 50316 | 49762 |
| EGDN | 2 | 49494 | 50174 |
| RDX | 3 | 74861 | 72544 |
| 1,3,5-TNB | 4 | 96123 | 97881 |
| 1,3-DNB | 5 | 114878 | 113353 |
| NB | 6 | 129999 | 131215 |
| 3,5-DNA | 7 | 135268 | 134113 |
| Tetryl | 8 | 158320 | 164861 |
| 2,4,6-TNT | 9 | 157171 | 160916 |
| 4-A-2,6-DNT | 10 | 177625 | 167242 |
| 2-A-4,6-DNT | 11 | 172478 | 51819 |
| 2,6-DNT | 12 | - | - |
| 2-NT | 13 | 206239 | 200383 |

| 4-NT | 14 | 214348 | 208222 |
| 3-NT | 15 | 219374 | 213310 |
| PETN | 16 | 228438 | 225669 |

**TABLE VI**
INFLUENCE OF COLUMN TEMPERATURE ON N (CONTINUED)

| T (°C) | | 30 | 32 | 35 | 38 |
|---|---|---|---|---|---|
| Compound | Peak order | | N | | |
| C1 | 1 | 49786 | 49370 | 47722 | 45752 |
| C2 | 2 | 51625 | 52159 | 51381 | 49621 |
| C3 | 3 | 71402 | 70862 | 68289 | 65721 |
| C4 | 4 | 99265 | 98391 | 95914 | 94143 |
| C5 | 5 | 112891 | 111894 | 109241 | 104377 |
| C6 | 6 | 130194 | 131786 | 126031 | 99043 |
| C7 | 7 | 133100 | 132649 | 115456 | - |
| C8 | 8 | 164761 | 161625 | 154060 | 147515 |
| C9 | 9 | 172095 | 168921 | 170255 | 129397 |
| C10 | 10 | 330984 | 377306 | 57513 | 63760 |
| C11 | 11 | 223214 | 315890 | - | - |
| C12 | 12 | 274220 | 217809 | 183513 | 168383 |
| C13 | 13 | 202108 | 200554 | 193880 | 185785 |
| C14 | 14 | 207674 | 208074 | 198628 | 193769 |
| C15 | 15 | 210452 | 210788 | 205318 | 199153 |
| C16 | 16 | 226431 | 216383 | 219000 | 219600 |

**TABLE VII**
INFLUENCE OF COLUMN TEMPERATURE ON R.

| T (°C) | | 25 | 28 | 30 | 32 | 35 | 38 |
|---|---|---|---|---|---|---|---|
| Compound | Peak order | | | R | | | |
| HMX | 1 | - | - | - | - | - | - |
| EGDN | 2 | 10.0 | 10.8 | 11.4 | 11.9 | 12.4 | 12.8 |
| RDX | 3 | 5.9 | 5.3 | 4.9 | 4.5 | 3.9 | 3.3 |
| 1,3,5-TNB | 4 | 10.4 | 10.9 | 11.2 | 11.5 | 11.8 | 12.1 |
| 1,3-DNB | 5 | 5.6 | 5.6 | 5.6 | 5.5 | 5.4 | 5.3 |
| NB | 6 | 2.9 | 3.0 | 3.2 | 3.3 | 3.4 | 3.5 |
| 3,5-DNA | 7 | 2.8 | 2.2 | 1.8 | 1.4 | 0.8 | - |
| Tetryl | 8 | 2.7 | 3.2 | 3.5 | 3.8 | 4.2 | 4.5 |
| 2,4,6-TNT | 9 | 2.7 | 2.9 | 3.1 | 3.2 | 3.4 | 3.3 |
| 4-A-2,6-DNT | 10 | 2.3 | 1.9 | 2.0 | 1.8 | 1.1 | 0.8 |
| 2-A-4,6-DNT | 11 | 2.3 | 0.7 | 0.7 | 0.8 | - | - |
| 2,6-DNT | 12 | - | - | 0.9 | 1.2 | 1.1 | 1.5 |
| 2-NT | 13 | 3.5 | 2.6 | 4.2 | 4.1 | 4.0 | 4.1 |
| 4-NT | 14 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
| 3-NT | 15 | 1.7 | 1.8 | 1.8 | 1.8 | 1.8 | 1.8 |
| PETN | 16 | 2.3 | 2.6 | 2.8 | 2.9 | 3.1 | 3.3 |

**TABLE VIII**
INFLUENCE OF COLUMN TEMPERATURE ON α.

| T (°C) | | 25 | 28 | 30 | 32 | 35 | 38 |
|---|---|---|---|---|---|---|---|
| Compound | Peak order | | | α | | | |
| HMX | 1 | - | - | - | - | - | - |
| EGDN | 2 | 1.25 | 1.27 | 1.28 | 1.3 | 1.32 | 1.34 |
| RDX | 3 | 1.12 | 1.11 | 1.1 | 1.09 | 1.08 | 1.07 |
| 1,3,5-TNB | 4 | 1.18 | 1.19 | 1.2 | 1.2 | 1.22 | 1.23 |
| 1,3-DNB | 5 | 1.08 | 1.08 | 1.08 | 1.08 | 1.08 | 1.08 |
| NB | 6 | 1.04 | 1.04 | 1.04 | 1.04 | 1.05 | 1.05 |
| 3,5-DNA | 7 | 1.04 | 1.03 | 1.02 | 1.02 | 1.01 | - |
| Tetryl | 8 | 1.03 | 1.04 | 1.04 | 1.05 | 1.05 | 1.06 |
| 2,4,6-TNT | 9 | 1.03 | 1.03 | 1.03 | 1.04 | 1.04 | 1.04 |
| 4-A-2,6-DNT | 10 | 1.03 | 1.02 | 1.02 | 1.02 | 1.02 | 1.01 |
| 2-A-4,6-DNT | 11 | 1.03 | 1.01 | 1.01 | 1.01 | - | - |
| 2,6-DNT | 12 | - | - | 1.01 | 1.01 | 1.02 | 1.02 |
| 2-NT | 13 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| 4-NT | 14 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 3-NT | 15 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| PETN | 16 | 1.02 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 |

### III.4. COLUMN EFFECTIVENESS OPTIMIZATION (N): EFFECT OF MOBILE PHASE FLOW RATE

A further separation improvement may be possible by varying some column conditions (such as the column length, the flow rate, and the particle size), in order to improve the column plate number N. The mobile phase flow rate effect is indeed investigated. The previously optimized conditions such as the capacity factor and the column temperature (32°C) are maintained. The flow rate variation is performed in an inverse trend to the duration of the elution gradient. Thus, a decrease in flow rate by one-half will correspond to double of the elution gradient duration. The chromatographic parameters obtained are presented in Table IX, Table X, Table XI, Table XII and Table XIII.
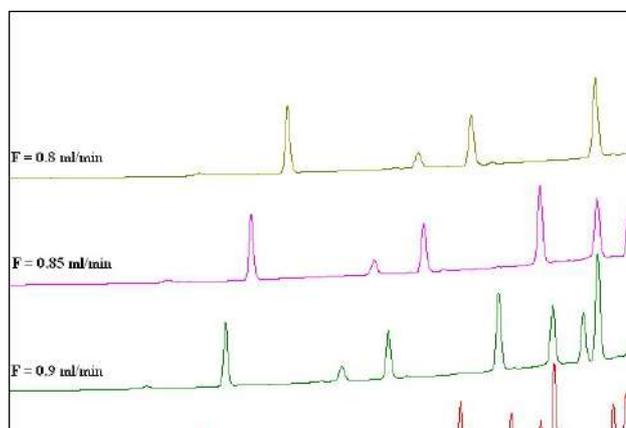


**Fig. 4:** Effect of flow rate on the separation performance. Separation conditions: elution mode: linear gradient from 5% to 100% of MeOH in $H_2O$ in 15 min, injection volume = 5 µL, column temperature: 32°C.

As can be seen from Fig. 4, the 16 compounds are all separated whatever the applied flow rate. It should be noted that at a flow rate of 0.7 mL/min, the time required for the separation of the sixteen substances is much longer, with an increase of roughly 50%. However, the capacity factors of the different separated substances remain within the optimum range. As was expected, the number of theoretical plates increases with a decrease of the flow rate.

Concerning the resolutions, it seems that their values are fairly close within the investigated domain of flow rates. Based on these observations, a compromise between $t_R$, R, and N was considered satisfactory for a flow rate of 0.9 mL/min.

**TABLE IX**
EFFECT OF LIQUID PHASE FLOW RATE ON $t_R$

| Flow(mL/min) | 0.8 | 0.85 | 0.9 | 0.95 | 1.0 |
|---|---|---|---|---|---|
| Compound | | | $t_R$ | | |
| HMX | 8.85 | 8.28 | 7.86 | 7.45 | 7.06 |
| EGDN | 10.95 | 10.25 | 9.72 | 9.22 | 8.73 |
| RDX | 11.81 | 11.05 | 10.47 | 9.95 | 9.41 |
| 1,3,5-TNB | 13.79 | 12.91 | 12.24 | 11.63 | 11.01 |
| 1,3-DNB | 14.76 | 13.82 | 13.10 | 12.45 | 11.79 |
| NB | 15.32 | 14.34 | 13.59 | 12.91 | 12.24 |
| 3,5-DNA | 15.58 | 14.58 | 13.82 | 13.13 | 12.45 |
| Tetryl | 16.17 | 15.14 | 14.36 | 13.64 | 12.95 |

| | | | | | |
|---|---|---|---|---|---|
| 2,4,6-TNT | 16.69 | 15.63 | 14.82 | 14.08 | 13.36 |
| 4-A-2,6-DNT | 16.94 | 15.86 | 15.04 | 14.28 | 13.56 |
| 2-A-4,6-DNT | 17.03 | 15.95 | 15.12 | 14.36 | 13.63 |
| 2,6-DNT | 17.18 | 16.09 | 15.25 | 14.49 | 13.76 |
| 2-NT | 17.79 | 16.67 | 15.80 | 15.01 | 14.25 |
| 4-NT | 18.02 | 16.88 | 16.00 | 15.20 | 14.44 |
| 3-NT | 18.30 | 17.14 | 16.25 | 15.44 | 14.66 |
| PETN | 18.77 | 17.57 | 16.67 | 15.83 | 15.04 |

**TABLE X**
EFFECT OF LIQUID PHASE FLOW RATE ON k.

| Flow (mL/min) | | 0.8 | 0.85 | 0.9 | 0.95 | 1.0 |
|---|---|---|---|---|---|---|
| Compound | Peak order | | | k | | |
| HMX | 1 | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
| EGDN | 2 | 4.9 | 4.8 | 4.9 | 4.9 | 4.8 |
| RDX | 3 | 5.3 | 5.3 | 5.3 | 5.3 | 5.3 |
| 1,3,5-TNB | 4 | 6.4 | 6.3 | 6.4 | 6.4 | 6.4 |
| 1,3-DNB | 5 | 6.9 | 6.9 | 6.9 | 6.9 | 6.9 |
| NB | 6 | 7.2 | 7.2 | 7.2 | 7.2 | 7.2 |
| 3,5-DNA | 7 | 7.3 | 7.3 | 7.3 | 7.3 | 7.3 |
| Tetryl | 8 | 7.7 | 7.6 | 7.6 | 7.7 | 7.7 |
| 2,4,6-TNT | 9 | 7.9 | 7.9 | 7.9 | 7.9 | 7.9 |
| 4-A-2,6-DNT | 10 | 8.1 | 8.0 | 8.0 | 8.1 | 8.1 |
| 2-A-4,6-DNT | 11 | 8.1 | 8.1 | 8.1 | 8.1 | 8.1 |
| 2,6-DNT | 12 | 8.2 | 8.1 | 8.2 | 8.2 | 8.2 |
| 2-NT | 13 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 |
| 4-NT | 14 | 8.6 | 8.6 | 8.6 | 8.7 | 8.7 |
| 3-NT | 15 | 8.8 | 8.7 | 8.8 | 8.8 | 8.8 |
| PETN | 16 | 9.0 | 9.0 | 9.0 | 9.1 | 9.1 |

**TABLE XI**
EFFECT OF LIQUID PHASE FLOW RATE ON N

| Flow (mL/min) | | 0.8 | 0.85 | 0.9 | 0.95 | 1.0 |
|---|---|---|---|---|---|---|
| Peak order | | | | N | | |
| C1 | 1 | 64723 | 63347 | 60882 | 57707 | 55814 |
| C2 | 2 | 58826 | 58660 | 57767 | 57622 | 55733 |
| C3 | 3 | 90062 | 86259 | 85604 | 82398 | 80347 |
| C4 | 4 | 124956 | 119581 | 117616 | 111927 | 109806 |
| C5 | 5 | 136523 | 133911 | 132714 | 128401 | 123456 |
| C6 | 6 | 159956 | 155109 | 155118 | 151431 | 146315 |
| C7 | 7 | 171166 | 164257 | 161346 | 155204 | 149189 |
| C8 | 8 | 200065 | 186948 | 198195 | 189478 | 182309 |
| C9 | 9 | 211432 | 198988 | 199255 | 193421 | 186085 |
| C10 | 10 | 207915 | 207383 | 219216 | 221841 | 211009 |
| C11 | 11 | 176601 | 179830 | 198470 | 197487 | 129095 |
| C12 | 12 | 219183 | 228468 | 270426 | 265321 | 214104 |
| C13 | 13 | 256388 | 248446 | 240285 | 233506 | 222446 |
| C14 | 14 | 263325 | 255489 | 248564 | 240460 | 230839 |
| C15 | 15 | 266250 | 264247 | 252705 | 245534 | 238434 |
| C16 | 16 | 319770 | 305640 | 288809 | 277017 | 264889 |

**TABLE XII**
EFFECT OF LIQUID PHASE FLOW RATE ON R

| Flow (mL/min) | | 0.8 | 0.85 | 0.9 | 0.95 | 1.0 |
|---|---|---|---|---|---|---|
| Compound | Peak order | | | R | | |
| HMX | 1 | - | - | - | - | - |
| EGDN | 2 | 13.1 | 13.1 | 12.8 | 12.7 | 12.5 |
| RDX | 3 | 5.1 | 5.0 | 4.9 | 5.0 | 4.8 |
| 1,3,5-TNB | 4 | 12.7 | 12.4 | 12.4 | 12.1 | 12.1 |
| 1,3-DNB | 5 | 6.1 | 6.1 | 6.0 | 5.9 | 5.8 |
| NB | 6 | 3.5 | 3.5 | 3.5 | 3.4 | 3.4 |
| 3,5-DNA | 7 | 1.7 | 1.7 | 1.7 | 1.6 | 1.6 |
| Tetryl | 8 | 4.0 | 3.9 | 4.0 | 4.0 | 4.0 |
| 2,4,6-TNT | 9 | 3.5 | 3.5 | 3.5 | 3.4 | 3.4 |

| 4-A-2,6-DNT | 10 | 1.7 | 1.6 | 1.7 | 1.7 | 1.7 |
|---|---|---|---|---|---|---|
| 2-A-4,6-DNT | 11 | 0.6 | 0.6 | 0.6 | 0.6 | 0.5 |
| 2,6-DNT | 12 | 1.0 | 1.0 | 1.1 | 1.1 | 0.9 |
| 2-NT | 13 | 4.3 | 4.3 | 4.4 | 4.4 | 4.1 |
| 4-NT | 14 | 1.6 | 1.6 | 1.6 | 1.6 | 1.5 |
| 3-NT | 15 | 2.0 | 2.0 | 1.9 | 1.9 | 1.9 |
| PETN | 16 | 3.4 | 3.3 | 3.3 | 3.2 | 3.2 |

**TABLE XIII**
EFFECT OF LIQUID PHASE FLOW RATE ON α

| Flow (mL/min) | | 0.8 | 0.85 | 0.9 | 0.95 | 1.0 |
|---|---|---|---|---|---|---|
| Compound | Peak order | | | α | | |
| HMX | 1 | - | - | - | - | - |
| EGDN | 2 | 1.30 | 1.30 | 1.30 | 1.30 | 1.30 |
| RDX | 3 | 1.10 | 1.10 | 1.10 | 1.09 | 1.09 |
| 1,3,5-TNB | 4 | 1.20 | 1.20 | 1.20 | 1.20 | 1.20 |
| 1,3-DNB | 5 | 1.10 | 1.10 | 1.10 | 1.08 | 1.08 |
| NB | 6 | 1.00 | 1.00 | 1.00 | 1.04 | 1.04 |
| 3,5-DNA | 7 | 1.00 | 1.00 | 1.00 | 1.02 | 1.02 |
| Tetryl | 8 | 1.00 | 1.00 | 1.00 | 1.04 | 1.05 |
| 2,4,6-TNT | 9 | 1.00 | 1.00 | 1.00 | 1.04 | 1.04 |
| 4-A-2,6-DNT | 10 | 1.00 | 1.00 | 1.00 | 1.02 | 1.02 |
| 2-A-4,6-DNT | 11 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 |
| 2,6-DNT | 12 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 |
| 2-NT | 13 | 1.00 | 1.00 | 1.00 | 1.04 | 1.04 |
| 4-NT | 14 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 |
| 3-NT | 15 | 1.00 | 1.00 | 1.00 | 1.02 | 1.02 |
| PETN | 16 | 1.00 | 1.00 | 1.00 | 1.03 | 1.03 |

III.5. EFFECT OF CHANGES IN THE GRADIENT

The usual goal of a change of the initial-%B is to shorten the run time, by removing the empty space in the early part of the gradient chromatogram. In the following section, the effects of a change of initial-%B are investigated by varying gradient time $t_G$ in proportion to $\Delta\varnothing$, thus holding $(\Delta\varnothing/t_G)$ constant.

Fig. 5 illustrates the effects of a change in initial-% methanol during the separation of the explosives mixture.

The effect of the initial %MeOH on the capacity factors (k), the resolutions (R), the number of theoretical plates (N) and selectivity (α) of recorded peaks is shown in Table XIV, Table XV, Table XVI and Table XVII respectively.

An increase in initial %MeOH results in a rather rapid elution of the first compounds, resulting in a decrease of the capacity factor values. In this case the acceptable values of (k) correspond to initial% in MeOH ≤ 30%, when values of (k) are varied between 1.4 and 7.1.

Regarding the peak resolutions, the obtained results show that an increase of the initial % in methanol causes a slight decrease of the resolutions. Once again, resolutions of peaks (C10, C11, and C12), are rather low. These peaks are those of compounds 2-A-2,4-DNT, 4-A-2,6-DNT and 2,6-DNT respectively, which are hardly separated on a $C_{18}$ column.
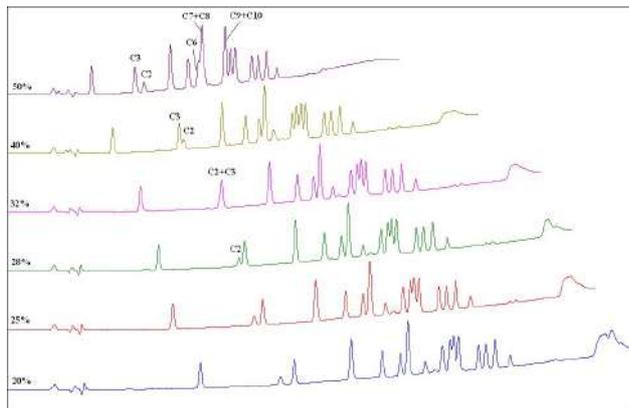
**Fig. 5:** Effect of a change in initial %Methanol for the gradient separation of the explosives mixture sample. Conditions: (150 × 4.6 mm, 4μm) $C_{18}$ Poroshell 120 column; temperature: 32°C; injection volume = 5 μL, Flow rate: 0.9 mL/min; initial % MeOH: 20, 25, 28, 32, 40 and 50.

It can be inferred that, by increasing the initial percentage % of the organic phase (MeOH), especially starting from 40%, the dilution order varied between EGDN (C2) and RDX (C3). As it is shown in Fig. 5, at 50 %, the peaks of 3,5-DNA and Tetryl overlap which induced the decreasing of the resolution.

**TABLE XIV**
EFFECT OF INITIAL % IN MeOH ON (k) OF RECORDED PEAKS

| Initial %MeOH | 20 | 25 | 28 | 32 | | 40 | 50 |
|---|---|---|---|---|---|---|---|
| | Peak order | | k | | Peak order | | k |
| C1 | 1 | 2.1 | 1.7 | 1.4 | 1.1 | 1 | 0.7 | 0.3 |
| C2 | 2 | 3.4 | 3.0 | 2.7 | 2.4 | 3 | 1.8 | 1.2 |
| C3 | 3 | 3.6 | 3.1 | 2.8 | - | 2 | 1.8 | 1.0 |
| C4 | 4 | 4.5 | 4.0 | 3.6 | 3.2 | 4 | 2.4 | 1.6 |
| C5 | 5 | 5.0 | 4.4 | 4.1 | 3.7 | 5 | 2.8 | 1.9 |
| C6 | 6 | 5.3 | 4.7 | 4.4 | 3.9 | 6 | 3.0 | 2.1 |
| C7 | 7 | 5.4 | 4.8 | 4.5 | 4.0 | 7 | 3.1 | 2.1 |
| C8 | 8 | 5.7 | 5.1 | 4.7 | 4.2 | - | 3.3 | - |
| C9 | 9 | 6.0 | 5.4 | 5.0 | 4.5 | 8 | 3.6 | 2.5 |
| C10 | 10 | 6.1 | 5.5 | 5.1 | 4.6 | - | 3.6 | - |
| C11 | 11 | 6.2 | 5.5 | 5.2 | 4.7 | 9 | 3.7 | 2.6 |
| C12 | 12 | 6.3 | 5.6 | 5.3 | 4.8 | 10 | 3.8 | 2.7 |
| C13 | 13 | 6.6 | 6.0 | 5.6 | 5.1 | 11 | 4.1 | 2.9 |
| C14 | 14 | 6.7 | 6.1 | 5.7 | 5.2 | 12 | 4.2 | 3.0 |
| C15 | 15 | 6.8 | 6.2 | 5.8 | 5.3 | 13 | 4.3 | 3.2 |
| C16 | 16 | 7.1 | 6.5 | 6.1 | 5.6 | 14 | 4.5 | 3.3 |

**TABLE XV**
EFFECT OF INITIAL % IN METHANOL ON N OF RECORDED PEAKS

| Initial %MeOH | 20 | 25 | 28 | 32 |
|---|---|---|---|---|
| Compound | Peak order | | N | |
| C1 | 1 | 25173 | 18389 | 14469 | 10623 |
| C2 | 2 | 38054 | 31584 | 31527 | 20142 |
| C3 | 3 | 47298 | 40452 | 35124 | - |
| C4 | 4 | 68290 | 55231 | 50332 | 41692 |
| C5 | 5 | 84048 | 70237 | 60910 | 52399 |
| C6 | 6 | 98017 | 81500 | 72286 | 62526 |
| C7 | 7 | 99605 | 83001 | 74647 | 63395 |
| C8 | 8 | 125853 | 105483 | 90365 | 79287 |
| C9 | 9 | 122221 | 105185 | 95014 | 84315 |
| C10 | 10 | 127185 | 101584 | 86813 | 70866 |
| C11 | 11 | 98589 | 77992 | 67183 | 63831 |
| C12 | 12 | 128035 | 111100 | 102257 | 90318 |
| C13 | 13 | 154715 | 132560 | 120192 | 103323 |
| C14 | 14 | 161318 | 136848 | 124075 | 107459 |
| C15 | 15 | 165945 | 141695 | 127752 | 111521 |
| C16 | 16 | 189782 | 161170 | 147357 | 127396 |

**TABLE XV**

**EFFECT OF INITIAL % IN METHANOL ON N OF RECORDED PEAKS (CONTINUED)**

| Initial %MeOH | | 40 | 50 |
|---|---|---|---|
| Compound | Peak order | N | N |
| C1 | 1 | 7674 | 5890 |
| C2 | 3 | 16425 | 13551 |
| C3 | 2 | 19210 | 10627 |
| C4 | 4 | 30198 | 18041 |
| C5 | 5 | 38452 | 23351 |
| C6 | 6 | 44241 | 42438 |
| C7 | 7 | 44456 | 18294 |
| C8 | 8 | 37916 | - |
| C9 | 9 | 37737 | 30126 |
| C10 | 10 | 37272 | - |
| C11 | 11 | 54038 | 35293 |
| C12 | 12 | 62762 | 36953 |
| C13 | 13 | 76108 | 46346 |
| C14 | 14 | 79384 | 48612 |
| C15 | 15 | 81875 | 51230 |
| C16 | 16 | 91153 | 56612 |

**TABLE XVI**
EFFECT OF INITIAL % IN METHANOL ON (R) OF RECORDED PEAKS

| Initial %MeOH | 20 | 25 | 28 | 32 | | 40 | 50 |
|---|---|---|---|---|---|---|---|
| Compound | Peak order | | R | | Peak order | | R |
| C1 | 1 | - | - | - | - | 1 | - | - |
| C2 | 2 | 15.4 | 15.6 | 15.9 | 14.7 | 3 | 0.8 | 1.9 |
| C3 | 3 | 2.5 | 1.6 | 1.1 | - | 2 | 13.9 | 9.5 |
| C4 | 4 | 10.8 | 10.3 | 10.0 | 8.7 | 4 | 7.4 | 5.6 |
| C5 | 5 | 5.9 | 5.8 | 5.7 | 5.5 | 5 | 4.8 | 3.7 |
| C6 | 6 | 3.6 | 3.5 | 3.4 | 3.3 | 6 | 2.9 | 2.5 |
| C7 | 7 | 1.5 | 1.4 | 1.4 | 1.3 | 7 | 1.1 | 0.7 |
| C8 | 8 | 3.6 | 3.3 | 3.1 | 2.8 | - | 1.7 | - |
| C9 | 9 | 3.6 | 3.6 | 3.7 | 3.8 | 8 | 3.4 | 4.4 |
| C10 | 10 | 1.5 | 1.5 | 1.4 | 1.3 | - | 0.7 | - |
| C11 | 11 | 0.6 | 0.6 | 0.7 | 0.7 | 9 | 0.8 | 1.1 |
| C12 | 12 | 0.9 | 0.9 | 0.9 | 0.9 | 10 | 1.0 | 1.0 |
| C13 | 13 | 4.1 | 4.1 | 4.2 | 4.1 | 11 | 4.0 | 3.7 |
| C14 | 14 | 1.6 | 1.5 | 1.5 | 1.5 | 12 | 1.5 | 1.4 |
| C15 | 15 | 2.0 | 1.9 | 1.9 | 1.9 | 13 | 1.9 | 1.8 |
| C16 | 16 | 3.2 | 3.1 | 3.1 | 3.0 | 14 | 2.8 | 2.2 |

**TABLE XVII**
EFFECT OF INITIAL % IN METHANOL ON α OF RECORDED PEAKS

| Initial % MeOH | 20 | 25 | 28 | 32 | | 40 | 50 |
|---|---|---|---|---|---|---|---|
| | Peak order | | α | | Peak order | | α |
| C1 | 1 | - | - | - | - | 1 | - | - |
| C2 | 2 | 1.62 | 1.79 | 1.93 | 2.17 | 3 | 1.04 | 1.14 |
| C3 | 3 | 1.07 | 1.05 | 1.03 | - | 2 | 2.61 | 3.12 |
| C4 | 4 | 1.25 | 1.28 | 1.29 | 1.32 | 4 | 1.34 | 1.36 |
| C5 | 5 | 1.11 | 1.12 | 1.13 | 1.14 | 5 | 1.16 | 1.18 |
| C6 | 6 | 1.06 | 1.06 | 1.07 | 1.07 | 6 | 1.08 | 1.09 |
| C7 | 7 | 1.02 | 1.02 | 1.02 | 1.03 | 7 | 1.03 | 1.02 |
| C8 | 8 | 1.05 | 1.05 | 1.05 | 1.05 | - | 1.05 | - |
| C9 | 9 | 1.05 | 1.05 | 1.06 | 1.07 | 8 | 1.09 | 1.18 |
| C10 | 10 | 1.02 | 1.02 | 1.02 | 1.02 | - | 1.02 | - |
| C11 | 11 | 1.01 | 1.01 | 1.01 | 1.01 | 9 | 1.02 | 1.03 |
| C12 | 12 | 1.01 | 1.01 | 1.02 | 1.02 | 10 | 1.02 | 1.03 |
| C13 | 13 | 1.05 | 1.06 | 1.06 | 1.07 | 11 | 1.08 | 1.10 |
| C14 | 14 | 1.02 | 1.02 | 1.02 | 1.02 | 12 | 1.03 | 1.04 |
| C15 | 15 | 1.02 | 1.02 | 1.03 | 1.03 | 13 | 1.03 | 1.04 |
| C16 | 16 | 1.04 | 1.04 | 1.04 | 1.04 | 14 | 1.05 | 1.05 |
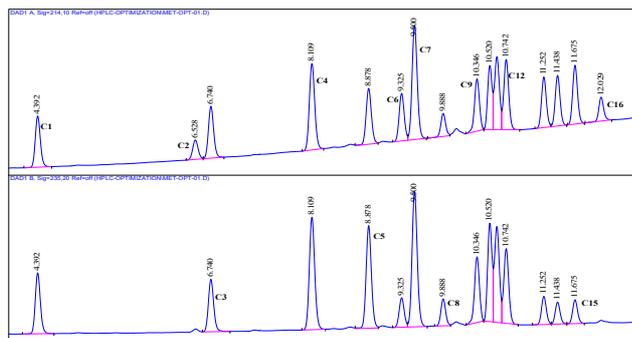
**Fig. 6:** Chromatogram of separation of 16 explosives after optimization steps.

Based on the above optimization approaches, a compromise given below was obtained for the best separation conditions (Fig. 6):

- Column: Agilent Poroshell 120 EC-120 $C_{18}$ (150 x 4.6 mm, 4μm);
- Mobile phase: MeOH – $H_2O$;
- Column Temperature: 32°C;
- Liquid phase flow rate: 0.9 mL/min;
- Elution mode: linear gradient 25 à 100% in MeOH during 15 min.

## IV. CONCLUSION

An Agilent Poroshell 120 EC-120 $C_{18}$ (4.6 x 150 mm, 4 μm) was used to provide a means for a rapid screening of a mixture of 16 commercial and military grade explosives. The optimization of relevant chromatographic parameters allowed achieving a fairly acceptable separation of all the compounds in less than 15 minutes except for the amino-dinitrotoluene isomers (2-A-4,6-DNT and 4-A-2,6-DNT and 2,6-DNT) which could not be resolved by the $C_{18}$ column whatever the conditions that were chosen. This method is intended to be applied to the analysis of soils contaminated with explosive residues. The optimized chromatographic technique will be used after solid phase extraction preparation method, in order to purify the samples by eliminating all sorts of contaminations. This activity is currently under way.

## REFERENCES

[1] Mauriche Marshall, Jimmie C. Oxley, "Aspects of Explosives Detection", 2009, Elsevier.
[2] W. Abu-Qare and B. Abou-Donia, Simultaneous analysis of sarin, pyridostigmine bromide and their metabolites in rat plasma and urine using HPLC, Chromatographia, 53: 251. 2001. https://doi.org/10.1007/BF02490419
[3] L. Shanmao, L. Wei, Z. Boli, Y. Lijun and W. Qinpel, The Determination of Degradation Products of Lewisite and/or Mustard Gas in Water by High Performance Liquid Chromatography, S. Afr. J. Chem., 58: 82. 2005.
[4] Dhingra Gaurav , Ashok Kumar Malik, P.K.Rai, Development of a new SPME–HPLC–UV method for the analysis of nitro explosives on reverse phase amide column and application to analysis of aqueous samples, Journal of Hazardous Materials, Vol 172, Issues 2–3, 30 December 2009, https://doi.org/10.1016/j.jhazmat.2009.08.039
[5] Dhingra Gaurav, Ashok Kumar Malik & P. K. Rai : High-Performance Liquid Chromatographic Methods for the Analysis of Explosives, Critical Reviews in Analytical Chemistry, 37:4, 2007. http://dx.doi.org/10.1080/10408340701244698
[6] Ronita L. Marple, William R. LaCourse, A platform for on-site environmental analysis of explosives using high performance liquid chromatography with UV absorbance and photo-assisted

[7] electrochemical detection, Talanta, Volume 66, Issue 3, 30 April 2005, https://doi.org/10.1016/j.talanta.2004.11.034
[7] E.W.J. Hooijschuur, C.E. Kientz ,U.A.Th. Brinkman, Application of microcolumn liquid chromatography and capillary electrophoresis with flame photometric detection for the screening of degradation products of chemical warfare agents in water and soil, J. Chromatogr. A, 928 : 187-199. 2001.
[8] A.L.Russell, J.M.Seiter, J.G.Coleman, B.Winstead, A.J.Bednar, Analysis of munitions constituents in IMX formulations by HPLC and HPLC–MS (2014), Talanta, Vol 128, 1 October 2014, pp. 524-530, https://doi.org/10.1016/j.talanta.2014.02.013
[9] XiaomaXu, MattijsKoeberg, Chris-Jan Kuijpers, Eric Kok, Development and validation of highly selective screening and confirmatory methods for the qualitative forensic analysis of organic explosive compounds with high performance liquid chromatography coupled with (photodiode array and) LTQ ion trap/Orbitrap mass spectrometric detections (HPLC-(PDA)-LTQOrbitrap) (2014), Science & Justice, Vol 54, Issue 1, January 2014, pp. 3-21, https://doi.org/10.1016/j.scijus.2013.08.003
[10] Michael Astratov, Alfred Preiβ, KarstenLevsen, GeroldWünsch, Identification of pollutants in ammunition hazardous waste sites by thermospray HPLC/MS (1997), International Journal of Mass Spectrometry and Ion Processes, Volumes 167–168, pp. 481-502, https://doi.org/10.1016/S0168-1176(97)00089-X
[11] R. Kinghorn, C. Milner, Analysis of Trace Residues of Explosive Materials by Time-of-Flight LC/MS, Agilent Technologies Application Note, publication number 5989-2449EN, 2005.
[12] P. Stone, J. Zweigenbaum, An Application Kit for the Screening of Samples for Analytes of Forensic and Toxicological Interest using TOF or Q-TOF LC/MS with a Personal Forensics/Toxicology Database, Agilent Technologies Application Note, publication number 5990-4252EN, 2009, http://www.agilent.com/cs/library/applications/5990-4252en.pdf
[13] E. Michael Thurman, I. Ferrer, Adapting EPA Method 8330B for Analysis of Explosives in Water to SPE and LC/MS/MS, Agilent Technologies Application Note, Part Number 5991-0676EN, 2012, http://www.agilent.com/cs/library/applications/5991-0676EN.pdf
[14] Michael J. Lang, Susan E. Burns, Improvement of EPA method 8330: complete separation using a two-phase approach, Journal of Chromatography A, 849 (1999) 381–388, https://doi.org/10.1016/S0021-9673(99)00591-9
[15] MohamadAfiq Mohamed Huri, UmiKalthom Ahmad, Roliana Ibrahim, Mustafa Omar, A review of explosive residue from forensic chemistry perspective, Malaysian Journal of Analytical Sciences, Vol 21 No 2 (2017) pp.267 – 282, https://doi.org/10.17576/mjas-2017-2102-01
[16] Priyanka Munjal et al. Identification and analysis of organic explosives from post-blast debris by nuclear magnetic resonance, Journal of Hazardous Materials, Volume 403, 5 February 2021, 124003
[17] Tilo Daniel Schachel, Alexandra Stork, Rasmus Schulte-Ladbeck, Torsten Vielhaber, Uwe Karst, Identification and differentiation of commercial and military explosives via high performance liquid chromatography – high resolution mass spectrometry (HPLC-HRMS), X-ray diffractometry (XRD) and X-ray fluorescence spectroscopy (XRF): Towards a forensic substance database on explosives, Forensic Science International, Volume 308 , mars 2020 , 110180, https://doi-org.sndl1.arn.dz/10.1016/j.forsciint.2020.110180
[18] Lloyd R. Snyder, Joseph J. Kirkland, John W. Dolan, Introduction to Modern Liquid Chromatography, Wiley. 2009.
[19] Xiaodong Liu et al., A Total Solution to Baseline Separation of 14 Explosives in U.S. EPA Method 8330, 2007. Available on line: http://www.cromlab.es/Articulos/Columnas/HPLC/Thermo/Acclaim/EXP/56197-ALN-Reprint.pdf

**Benmalek BOULESNAM** was born on February 21, 1979 in Algeria. In December 2005, he holds a master's degree in Chemical Engineering from the Ecole Nationale Polytechnique and is currently preparing a doctoral thesis on the fast analysis of complex mixtures of explosives. He works within the National Institute of Criminalistics and Criminology (INCC) of the Gendarmerie Nationale as a Judicial Expert in Analysis and Investigation of Fire & Explosives. Since 2018, he has also held the position of Head of the Laboratory for the Analysis of hazardous Chemicals at the INCC.

**Fahima HAMI** was born on January 07, 1988 in Tizi-Ouzou. In December 2018, she holds a master's degree in Chemical Engineering from the Military Polytechnic School of Bordj El Bahri. She works within the National Institute of

Criminalistics and Criminology (INCC) of the Gendarmerie Nationale as a Judicial Expert in Trace Analysis.



**Djalal Trache**, born on November 17, 1982 in M'Chedallah (Bouira), has been working as an Associate Professor at Ecole Militaire Polytechnique (EMP), Algeria, since 2016. He is an Associate Editor for Applied Nanoscience and Journal of Nanostructure in Chemistry (Springer) and a member of the editorial board of several reputed journals. He has made several presentations at national and international conferences, published over 100SCI scientific papers in the field of chemical sciences/materials science in mainstream journals such as ChemEng J, Green Chem, CarbohydrPolym, Nanoscale, and Fuel. He has also published eight book chapters and three books. His total citation is more than 3000 times, including 6 ESI highly cited papers with an h index of 27. He is a reviewer for more than 80 international journals in prestigious publishers such as Springer, Nature, Science Direct, Wiley, Taylor and Francis, ACS, and RSC. He has particular expertise in energetic materials, bio-based materials, polymer composites, and their characterizations. He also has interests in nanomaterials and their applications, phase equilibria and kinetics. Besides, he has successfully supervised many engineer, MSc and doctoral students.



**Toudert AHMED ZAID**, born on August 15, 1954 in Ain El Hammam (Tizi-Ouzou), is Professor at the Chemical Engineering Department (ENP, Algiers). He has an experience of 20 years in the domain of surfactants and dispersed systems. He has authored or co-authored more than thirty papers mainly in this research area.