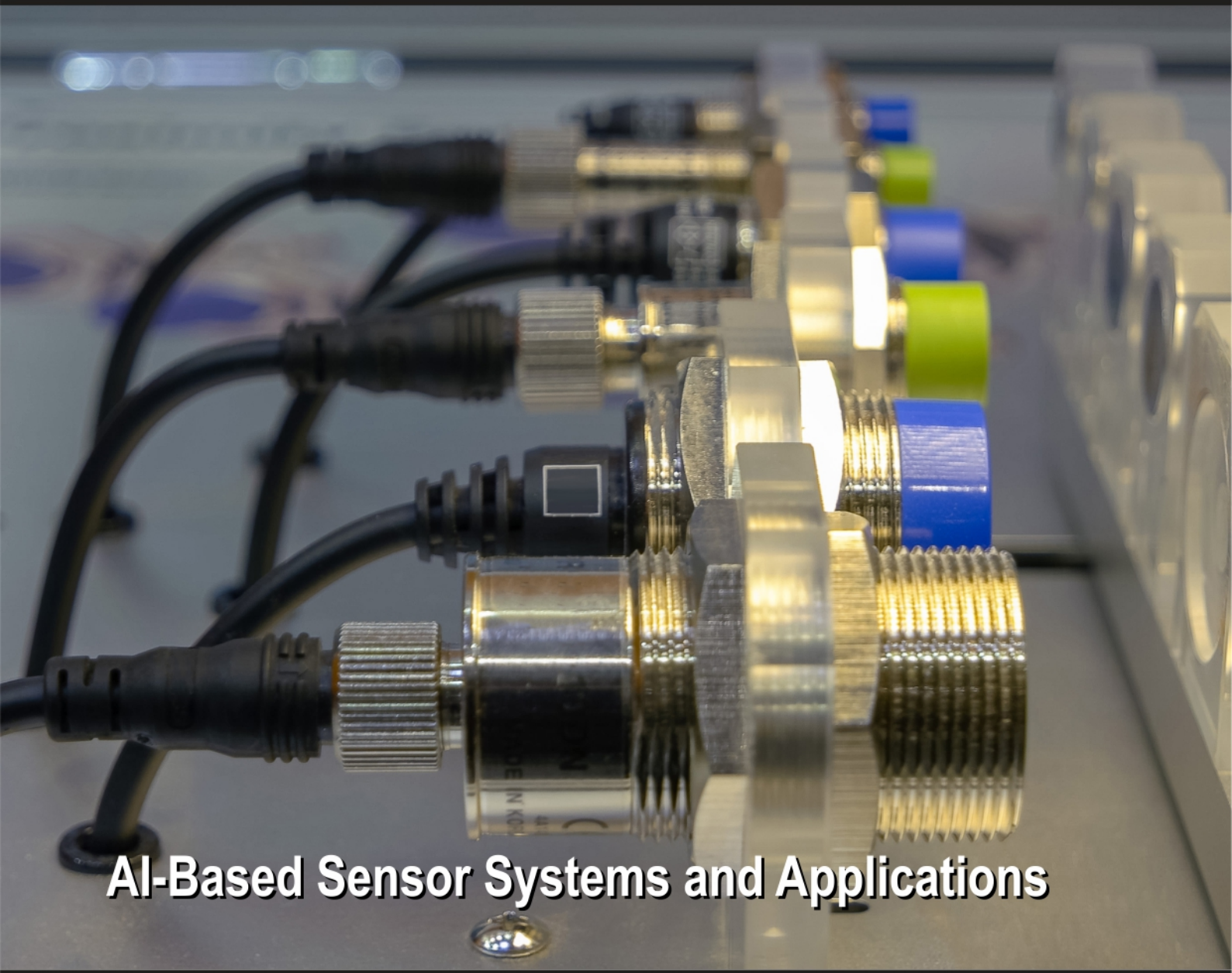
ISSN 1726-5479

SENSORS & TRANSDUCERS

vol. 268
1/25



AI-Based Sensor Systems and Applications

International Frequency Sensor Association Publishing

IFSA

Sensors & Transducers

International Official Open Access Journal of the International Frequency Sensor Association (IFSA) Devoted to Research and Development of Sensors and Transducers

Volume 268, Issue 1, April 2025

IFSA Publishing: Barcelona • Toronto

Published by International Frequency Sensor Association (IFSA) Publishing. Printed in the USA.

Sensors & Transducers

Volume 268, Issue 1
April 2025

www.sensorsportal.com

e-ISSN 1726-5479
ISSN 2306-8515

Sensors & Transducers journal is an open access, peer review international journal published monthly by International Frequency Sensor Association (IFSA).
Available in both: print and electronic (printable pdf) formats.

Sensors & Transducers Journal

Contents

Volume 268
Issue 1
April 2025

www.sensorsportal.com

ISSN 2306-8515
e-ISSN 1726-5479

Research Articles

Flexible Electronic Skin Sensors: Review
Issahaku Walaman-I, Ning Yang, Jing Bai and Debao Zhou 1

Identification and Online-correction of Periodic Motor Torque Fluctuations on Electromechanical Axes
Chris Schoeberlein, Armin Schleinitz, Holger Schlegel and Martin Dix 12

Vehicle Localization Algorithm with Lane Discrimination Based on Inertial and Geomagnetic Sensor Data for GNSS-Denied Environments and Its Evaluation for Different Vehicles and Drivers
Takayoshi Yokota 19

Robotic Automation for Nuclear Decommissioning: Development of a Tool Carrier System with Milling and Clearance Measurement Capabilities
Siavash Kazemi 28

A COTS-based Lightweight, Low-power and Versatile Companion Computer for Nano UAVs
Rintaro Okudera, Yixiao LI, Yutaka Matsubara and Hiroaki Takada 37

A Hybrid Machine Learning and Physics-based Approach for Accurate Energy Consumption Modeling of Electric Buses in Public Transport
Lucas Adam, Robert Pellerin, and Bruno Agard 45

Adoption of Post-Quantum Cryptography and FIPS Standards in Existing as well as Emerging Communication Technologies
Vaghawan Ojha, Sumit Chauhan, Shanita Yarahmadhian and David Carvalho 59

Transfer Learning-driven Comparative Analysis of GA, PSO, and NSGA-II Over FIS for Enhanced Energy Efficiency in Assisted Living Settings
Anita Xhemali , Elma Zanaj , Gledis Basha and Lorena Balliu 66

Authors are encouraged to submit article in MS Word (doc) and Acrobat (pdf) formats
by e-mail: editor@sensorsportal.com. Please visit journal's webpage with preparation instructions:
https://sensorsportal.com/st_guide_for_authors.html

International Frequency Sensor Association (IFSA).

11th International Conference on Sensors and Electronic Instrumentation Advances

https://www.seia-conference.com/

24-26 September 2025
Ponta Delgada, San Miguel (Azores Islands), Portugal

SÃO MIGUEL SEIA'25

IFSA

The 11th annual SEIA' 2025 Conference is a key platform for sharing the latest in sensor and electronic instrumentation advancements, including AI-based sensor systems. It attracts global experts from academia and industry to discuss innovations in sensors, electronics, and measurement technologies. Since 2015, the event has been held in various locations, with the 2025 edition scheduled for 24-26 September in Ponta Delgada, San Miguel (Azores Islands), Portugal, alongside the 7th IFSA Frequency and Time Conference and the 7th MicDAT Conference under the same conference umbrella. Organized by the International Frequency Sensor Association (IFSA) - one of the main professional association serving the sensor industry and academy since 1999, the SEIA' 2025 also emphasizes networking, featuring social events like a welcome cocktail, Gala Dinner, and farewell cocktail.

Sensor Technology

- Accelerometers
- Inclinometers
- Gyroscopes
- Mechanical Sensors
- Optical Sensors
- Optical Fiber Sensors
- Photonic Sensors
- Chemical Sensors
- Biosensors
- Immunosensors
- BioMEMS
- Temperature Sensors
- Pressure Sensors
- Acoustic Sensors
- Electromagnetic Sensors
- Gas Sensors
- Humidity Sensors
- Infrared Sensors, Devices and Thermography
- Radiation Sensors
- Multi Sensor Fusion
- Smart Sensors
- Intelligent Sensors
- Artificial Intelligence based Sensors and Sensor Systems
- Virtual Sensors
- Sensor Interfacing and Signal Conditioning
- Sensor Calibration
- Nanomaterials and Electronics Technology for Sensors
- Semiconductor Materials for Sensors
- Polymer Materials for Sensors
- MEMS and NEMS
- Remote Sensors and Telemetry
- Sensor Applications

Sensor Instrumentation and Measuring Technology

- Metrology and Measurement Science
- Methods of Measurements
- Calibrations and Standards
- Measurement of Electrical Quantities
- Time and Frequency Measurements
- Measurement of Force, Mass, Torque, Inclination and Acceleration
- Magnetic Measurements
- Hardness Measurements
- Measurement of Geometrical Quantities
- Temperature and Thermal Measurements
- Pressure and Vacuum Measurements
- Vibration and Noise Measurement
- Flow Measurements
- Chemical Measurements
- Quantum Measurements and Photonics
- Acoustics and the Ultrasonic Measurements
- Environmental Measurements
- Power and Energy Measurements
- Measurement of Human Functions
- Measurements in Biology and Medicine
- Mathematical Tools for Measurements
- Optical and Radiation Measurements
- Microwave Measurements
- Virtual Instruments and Data Acquisition Systems
- Soft Measurements
- Measurement Systems
- Distributed Measurements
- Analog-to-Digital Converters, Digital and Mixed Signal Processing
- Waveform Analysis and Measurements
- Scientific and Industrial Instrumentation
- Cyber-Physical Systems and IoT
- Experimental Mechanics
- Measurement in Robotics
- Metrology in Food and Nutrition
- Intelligent and Computer Vision Instruments
- Reliability of Instrument and Measurement Systems
- Nanometrology
- Technical Diagnostics and Testing
- Nondestructive Testing
- Blockchain Applications in Metrology and Measurements
- Education and Training in Measurement and Instrumentation

Special Sessions:

Authors are welcome to propose and manage special sessions during the SEIA' 2025. Each special session will contain 4-6 papers in a related field as specified above.

Session organizers will get: Certificate of Appreciation; Free Registration; Special Publishing Theme within Conference Proceedings.

Contribution Types:

- Keynote Presentations
- Invited Talks
- Special Sessions
- Regular Sessions
- Poster Sessions
- Special Sessions
- Online Sessions
- Exhibition

Deadlines:

Submission (2-page extended abstract):	**30 May 2025**
Notification of acceptance:	**20 June 2025**
Registration:	**25 July 2025**
Camera ready and late registration:	**25 August 2025**

11th International Conference on Sensors and Electronic Instrumentation Advances

SÃO MIGUEL SEIA'25

https://www.seia-conference.com/

One event - three different publications !

1) All registered abstracts will be published in the conference proceedings (with the ISBN and DOI). Some previous SEIA Conference Series Proceedings are indexed in Conference Proceedings Citation Index (CPCI), Web of Science by Clarivate Analytics.

2) Authors will be invited to submit full-page extended papers to special issue of open access journal: Sensors & Transducers (ISSN: 2306-8515, e-ISSN: 1726-5479), published by IFSA Publishing, and indexed in Scopus.

3) The papers, presented at SEIA' 2025 may be extended to open access book chapters for the Book Series on 'Advances in Sensors', Vol. 9/10, 'Advances in Biosensors', Vol. 4 or 'Advances in Measurements and Instrumentation', Vol. 3/4. Some previous Book Series's volumes are indexed in the Book Citation Index (Wed of Science) by Clarivate Analytics.

Conference Hotel

The SEIA' 2025 Conference will take place in the Octant Hotel.
Address: Av. Dr. João Bosco Mota Amaral, 4 9500-771 Ponta Delgada, San Miguel, Portugal.

Social Programme

- Welcome Cocktail: 23 September 2025 (20:00-21:30).
- Gala Dinner: 25 September 2025 (20:00-23:00)
- Farewell Cocktail: 26 September 2025, Friday (during the poster session)

Organizing Committee

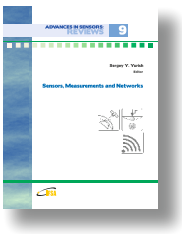
Chairman
Prof., Dr. Sergey Y. Yurish (IFSA, Spain),
e-mail: seia@sensorsportal.com

Advisory Chairman
Prof. Vijyakumar Varadarajan (Ajeenkya D Y Patil University India & La Trobe University, Australia)

Conference and Publication Manager
Tetyana Zakharchenko (IFSA Publishing, S.L., Spain)

Steering Committee:

Prof. Gennaro Conte (University Roma Tre, Italy)
Dr. Pavel Shuk (Saint-Gobain NorPro, USA)
Dr. Marius Gheorghe (Ideal Aerosmith, Inc., USA)
Dr. Paolo Dabove (Politecnico di Torino, Italy)
Prof. George Kiriakidis (European Materials Research Society (E-MRS), France)
Prof., Dr. Arkady Zhukov (Univ. of the Basque Country, Spain)

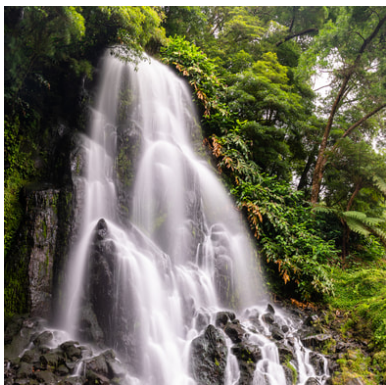


IFSA Publishing

F6S network partner

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY



Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 1-11

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Flexible Electronic Skin Sensors: Review

[1, *] Issahaku WALAMAN-I, [2] Ning YANG, [1] Jing BAI and [1] Debao ZHOU

[1] University of Minnesota Duluth, Engineering 140, 1303 Ordean Court, Duluth, MN, 55812, USA
[2] Tiangong University, No. 399 Bin Shui Xi Road, Xi Qing District, Tianjin, 300387, P.R. China
[1] Tel.: (218) 726 6648
[*] E-mail: walam001@d.umn.edu

Received: 5 Nov. 2024 /Revised: 20 Jan. 2025 /Accepted: 5 March 2025/ Published:30 April 2025

Abstract: Traditional rigid electronics face critical challenges of non-conformability to multiple shapes. Flexible and stretchable electronics offer unique solutions to such critical challenges by allowing conformability to various shapes. Electronic skin (E-skin) sensors have emerged from the field of flexible electronics to mimic human skin's functional properties. By mimicking such properties, e-skin sensors have potential applications in several fields ranging from healthcare to robotics. This paper provides a comprehensive review of advancements in e-skin sensor technologies, highlighting sensing mechanisms used to convert external stimuli to electric signals, various sensor structures for effectively elevating sensing capabilities, and the materials utilized to achieve flexible and stretchable e-skins. We also discuss the potential application areas of e-skin sensor technologies and the challenges affecting e-skin sensor technologies. Achieving multi-functional e-skin sensors capable of simultaneously detecting various stimuli while maintaining mechanical properties like the human skin presents unique opportunities for future innovations in diverse fields.

Keywords: Flexible electronics, E-skin sensors, Substrates, Electrodes, Sensing mechanisms.

1. Introduction

The field of flexible electronics is an emerging research area where electronic circuits are mounted on flexible substrates to enable their application on curvilinear surfaces. Current electronic devices are mainly realized by mounting electronic components on rigid substrates [1]. Flexible and stretchable electronic devices provide an alternative to their rigid counterparts as they conform to curvilinear surfaces and detect signals more accurately and precisely. Work towards these uniquely structured electronic devices began approximately three decades ago and has seen significant progress over recent years [2].

Electronic skin (E-skin) sensors evolved from flexible electronics mainly due to the need for flexible, sensitive, and stable sensors that could be applied in various areas. These sensors are designed to satisfy several requirements for conformable pressure sensing on curvilinear surfaces. Their functional requirements demand that they achieve high-performance sensing even under different deformation modes, including bending, twisting, and stretching. Due to their flexibility and stretchability, e-skins have potential applications in wearable electronics, human health monitoring, robotics, and medical devices [2].

Despite the promise these sensors show in diverse fields, they are still plagued by challenges in fulfilling design requirements of high performance and reliability while being subjected to various modes of deformation [2]. The drive towards adapting e-skins in healthcare monitoring and robotics poses further challenges for these sensors to achieve multimodal sensing capabilities, thus sensing more than a single source of stimuli [3]. However, there have been growing efforts to reduce the limitations of e-skin sensors, as will be discussed later in this paper. Addressing these challenges requires interdisciplinary collaboration across several fields.

2. Sensing Mechanisms for E-skins

The mode of signal transduction is an essential consideration for developing e-skin sensors. By studying human skin composition, sensing mechanisms for e-skins have advanced. Some sensing mechanisms widely employed for developing e-skins with good sensing capabilities generally include piezoresistive, capacitive, piezoelectric, and triboelectric sensing [4]. Other sensing mechanisms also utilized for signal transduction include iontronic [5], magnetoresistive [6], optical [7], thermal [8], and hybrid [9] mechanisms. These mechanisms have demonstrated promise in achieving sensors with good sensing capabilities, as will be further discussed.

2.1. Piezoresistive Sensing

Piezoresistive sensing relies on a change in a material's resistance to achieve signal transduction. When mechanical stress is exerted on the material, the material deforms, and a change in its resistance is observed. The deformation is mapped to the corresponding applied force or pressure. Thus, a proportional relationship is established between the change in resistance and applied deformation [10]. This sensing mechanism is illustrated in Fig. 1. The variation in resistance could be attributed to the natural resistivity of material under variable pressure [11]. The resistance of the material is fundamentally governed by the following equation.

$$R = \frac{\rho l}{A} \tag{1}$$

where R is the material resistance, ρ is its resistivity, A is the cross-sectional area and l is the length of the material [11].

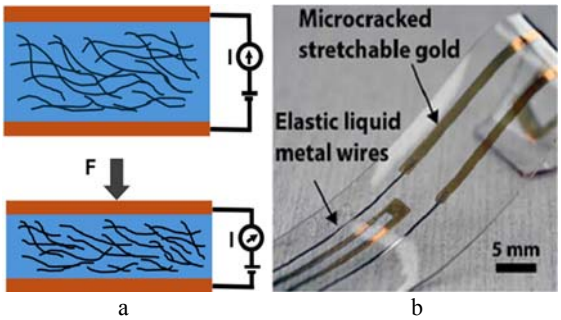


Fig. 1. a. Illustration of piezoresistive sensing. b. Magnified image of sample piezoresistive e-skin, Reproduced with permission. [16] Copyright 2015 John Wiley & Sons, Inc.

Piezoresistive sensing is comparatively a preferable sensing mechanism since changes produced by stimuli such as pressure, temperature, and strain can easily be mapped to a corresponding resistance change and detected by peripheral circuitry [12]. E-skin sensors with this sensing mechanism have advantages, including low cost of fabrication, low power consumption, good stability, and a simple signal-sensing mechanism [13].

Piezoresistive e-skin sensors made from carbon, metal, and conductive polymer materials have demonstrated excellent mechanical and electrical properties [10]. These sensors have been utilized for applications, including human health monitoring and soft robotics, to advance sensing technology. Liu et al. developed a piezoresistive textile-based pressure sensor using Carbon nanotubes (CNTs) and graphene-polymer nanocomposites, achieving a sensitivity reaching 14.4 kPa^{-1} with a low detection limit of 2 Pa [14].

Various methods have also been explored to enhance the sensing properties of piezoresistive e-skin sensors using new materials and structures. More recently, Zhang et al. developed a piezoresistive e-skin sensor using a PDMS substrate and Waterborne Polyurethane (WPU) and silver nanowire composite, achieving a remarkable sensitivity of 1.04 × 106 kPa^{-1} in the pressure range from 0 to 27kPa [15]. The performance characteristics of various piezoresistive e-skin sensors are summarized in Table 1. Performance characteristics examined for these sensors generally include determining the sensor's sensitivity, the spatial resolution of the sensor, how fast it responds to a stimulus, how easily it recovers after it deforms, as well as the life cycle of the sensor based on cyclic tests. The kinds of materials used to fabricate the sensor were also considered. These performance characteristics of e-skin sensors combine to give an idea of how close they are to the sensing capabilities of human skin.

2.2. Capacitive Sensing

This sensing mechanism relies on the capacitance change in a material under mechanical deformation. Capacitive e-skin sensors rely on the principle of capacitance, thus the ability of a system to store electric charge. They comprise a dielectric layer sandwiched between two flexible electrodes and enclosed by a substrate [20]. When pressure is applied to the sensor, it changes the material's area and reduces the distance between the electrodes, causing a variation in the capacitance of the sensor [11]. This relationship is shown in the following equation and illustrated in Fig. 2.

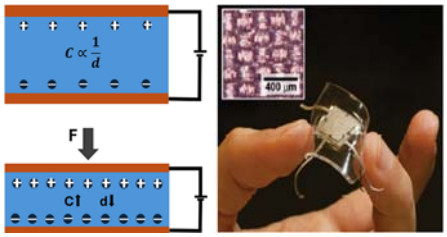
$$C = \frac{\varepsilon A}{d}, \tag{2}$$

where the capacitance has a direct relationship with area and an inverse relationship with distance. A is the area of the electrode material, whereas d is the gap distance between the electrodes, and ε is the permittivity of the dielectric material.

Table 1. Summary of results of piezoresistive sensing for e-skin applications.

Reference	Materials	Sensitivity	Spatial Resolution	Response Time	Recovery Time	Stability
[14]	Polyester/Nylon/ Ni/CNT[e]/VHB[f]	14.4 kPa^{-1} @ < 3.5 kPa	n/a	24 ms @ 2.9 kPa	30 ms @ 2.9 kPa	1000 cycles
[15]	PDMS[a]/WPU[b]/ AgNWs[c]	1.04 x 10^6 kPa^{-1} @ 0-27 kPa, 9.5 x 10^4 kPa^{-1}@ 27 -33 kPa	n/a [d]	160 ms	160 ms	> 1000 cycles
[17]	PDMS/MWCNTs[g]	15.1 kPa^{-1} @ < 0.5 kPa	1 cm × 1.5 cm sensile, 10 × 10 array	~ 0.04 s	~ 0.04 s	> 1000 cycles
[18]	PDMS/GNPs[h]/ PEO[i]/MWCNTs	6.56 MPa^{-1} @ < 66 kPa, 0.335 kPa^{-1} @ 100 kPa	n/a	171 ms	110 ms	1000 cycles
[19]	PDMS/PPy[j]/Au	1907.2 kPa^{-1} @ 0-100 Pa	5 × 5 array	50 µs	6 ms	> 15 000 cycles

a. Polydimethyl Siloxane b. Water-borne Polyurethane c. Silver Nanowires d. Not available e. Carbon Nanotubes f. Very High Bonding g. Multi-walled Carbon Nanotubes h. Graphene Nanoplatelets i. Polyethylene Oxide j. Polypyrrole



a b

Fig. 2. a) Illustration of capacitive pressure sensing; b) Sample capacitive e-skin with an optical microscope view of a conductive textile electrode. Reproduced with permission. [24] Copyright 2014 John Wiley & Sons, Inc.

Capacitive e-skin sensors often present some advantages, including high-performance pressure sensing, low hysteresis, reduced response times, good stability, and a wide detection range [20]. Leveraging these characteristics, there have been significant efforts to develop capacitive pressure sensors mirroring these features. Kang et al. developed a capacitive e-skin sensor using porous PDMS and Indium Tin Oxide (ITO)/Polyethylene Terephthalate (PET) films which achieved a sensitivity of 0.63 kPa^{-1} and a low detection limit of 2.42 Pa [21]. Li et al. developed a flexible capacitive pressure sensor using bionic hybrid microstructures, achieving sensitivity reaching 3.17 kPa^{-1} over a wide detection range [22]. Efforts have also been made to integrate these sensors into proximity detection for robotics and human-machine interfaces. Wu et al. further discuss the piezocapacitive transduction mechanism for proximity sensing and its applications in human-robot collaboration, and human-machine interfaces [23]. A summary of the performance characteristics of various capacitive e-skins sensors is shown in Table 2.

Table 2. Summary of results of capacitive sensing for e-skin applications.

Reference	Materials	Sensitivity	Spatial Resolution	Response Time	Recovery Time	Stability
[21]	PDMS/ITO[a]/ PET[b]	0.63 kPa^{-1}, 0.009 kPa^{-1} @ > 90 kPa	0.9 × 0.9 cm sensile, 15 × 15 array (42 × 42 mm)	40 ms	< 1 s	> 10 000 cycles
[22]	PDMS/Ionic gel	3.17 kPa^{-1} @ 0-90 kPa	n/a	n/a	n/a	n/a
[24]	PDMS/FVMQe/ Cu/Sn	0.91 kPa-1 @ 0.5 kPa	n/a	< 40 ms	n/a	n/a
[25]	PDMS/AgNPs[c]/ SBS[d]	0.21 kPa^{-1} @ < 2 kPa	0.16 mm^2 pixel, 6 × 6 array (2 × 1.5 cm^2)	~ 40 ms	~ 10 ms	10 000 cycles
[26]	PDMS/Ag/Ni	0.159 kPa^{-1} @ 0-1 kPa	n/a	n/a	n/a	9200 cycles

a) Indium Tin Oxide; b) Polyethylene Terephthalate; c) Silver Nanoparticles; d) Poly(styrene-butadiene-styrene); e) Flourosilicone.

2.3. Piezoelectric Sensing

Piezoelectric sensing generally utilizes the piezoelectric effect principle, where certain materials can generate electric charges when subjected to mechanical stress. When pressure or strain is exerted on the material, it creates a deformation that causes the positions of opposite charges within the material to shift, thus producing an electric charge. The electric charge is proportional to the applied mechanical deformation. Table 3 highlights the performance characteristics of various e-skin sensors with

piezoelectric sensing. The operating principle for piezoelectric sensing is illustrated in Fig. 3.

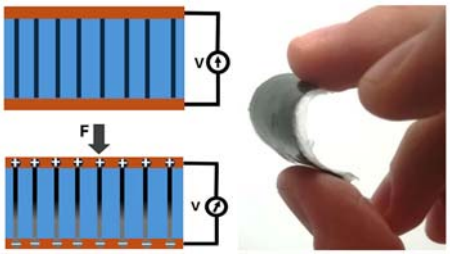
Piezoelectric sensors generally comprise piezoelectric material as the active layer, conductive layers that transmit charges, and a substrate to encapsulate the structure and provide flexibility. Sensors developed using this mechanism have shown some advantages, including self-powering capabilities, high sensitivity, and detection of dynamic pressure in real-time, making them ideal for applications in robotics and prosthetics [27, 28].

By leveraging these advantages, Guan et al. developed a piezoelectric elastomer sensor that achieved an electrical signal output of 253 mV with applied pressure and a response time of 15 ms, demonstrating its ability to detect signals quickly [27]. Sultana et al. also developed a flexible piezoelectric e-skin using an electrospun poly(l-lactic acid) PLLA nanofiber membrane, achieving a sensitivity reaching 22 VN^{-1}, stable over 37500 cycles [29].

Table 3. Summary of results of piezoelectric sensing for e-skin applications.

Reference	Materials	Sensitivity	Power Density	Open Circuit Voltage	Spatial Resolution	Stability
[27]	PVDF[a]/PAN[b]	5.6503 kPa^{-1} @ 40.59 kPa	n/a	253 mV	10 mm × 10 mm × 1 mm	10 000 cycles
[29]	PLLA[c]/Cu/Ni	22 V N^{-1}	0.07 mW cm^{-2}	1 V @ 15 Hz	n/a	375 000 cycles
[31]	PDMS/Ecoflex /PVDF	0.33 mV μ ε^{-1}	n/a	n/a	3.5 cm × 1.5 cm	n/a
[32]	$Si0_2$[d]/PVDF	5.6503 kPa^{-1} @ 150-510 kPa	39.24 mW m^{-2}	n/a	n/a	n/a

a) Poly(vinylidene fluoride); b) Polyacrylonitrile; c) Poly(l-lactic acid); d) Silicon dioxide



a b

Fig. 3. a. Illustration of piezoelectric pressure sensing. b. Sample piezoelectric e-skin sensor: Reproduced under the terms of the CC BY license. [30] Copyright 2023, The Authors, published by MDPI.

2.4. Triboelectric Sensing

Triboelectric sensing relies on the principle of triboelectric effect where a material becomes electrically charged due to contact and separation from another material. The generated charge is detected as an electrical signal and thus used to detect mechanical stimuli. The operating principle of this sensing mechanism is illustrated in Fig. 4. Electronic skin sensors developed using triboelectric sensing generally comprise a triboelectric layer combined with electrodes and flexible substrates to provide mechanical support. It also offers advantages, including its ability to generate electrical signals without an external power supply, making it suitable for energy-harvesting applications [33].

Triboelectric sensors also exhibit high sensitivity, can detect dynamic pressure, and offer excellent flexibility, making them suitable for robotics and healthcare monitoring [28]. Pu et al. developed a triboelectric e-skin sensor using a hydrogel sealed between two elastomer films of PDMS or Very High Bonding (VHB), achieving remarkable stretchability of up to 1160 % with a peak power density of 35 $mW.m^{-2}$ [34]. Arief et al. also developed a Triboelectric e-skin using fluoroelastomer, Laser-Induced Graphene (LIG), and Titanium Dioxide (TiO_2) to achieve a power density of 715 $mW.m^{-2}$ and maintained a high electrical output while subjected to 200 % elongation [37]. Table 4 provides a summary of the performance characteristics of various triboelectric e-skin sensors, including the various materials used, the sensitivity of each sensor, the power density, open-circuit voltage, spatial resolution, and stability, which are all crucial parameters.

2.5. Other Sensing Mechanisms

Several other transduction mechanisms, including optical, magnetoresistive, iontronic, and thermal, have been used for e-skin sensors. The optical sensing mechanism uses light to detect various stimuli, including pressure, temperature, and strain. These sensors offer high-precision sensing, low signal interference, and are suited for use in extreme weather conditions [38].

Based on the promising properties of optical sensing, Zhang et al. developed a highly sensitive optical e-skin sensor using glass micro/nanofibers in thin layers of PDMS, which achieved a sensitivity reaching 1870 kPa^{-1} and a low detection limit of 7 mPa [7]. Magnetoresistive sensing measures change in a material's resistance in response to a magnetic field. These sensors can detect objects' position, orientation, and magnetic field properties, thus enhancing tactile perception [6]. This is especially useful in developing magnetoresistive sensors that enable touchless interaction and real-time pressure and temperature detection, as highlighted in Bermúdez and Makarov [39]. Melzer et al. developed a magnetoresistive sensor exhibiting high sensitivity and flexibility reaching strains exceeding 270 % and stable over 1000 cycles [40].

Iontronic sensing generally involves using ions as sensing elements to detect external stimuli. Iontronic sensors detect external stimuli such as pressure, temperature, or humidity by redistributing ions in response to applied stimuli [41]. They are useful in human-machine interaction and human health monitoring. Guo et al. developed an e-skin sensor based on iontronic sensing using photocurable hydrogel to achieve a sensitivity of 171 kPa^{-1} within the pressure range from 0 to 60 kPa [5].

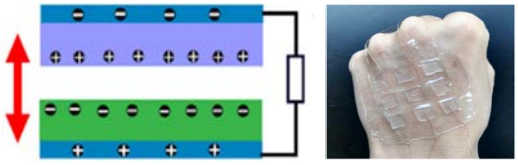


Fig. 4. a. Illustration of triboelectric pressure sensing based on vertical contact separation mode: Reproduced with permission. [11] Copyright 2021 John Wiley & Sons Inc., b. Sample triboelectric e-skin: Reproduced under the terms of the CC BY-NC license. [34] Copyright 2017, The Authors, published by American Association for the Advancement of Science (AAAS).

Table 4. Summary of results of triboelectric sensing for e-skin applications.

Reference	Materials	Sensitivity	Power Density	Open Circuit Voltage	Spatial Resolution	Stability
[34]	PDMS/VHB/ Hydrogel	0.013 kPa^{-1} @ > 70 kPa	35 mW m^{-2} @ 70 MΩ, 328 mW m^{-2} @ 7 MΩ	145 V, 182 V	1 cm × 1 cm sensile, 3 × 3 array	20 000 cycles
[35]	PDMS/P(VDF-TrFE)[a]/AgNWs	0.55 V kPa^{-1} @ 19.8 kPa	46.7 μW cm^{-2}	64 V	1 × 1.5 cm^2, 1.5 × 1.5 cm^2 sensile	10 000 cycles
[36]	PDMS/Ionogel	1.76 V N^{-1} @ 0.1-1 N	n/a	1.38 V @ 8 Hz	2 × 1.5 cm	6000 cycles
[37]	PDMS/FKM[b]/ LIG[c]/TiO_2[d]	n/a	715 mW m^{-2}	148 V	10 mm × 10 mm	3000 cycles

a) Poly(vinylidene fluoride-trifluoroethylene); b) Fluoroelastomer; c) Laser-Induced Graphene; d) Titanium Dioxide.

3. Design and Material Composition of E-skin Sensors

E-skin sensors are primarily composed of flexible multilayer structures which are crucial for the sensor's overall performance. These structures include flexible substrates, sensing layers, as well as external or integrated circuitry. This basic composition aids in detecting and converting external stimuli into electrical signals. The structural design of e-skin sensors also involves other approaches to enhance their functional properties, including sensitivity, stretchability, and many more. These approaches range from integrating novel structures like wavy and helical designs to enhancing sensing properties through surface-structured architectures.

3.1. Flexible Structures

Human skin can withstand deformation of up to 30 %, hence stretchability becomes a crucial requirement for e-skin sensors [42]. E-skins fundamentally achieve stretchability through two widely accepted approaches, thus designing flexible structures from conventional materials or using innately stretchable materials to achieve stretchability [1, 2, 42]. Flexible structures are designed based on the rule that the bending strain of materials decreases with material thickness. Hence, various material types in the formats of ribbons and wires can achieve stretchability [42].

Innately stretchable materials also provide an alternative route to stretchable electronics. These materials have the innate property of stretchability and can recover their original state after applied deformation. They comprise elastomers and liquid metals with stretchability based on long polymer chains or weak intermolecular forces [42]. The development of e-skin sensors heavily relies on carefully selecting materials that satisfy core requirements and can conform to curvilinear surfaces while retaining their functionality under deformation. The growing interest in the field of flexible electronic devices has led to the emergence of a diverse range of materials to meet these core requirements.

3.1.1. Stretchable Substrates

Stretchable substrates serve as the platform for which functional components of e-skin sensors are integrated. E-skin sensors satisfy the requirements of conforming to curvilinear surfaces largely based on the ability of substrates to bend, stretch, and adhere to such surfaces, for instance, human skin.

3.1.1.1. Polymers

Polymers such as Polyimide, Polyaniline, Polyurethane, Polyethylene Terephthalate (PET), Acrylic, Hydrogenated Nitrile Rubber (HNBR), and Parylene have all been widely used as e-skin substrates due to their flexibility and stretchability. Polymers could generally also be used as encapsulation for electronic components of e-skin sensors. Polyurethane offers a large tear strength and resistance to abrasion compared to silicon rubbers and is thus ideal for substates which will be subject to impact and scratches [42].

Acrylics can be used as passive substrates, while acrylic elastomers are softer than PDMS and offer high stretchability [42]. As shown in [43], the authors demonstrate using an acrylic elastomer as the substrate for fabricating compliant and stretchable electrodes with stretchability reaching 300 %.

3.1.1.2. Silicon-based Elastomers

These materials exhibit good mechanical adaptability and can withstand continuous deformation while recovering their elastic properties, making them great choices for e-skin applications [44]. They tend to possess high electrical conductivity, low glass transition temperature, large thermal coefficient of expansion, and high flexibility [42]. Polydimethylsiloxane (PDMS), Flourosilicone, Ecoflex, and Dragon skin are among the commonly used elastomers for e-skins. PDMS is readily accessible and can be made using available laboratory techniques; hence, it tends to be a widely used elastomer for e-skin sensors [42]. Jiang et al. demonstrated using a PDMS substrate with a thin gold layer, achieving a stretchability of up to 300 % while remaining highly conductive [45]. Chen et al. also developed a strain sensor that comprised a PDMS film, graphene/silver nanoparticle network, and thermoplastic polyurethane. It achieved a high stretchability of up to 1000 % [46]. Ecoflex is also commonly used as a substrate material for e-skins. Based on its Youngs modulus exceeding 125 kPa, it exhibits mechanical compliance like that of the human skin [47]. It is also biocompatible, highly stretchable, and easily adheres to the skin [48].

3.1.2. Sensing Layers

These are crucial components in designing e-skin sensors as they are the primary functional components that interact with external stimuli and enable their conversion into electrical signals. They form the active materials of the e-skin sensor. They depend on the sensing mechanism for transducing stimuli into electrical signals that can be measured and processed. Sensing layers for e-skins generally comprise conductive materials such as conductive polymers in an elastomer matrix (for instance, CNTs in PDMS), piezoelectric materials, dielectric materials, triboelectric materials, and ionic conductors. Each of these materials plays a significant role in the sensing capabilities of e-skin sensors. Some sensing layers, like conductive polymers and ionic conductors, serve the dual purpose of also being used as electrodes to ensure reliable signal transmission while under deformation.

3.1.2.1. Conductive Polymers

These are kinds of polymers that possess a unique property of exhibiting metallic conductivity. They are usually elastic and can recover their original state after applied deformation [49]. This, coupled with their conductive properties, makes them suitable for various applications. In the case of e-skins, these polymers are usually made by combining stretchable polymers with conductive fillers like nanoparticles, nanotubes, or nanowires [50]. Nanoparticles tend to ensure greater mobility when used in a polymer matrix compared to nanotubes and nanowires [51]. Conductive polymers can generally be categorized into 1-dimensional (1D), 2-dimensional (2D), and 3-dimensional (3D) conductive polymers based on their composition. Examples of 1D, 2D, and 3D conductive polymers are polyacetylene, polythiophene, and polypyrrole, respectively [52].

Dinh et al. reported on an electrically conductive multiwalled carbon nanotube and polymer nanocomposite using various polymerization techniques. They achieved conductive films with stretchability ranging from 400 % to 1600 % depending on the fabrication process used [53]. Wang et al. reported an elastic conductor made of polyurethane and silver nanowires which had a high conductivity of 9190 S cm-1 and stretchability reaching 310 % [54]. Kim et al. also demonstrated a stretchable conductor made of polyurethane containing spherical nanoparticles. The authors utilized two methods of fabrication, thus a Layer by Layer (LBL) deposition and Vapor-Assisted Flocculation. The LBL method yielded a polyurethane nanoparticle composite reaching a conductivity of 11000 S cm^{-1} and a maximum strain of up to 115 % after lamination. In comparison, the VAP method yielded a conductivity of 1800 S cm^{-1} and a maximum strain reaching 486 % after lamination [51].

3.1.2.2. Ionic Conductors

Ionic conductors are useful in electronic skin sensors as they facilitate sensing by converting stimuli

to electrical signals through ionic conduction. They offer important properties of biocompatibility and flexibility, making them suitable for e-skin sensor applications to detect various stimuli, including temperature and pressure. High ion conductivity and mechanical flexibility are highly sought-after properties of e-skins, thus elevating the significance of ionic conductors for e-skins. Ionic conductors serve the dual purpose of being used as stretchable electrodes, sensing layers, or performing a hybrid role of being both a sensing layer and an electrode in more advanced e-skin sensor designs. Stretchable ionic conductors, including hydrogels, ionogels, and liquid-free ion-conductive elastomers, have been utilized for e-skin applications [55].

Hydrogels, a type of ionic conductor used in developing electronic skin sensors capable of detecting various stimulus, have shown significant promise. Hao et al. demonstrated a conductive hydrogel sensor with conductivity reaching 5.6 mS cm^{-1} and remarkable stretchability of up to 4000 % [56]. Guo et al. developed an electronic sensor using a photocurable ionic hydrogel with high sensitivity reaching 171 kPa^{-1} for pressures from 0 to 60 kPa [5]. Ionic gels have also recently yielded high conductivity when used for e-skins. Niu et al. developed a micropyramid electronic skin array utilizing polyimide, copper, and gold electrode arrays with an ionic gel yielding sensitivities reaching 655.3 kPa^{-1} at pressure ranges below 0.5 kPa [57]. The authors further developed a full-skin bionic electronic skin by combining PDMS with an ionic gel resulting in a sensor with a high sensitivity reaching 8053.1 kPa^{-1} for pressure ranges below 1 kPa [58].

3.1.3. Stretchable Electrodes

Electrodes are a crucial component of e-skin sensors as they facilitate the detection of physical signals. They are used for detecting and transmitting a variety of inputs, including pressure, temperature, and strain, mainly by detecting changes in electrical properties. In the case of electronic skins, electrodes used in these sensors tend to have properties of flexibility and stretchability, thus enabling them to maintain electrical conductivity while being subject to mechanical deformation [59]. They achieve these features based on their structural design using stretchable conductive materials. Various materials have been explored for developing e-skin electrodes.

3.1.3.1. Metallic Electrodes

Metallic electrodes are usually made from conductive metals such as gold, copper, aluminum, chromium, platinum, and many others. For electronic skin sensors, electrical conductivity, biocompatibility, and cost are essential considerations for the choice of material for developing electrodes. Biocompatibility is especially important with regard to healthcare-related applications. Metals usually offer excellent electrical conductivity; hence metallic electrodes can be used to detect and measure electrical signals due to some form of applied stimuli. Thin films from metallic materials can be designed into special formats such as fractal, helical, and pre-strained bulking to achieve stretchability beyond 100 % [42]. Jiang et al. reported a thin, stretchable conducting electrode made of a 50 nm thin gold layer and a 1.2 μm PDMS film, which achieved stretchability reaching 300 % [60].

3.1.3.2. Liquid Metal-based Electrodes

Liquid Metals, based on their unique fluid-like properties and electrical conductivity, have also been explored for applications in e-skin sensors. These materials are innately deformable and can be integrated into stretchable materials such as elastomers to create conductive pathways that maintain electrical conductivity under deformation. Liquid-metal-based devices can withstand large amounts of strain up to 800 % and offer significant potential in e-skin sensors as they can sense pressure linearly over a broader range [42].

3.1.3.3. Carbon-based Electrodes

Carbon-based electrodes present unique properties of high electrical conductivity making them attractive materials in developing electrodes for e-skin sensors. They offer high conformability and enhance the functionality and performance of e-skin sensors. Examples of these include carbon particles, carbon nanotubes, carbon fibers, and graphene. Carbon nanotube fibres tend to offer significant advantages over carbon nanotube composites, especially by avoiding aggregation problems, which are typically common in randomly dispersed carbon nanotubes. They also provide high mechanical strength and low voltage operation and are cost-effective [61].

Graphene has also been explored in developing piezoresistive sensors. It is a lightweight material with excellent mechanical properties, simple fabrication methods, and electrical properties with electrical conductivity reaching 200000 $cm^2\ V^{-1}\ s^{-1}$ [13]. Its versatility allows it to be used alongside elastomers such as PDMS to exhibit reversible deformation.

3.2. Surface Architectures for E-skins

E-skins were previously designed to be planar [62], thus in a flat, two-dimensional form, making them suitable for various applications. However, the introduction of surface-structured architectures, which include intricate designs such as micro-structured and nano-structured surfaces, enables e-skins to further mimic human skin functionalities and achieve unique performances. Examples of surface-structured architectures are illustrated in Fig. 5. Research has indicated that different microstructured geometries

respond uniquely to multidirectional mechanical forces [63]. Various microstructures have been investigated, including microdomes [64], micropyramids [57], micropillars [65], microcracks [45], and microchannels [66]. These structural shapes heighten the sensitivity of e-skin sensors by increasing the mode of contact under mechanical deformation.

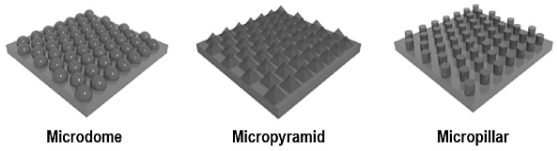


Fig. 5. Examples of surface-structured architecture for e-skins: Reproduced under the terms of the CC BY license. [63] Copyright 2018, The Authors, published by Nature.

Micropillar-shaped e-skins have been found to exhibit high sensitivity to shear stresses, while microdomes are optimal for normal, bending, and tensile stresses [63]. Microstructuring of dielectrics for capacitive e-skins has also been found to be a common strategy to significantly improve the sensor's performance, with micropyramids tending to exhibit the best sensitivity and performance for this use case [67]. Lee et al. highlights that the sensitivity of flexible pressure sensors improved by incorporating elastic pyramid-shaped microstructures [68]. Park et al. developed an ultrasensitive e-skin with an interlocked microdome structure, yielding a low detection limit of about 0.2 Pa and a rapid response and relaxation times of approximately 0.04 s [17].

To further understand the effect of the different microstructures on e-skin sensitivity, Park et al. analyzed the effect of pressure on single-sided and interlocked microstructured e-skins in comparison to planar shape [63]. For single-sided microdome, micropyramid, and micropillar structures, microdome shape yielded the highest linear sensitivity while all the other microstructured geometries had comparatively higher sensitivity than the planar-structured e-skin, confirming that different microstructured geometries influenced e-skin sensitivity [63]. These results are shown in Fig. 6.

Like microstructuring, nanostructures can also be integrated into e-skin architecture to improve overall performance. Nanostructures such as nanowires, nanoparticles, nanotubes, and nanofibers have been found to improve the functional properties of e-skins, and several studies have highlighted their significance. For instance, incorporating silicon dioxide nanoparticles in PVDF membranes has been found to achieve excellent pressure and temperature detection in e-skins, thus allowing multimodal sensing capabilities [32]. Ha et al. also demonstrated a bioinspired e-skin design of hierarchical micro and nano-structured Zinc Oxide (ZnO) nanowire arrays for sensitive detection of static and dynamic stimuli [9].

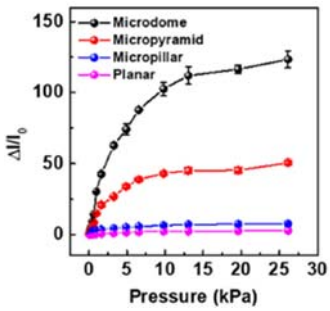


Fig. 6. Effect of single-sided geometry on e-skin pressure sensitivity: Reproduced under the terms of the CC BY license. [63] Copyright 2018, The Authors, published by Nature.

4. Applications for E-skin Sensors

By combining properties of the human skin, including flexibility, stretchability, and sensitivity, e-skin sensors have seen applications in healthcare monitoring, artificial intelligence devices, robotics, prosthetics, textiles, wearable devices, and human-machine interfaces. In the field of healthcare, the early detection of medical conditions is key to facilitating treatment. E-skin sensors have supported these efforts by monitoring vital life signals such as heart rate and body temperature, thus aiding in remote healthcare. Sun et al. explored the various efforts employed using polymer composites with functional properties to develop e-skins for health monitoring and artificial intelligence applications [69].

With the integration of machine learning, e-skin devices have advanced the field of robotics by enabling sense of touch, material perception, pressure, and temperature detection, thus improving the robot's interactions with objects, as shown in Fig. 7.

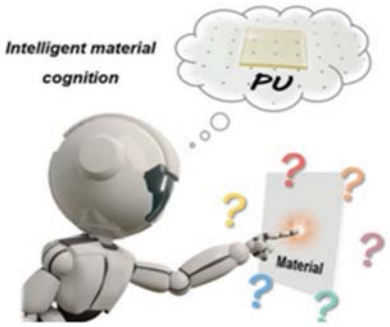


Fig. 7. Sample application of e-skin in Robotics for intelligent material cognition: Reproduced with permission. [58] Copyright 2022 John Wiley & Sons Inc.

Niu et al. developed an Artificial Intelligence (AI)-based e-skin with an advanced intelligent material recognition system to enable real-time cognition of materials through contact [58]. Dahiya [70] also emphasized the role e-skins play in manipulation or control-based tasks such as grasping objects and

estimating relevant contact parameters, including force and temperature. The breadth of applications of e-skins is continuously growing as more significant advancements are made.

5. Challenges

Despite the vast potential e-skins have in various fields, they are still plagued with several challenges, which researchers are actively working on mitigating. A key challenge affecting flexible electronic devices is the ability to remain highly sensitive while subjected to various modes of mechanical deformation, including bending and stretching [71]. The functional requirements of e-skins demand that they can conform to curvilinear surfaces while achieving high-performance sensing. E-skin sensors tend to lose sensitivity with increasing mechanical deformation, such as strain. One of the ways this could be addressed would be to develop e-skins that are highly conductive or capable of maintaining sensitivity while subjected to mechanical deformation. Following this, Lee et al. developed a bending-insensitive pressure sensor capable of detecting only normal pressure independent of mechanical stress to evaluate external stimuli on curvilinear surfaces. The authors achieved a sensor with a fast response while under complex bending conditions [68].

Additionally, e-skin sensors face the difficulty of reliable and long-lasting power sources, especially when used for continuous signal monitoring. This has necessitated the development of sensitive self-powered electronic skins to replicate high-performance sensing and continuous monitoring without external power supply [72]. Due to the complexities involved in the design and manufacturing processes of e-skins, the design of e-skin sensor arrays also requires addressing issues with flexibility and crosstalk, thus interference between adjacent sensor elements within the array. Overcoming the challenges associated with the development of e-skin sensors will help advance various fields of healthcare, robotics, wearable technology, and human-machine interface.

4. Conclusion

In this paper, we discuss e-skin sensor technology and the various efforts toward achieving unique flexible structures that can mimic the functionalities of human skin. We discuss the various sensing technologies employed in their development, their advantages, working principles, and research contributions utilizing sensing technologies to yield high-performance e-skins that mimic human skin functionalities. Carbon, Metal, and Conductive-polymer piezoresistive e-skin sensors have been found to realize excellent mechanical and electrical properties, while the use of capacitive e-skin sensors makes them suitable for proximity sensing applications. The self-powered nature of piezoelectric and triboelectric e-skins also solves a relevant challenge facing e-skin sensor advancement.

We further discuss sensor structure design and various stretchable materials used in developing e-skin sensor technologies. We emphasize the relevance of structure design in improving contact area for flexible pressure sensors, thus elevating the sensing capabilities. Research advancements have also realized e-skin sensors which show good sensitivity while being subjected to large amounts of strain of more than 300 %, highlighting their capability of conforming to curvilinear surfaces. We further discuss the potential application of e-skin sensors in healthcare, robotics, prosthetics, and human-machine interfaces. The breadth of application of e-skin sensors is continuously growing as technologies keep advancing and new materials and manufacturing methods are being developed. We finally address some significant challenges hindering e-skin sensor development. However, the limitations these sensors face are being addressed as the technology grows further. Electronic skin sensors, as discussed in this paper, hold significant promises for the future. With continued research efforts to mitigate their key challenges, they will usher in innovations in the fields of healthcare, robotics, and wearable technology.

References

[1]. Y. Sun, D. Zhou, J. Bai, E. J. Hauser, et al., Test of a tube-shaped artificial skin sensor array for colonoscopies, *IEEE Sensors Journal*, Vol. 18, Issue 6, 2018, pp. 2291-2298.

[2]. J. A. Rogers, et al., Materials and mechanics for stretchable electronics, *Science*, Vol. 327, 2010, pp. 1603-1607.

[3]. Q. Zhang, et al., Three-in-one portable electronic sensory system based on low-impedance laser-induced graphene on-skin electrode sensors for electrophysiological signal monitoring, *Advanced Materials Interfaces*, Vol. 10, Issue 3, 2023, 2201748.

[4]. A. Dos Santos, E. Fortunato, R. Martins, H. Águas, et al., Transduction mechanisms, micro-structuring techniques, and applications of electronic skin pressure sensors: a review of recent advances, *Sensors*, Vol. 20, Issue 16, 2020, 4598.

[5]. Y. Guo, F. Yin, Y. Li, G. Shen, et al., Incorporating wireless strategies to wearable devices enabled by a photocurable hydrogel for monitoring pressure information, *Advanced Materials*, Vol. 35, Issue 29, 2023, 2300183.

[6]. C. Becker, et al., A new dimension for magnetosensitive e-skins: active matrix integrated micro-origami sensor arrays, *Nature Communications*, Vol. 13, Issue 1, 2022, 757.

[7]. L. Zhang, et al., Ultrasensitive skin-like wearable optical sensors based on glass micro/nanofibers, *Opto-Electronic Advances*, Vol. 3, Issue 3, 2020, 190031.

[8]. J. Neto, R. Chirila, A. S. Dahiya, A. Christou, et al., Skin-inspired thermoreceptors-based electronic skin for biomimicking thermal pain reflexes, *Advanced Science*, Vol. 9, Issue 27, 2022, 2202155.

[9]. M. Ha, S. Lim, J. Park, D. S. Um, et al., Bioinspired interlocked and hierarchical design of ZNO nanowire arrays for static and dynamic pressure-sensitive electronic skins, *Advanced Functional Materials*, Vol. 25, Issue 19, 2015, pp. 2841-2849.

[10]. M. T. Innocent, et al., Piezoresistive fibers with large working factors for strain sensing applications, *ACS Applied Materials & Interfaces*, Vol. 14, Issue 5, 2022, pp. 7141-7151.

[11]. U. Pierre Claver, G. Zhao, Recent progress in flexible pressure sensors based electronic skin, *Advanced Engineering Materials*, Vol. 23, Issue 5, 2021, 2001398.

[12]. T. Xu, et al., High resolution skin-like sensor capable of sensing and visualizing various sensations and three dimensional shape, *Scientific Reports*, Vol. 5, 2015, 12997.

[13]. K. Y. Chen, Y. T. Xu, Y. Zhao, J. K. Li, et al., Recent progress in graphene-based wearable piezoresistive sensors: from 1D to 3D device geometries, *Nano Materials Science*, Vol. 5, Issue 3, 2023, pp. 247-264.

[14]. M. Liu, et al., Large-area all-textile pressure sensors for monitoring human motion and physiological signals, *Advanced Materials*, Vol. 29, Issue 41, 2017, 1703700.

[15]. C. Zhang, Y. Zhou, C. Ye, Silver nanowires/waterborne polyurethane composite film based piezoresistive pressure sensor for ultrasensitive human motion monitoring, *Nanotechnology*, Vol. 35, Issue 32, 2024, 325502.

[16]. A. P. Gerratt, H. O. Michaud, S. P. Lacour, Elastomeric electronic skin for prosthetic tactile sensation, *Advanced Functional Materials*, Vol. 25, Issue 15, 2015, pp. 2287-2295.

[17]. J. Park, et al., Giant tunneling piezoresistance of composite elastomers with interlocked microdome arrays for ultrasensitive and multimodal electronic skins, *ACS Nano*, Vol. 8, Issue 5, 2014, pp. 4689-4697.

[18]. Y. Luo, et al., Direct write of a flexible high-sensitivity pressure sensor with fast response for electronic skins, *Organic Electronics*, Vol. 67, 2019, pp. 10-18.

[19]. H. Li, K. Wu, Z. Xu, Z. Wang, et al., Ultrahigh-sensitivity piezoresistive pressure sensors for detection of tiny pressure, *ACS Applied Materials Interfaces*, Vol. 10, Issue 24, 2018, pp. 20826-20834.

[20]. N. Bai, et al., Graded intrafillable architecture-based iontronic pressure sensor with ultra-broad-range high sensitivity, *Nature Communications*, Vol. 11, Issue 1, 2020, 2409.

[21]. S. Kang, et al., Highly sensitive pressure sensor based on bioinspired porous structure for real-time tactile sensing, *Advanced Electronic Materials*, Vol. 2, Issue 12, 2016, 1600352.

[22]. L. Li, et al., Highly sensitive flexible capacitive pressure sensor based on bionic hybrid microstructures, in *Proceedings of the 17th IEEE International Conference on Nano/Micro Engineered and Molecular Systems (NEMS'22)*, 2022, pp. 291-295.

[23]. B. Wu, et al., Proximity sensing electronic skin: principles, characteristics, and applications, *Advanced Science*, Vol. 11, Issue 13, 2024, 2308749.

[24]. L. Viry, et al., Flexible three-axial force sensor for soft and highly sensitive artificial touch, *Advanced Materials*, Vol. 26, Issue 17, 2014, pp. 2659-2664.

[25]. J. Lee, et al., Conductive fiber-based ultrasensitive textile pressure sensor for wearable electronics, *Advanced Materials*, Vol. 27, Issue 15, 2015, pp. 2433-2439.

[26]. W. Asghar, et al., Piezocapacitive flexible e-skin pressure sensors having magnetically grown microstructures, *Advanced Materials Technologies*, Vol. 5, Issue 2, 2020, 1900892.

[27]. Y. Guan, et al., Soft, super-elastic, all-polymer piezoelectric elastomer for artificial electronic skin, *ACS Applied Materials Interfaces*, Vol. 15, Issue 1, 2023, pp. 1736-1747.

[28]. S. Cho, et al., Self-powered hybrid triboelectric–piezoelectric electronic skin based on P(VDF-TrFE) electrospun nanofibers for artificial sensory system, *Functional Composites and Structures*, Vol. 4, Issue 4, 2022, 045005.

[29]. A. Sultana, M. M. Alam, P. R. Ghosh, M. R. Middya, D. Mandal, Human skin interactive self-powered wearable piezoelectric bio-e-skin by electrospun poly-l-lactic acid nanofibers for non-invasive physiological signal monitoring, *Journal of Materials Chemistry B*, Vol. 5, Issue 35, 2017, pp. 7352-7359.

[30]. G. Selleri, et al., Self-sensing soft skin based on piezoelectric nanofibers, *Polymers*, Vol. 15, Issue 2, 2023, 340.

[31]. X. Yang, et al., Highly sensitive piezoelectric e-skin design based on electromechanical coupling concept, *Advanced Electronic Materials*, Vol. 9, Issue 5, 2023, 2201313.

[32]. E. Kar, P. Ghosh, S. Pratihar, M. Tavakoli, et al., SiO2 nanoparticles incorporated poly(Vinylidene) fluoride composite for efficient piezoelectric energy harvesting and dual-mode sensing, *Energy Technology*, Vol. 11, Issue 2, 2023, 2201130.

[33]. Z. Wang, et al., STEV: Stretchable triboelectric e-skin enabled proprioceptive vibration sensing for soft robot, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'23)*, 2023, pp. 588-593.

[34]. X. Pu, et al., Ultrastretchable, transparent triboelectric nanogenerator as electronic skin for biomechanical energy harvesting and tactile sensing, *Science Advances*, Vol. 3, Issue 5, 2017, e1700015.

[35]. M. Ha, et al., Skin-inspired hierarchical polymer architectures with gradient stiffness for spacer-free, ultrathin, and highly sensitive triboelectric sensors, *ACS Nano*, Vol. 12, Issue 4, 2018, pp. 3964-3974.

[36]. G. Zhao, et al., Transparent and stretchable triboelectric nanogenerator for self-powered tactile sensing, *Nano Energy*, Vol. 59, 2019, pp. 302-310.

[37]. I. Arief, J. H. Lee, T. D. Nguyen, J. C. Yeo, et al., Elastomeric microwell-based triboelectric nanogenerators by in situ simultaneous transfer-printing, *Materials Horizons*, Vol. 9, Issue 5, 2022, pp. 1468-1478.

[38]. C. Pendao, I. Silva, Optical fiber sensors and sensing networks: overview of the main principles and applications, *Sensors*, Vol. 22, Issue 19, 2022, 7554.

[39]. G. S. Cañón Bermúdez, D. Makarov, Magnetosensitive e-skins for interactive devices, *Advanced Functional Materials*, Vol. 31, Issue 39, 2021, 2105476.

[40]. M. Melzer, et al., Imperceptible magnetoelectronics, *Nature Communications*, Vol. 6, Issue 1, 2015, 6013.

[41]. Y. Xiong, J. Han, Y. Wang, Z. L. Wang, et al., Emerging iontronic sensing: materials, mechanisms, and applications, *Research*, Vol. 2022, 2022, 9785189.

[42]. Y. Zhao, X. Huang, Mechanisms and materials of flexible and stretchable skin sensors, *Micromachines*, Vol. 8, Issue 3, 2017, 69.

[43]. T. A. Kim, H. S. Kim, S. S. Lee, M. Park, Single-walled carbon nanotube/silicone rubber composites for compliant electrodes, *Carbon*, Vol. 50, Issue 2, 2012, pp. 444-449.
[44]. A. Panbude, P. Veluswamy, Silicone elastomer: encapsulating materials for flexible thermoelectric generator, *IEEE Sensors Journal*, Vol. 23, Issue 15, 2023, pp. 16608-16615.
[45]. Z. Jiang, et al., A 1.3-micrometre-thick elastic conductor for seamless on-skin and implantable sensors, *Nature Electronics*, Vol. 5, Issue 11, 2022, pp. 784-793.
[46]. S. Chen, Y. Wei, X. Yuan, Y. Lin, et al., A highly stretchable strain sensor based on a graphene/silver nanoparticle synergic conductive network and a sandwich structure, *Journal of Materials Chemistry C*, Vol. 4, Issue 19, 2016, pp. 4304-4311.
[47]. M. Amjadi, Y. J. Yoon, I. Park, Ultra-stretchable and skin-mountable strain sensors using carbon nanotubes-Ecoflex nanocomposites, *Nanotechnology*, Vol. 26, Issue 37, 2015, 375501.
[48]. T. Nishikawa, H. Yamane, N. Matsuhisa, N. Miki, Stretchable strain sensor with small but sufficient adhesion to skin, *Sensors*, Vol. 23, Issue 3, 2023, 1774.
[49]. L. Benny Mattam, A. Bijoy, D. Abraham Thadathil, L. George, et al., Conducting polymers: a versatile material for biomedical applications, *ChemistrySelect*, Vol. 7, Issue 37, 2022, e202202894.
[50]. M. Takakuwa, et al., Direct gold bonding for flexible integrated electronics, *Science Advances*, Vol. 7, Issue 52, 2021, eabk0383.
[51]. Y. Kim, et al., Stretchable nanoparticle conductors with self-organized conductive pathways, *Nature*, Vol. 500, 2013, pp. 59-63.
[52]. Y. Wang, W. Feng, Introduction of conductive polymers, in Conductive Polymers and Their Composites, *Springer Nature Singapore*, 2022, pp. 1-31.
[53]. L. N. M. Dinh, B. N. Tran, V. Agarwal, P. B. Zetterlund, Synthesis of highly stretchable and electrically conductive multiwalled carbon nanotube/polymer nanocomposite films, *ACS Applied Polymer Materials*, Vol. 4, Issue 3, 2022, pp. 1867-1877.
[54]. Y. Wang, T. Yokota, T. Somenya, Electrospun nanofiber-based soft electronics, *NPG Asia Materials*, Vol. 13, Issue 1, 2021, pp. 1-22.
[55]. W. Niu, X. Liu, Stretchable ionic conductors for soft electronics, *Macromolecular Rapid Communications*, Vol. 43, Issue 23, 2022, 2200460.
[56]. S. Hao, C. Shao, L. Meng, C. Cui, et al., Tannic acid-silver dual catalysis induced rapid polymerization of conductive hydrogel sensors with excellent stretchability, self-adhesion, and strain-sensitivity properties, *ACS Applied Materials Interfaces*, Vol. 12, Issue 50, 2020, pp. 56509-56521.
[57]. H. Niu, et al., Micropyramid array bimodal electronic skin for intelligent material and surface shape perception based on capacitive sensing, *Advanced Science*, Vol. 11, Issue 3, 2024, 2306267.
[58]. H. Niu, et al., Perception-to-cognition tactile sensing based on artificial-intelligence-motivated human full-skin bionic electronic skin, *Advanced Materials*, Vol. 34, Issue 31, 2022, 2202622.
[59]. H. Deng, J. Li, J. Wang, L. Zhou, Stretchable electrodes for wearable electronics, *SHS Web of Conferences*, Vol. 157, 2023, 04014.
[60]. Z. Jiang, et al., Highly stretchable metallic nanowire networks reinforced by the underlying randomly distributed elastic polymer nanofibers via interfacial adhesion improvement, *Advanced Materials*, Vol. 31, Issue 37, 2019, 1902953.
[61]. S. Lv, et al., Flexible, stretchable, wearable electronics based on aligned carbon nanotube fiber arrays for motion detection and human-machine interaction, *Sensors and Actuators: A. Physical*, Vol. 362, 2023, 117712.
[62]. X. Lin, et al., Biocompatible multifunctional e-skins with excellent self-healing ability enabled by clean and scalable fabrication, *Nano-Micro Letters*, Vol. 13, 2021, 200.
[63]. J. Park, et al., Tailoring force sensitivity and selectivity by microstructure engineering of multidirectional electronic skins, *NPG Asia Materials*, Vol. 10, Issue 4, 2018, pp. 163-176.
[64]. J. Park, M. Kim, Y. Lee, H. S. Lee, et al., Nanomaterials: fingertip skin-inspired microstructured ferroelectric skins discriminate static/dynamic pressure and temperature stimuli, *Science Advances*, Vol. 1, Issue 9, 2015, e1500661.
[65]. S. Zhang, et al., A facile and novel design of multifunctional electronic skin based on polydimethylsiloxane with micropillars for signal monitoring, *Journal of Materials Chemistry B*, Vol. 8, Issue 36, 2020, pp. 8315-8322.
[66]. D. M. Vogt, Y. L. Park, R. J. Wood, Design and characterization of a soft multi-axis force sensor using embedded microfluidic channels, *IEEE Sensors Journal*, Vol. 13, Issue 10, 2013, pp. 4056-4064.
[67]. W. Deng, et al., Microstructure-based interfacial tuning mechanism of capacitive pressure sensors for electronic skin, *Journal of Sensors*, Vol. 2016, 2016, 4049354.
[68]. S. Lee, et al., A transparent bending-insensitive pressure sensor, *Nature Nanotechnology*, Vol. 11, Issue 5, 2016, pp. 472-478.
[69]. Q. J. Sun, Q. T. Lai, Z. Tang, X. G. Tang, et al., Advanced functional composite materials toward e-skin for health monitoring and artificial intelligence, *Advanced Materials Technologies*, Vol. 8, Issue 5, 2023, 2201088.
[70]. R. Dahiya, E-skin: from humanoids to humans, *Proceedings of the IEEE*, Vol. 107, Issue 2, 2019, pp. 247-252.
[71]. L. Fliegans, J. Troughton, V. Divay, S. Blayac, et al., Design, fabrication and characterisation of multi-parameter optical sensors dedicated to e-skin applications, *Sensors*, Vol. 23, Issue 1, 2023, 233.
[72]. J. Liang, et al., Transparent electronic skin from the integration of strain sensors and supercapacitors, *Advanced Materials Technologies*, Vol. 8, Issue 4, 2023, 2201313.

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 12-18

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Identification and Online-correction of Periodic Motor Torque Fluctuations on Electromechanical Axes

[1, *] Chris SCHOEBERLEIN, [1] Armin SCHLEINITZ, [1] Holger SCHLEGEL and [1, 2] Martin DIX

[1] Professorship Production Systems and Processes (PSP), Chemnitz University of Technology, Reichenhainer Str. 70, 09126 Chemnitz, Germany
[2] Fraunhofer Institute for Machine Tools and Forming Technology IWU, 09126 Chemnitz, Germany
[1] Tel.: +49 371 30505
[*] E-mail: chris.schoeberlein@mb.tu-chemnitz.de

Received: 12 Mar. 2025 /Revised: 20 April 2025 /Accepted:25 April 2025 /Published: 30 April 2025

Abstract: Electromechanical axes generate high-precision motion profiles and positioning sequences in a wide range of industrial applications. The utilization of the drive internal signals for monitoring purposes, e.g. for estimating the axis load during a production process, is gaining more and more attention. One major drawback is the superimposition of the load-proportional motor torque by further operation-related disturbance effects. The paper presents a novel approach for the identification and correction of periodic disturbance fluctuations in the motor torque signal. Therefore, a novel method for automatic offline determination of the relevant spectral components using the least squares method is applied. The identified magnitude and phase values are stored in speed-related lookup tables. During normal operation, the sum of the individual sinoids is subtracted from the measured motor torque. The paper provides the overall description of the methodology and its functional verification on a single-axis drive test rig equipped with an industrial control.

Keywords: Electromechanical axis, Torque fluctuations, Monitoring, Offline identification, Online correction.

1. Introduction

Electromechanical axes are applied in a wide range of production systems to generate the required motion profiles and positioning sequences. Regardless of the specific application, the internal processed drive signals can also be used for higher-level load or process monitoring. For this purpose, different types of disturbance observers have been invented over the last two decades [1]. What almost all approaches have in common is that the estimation of external loads is based on an evaluation of the measured motor torque signal. Considering the torque equilibrium according to Eq. (1), it is noticeable that the motor torque signal (τ_m) is superimposed by further disturbance components. In addition to external load toques (τ_{le}), this includes friction (τ_f), gravitational (τ_g) and acceleration torques (τ_a) as well as periodic disturbance torques (τ_{ld}) [2].

$$\tau_m = \tau_a + \tau_f + \tau_g + \tau_{ld} + \tau_{le} \quad (1)$$

Basically, the causes of periodic torque fluctuations on electromechanical axes can be divided into three categories (see Fig. 1):

- Motor-related fluctuations;
- Encoder-related fluctuations;
- Fluctuations due to mechanical disturbances.

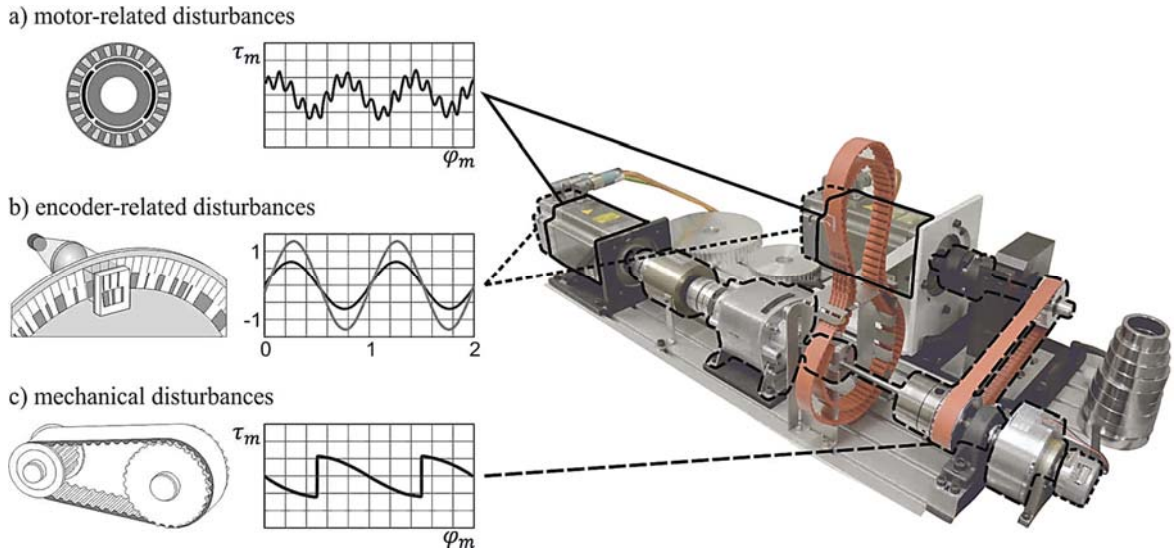


Fig. 1. Drive test rig and schematic representation of causes of periodic torque fluctuations [3].

Motor-related torque fluctuations occur as result of the electromechanical energy conversion in the servomotor. The torque fluctuations are periodic over one motor revolution, whereby their frequencies are multiples of the number of pole pairs and stator slots. For detailed research of the underlying effects and the resulting frequency components in the motor current signal, see [4]. The associated spectral components are primarily determined in the disassembled state of the motor, stored in form of correction tables and subtracted from the measured motor torque depending on the motor position in conventional operation (cf. [5, 6]). If the correction takes place directly in the drive control unit, this may lead to increased calculation time and further functional restrictions of the drive modules [7].

On the other hand, incremental motor encoders are subject to systematic and stochastic errors [8]. These lead to deviations of the measured position and speed values from the specified setpoint values and to an adjustment of the respective control variables. While stochastic errors (e.g. signal and quantization noise) are not deterministic and can only be reduced by additional hardware (e.g. shielding, oversampling, low-noise amplifiers) or suitable signal filtering methods, systematic encoder errors are generally reproducible [8]. Approaches for correcting systematic errors of incremental measurement systems and resolvers based on polynomials, correction tables and special controller structures are described in various publications (e.g. [8-12]). What almost all methods have in common is that the individual encoder track signals (i.e. separated into sine and cosine signals) must be available in high temporal resolution. Due to the modular design of modern control and drive systems, access to these signals at the required clock rate is usually not possible without further restrictions. However, depending on the drive control manufacturer, monitoring and correction functions for the raw encoder signals are occasionally implemented directly in the highly sampled encoder modules [13]. In addition, the underlying methods usually require experiments with decoupled mechanics or additional flywheel masses.

The third cause of periodic motor torque fluctuations is the mechanical system. A distinction can be made between operational and damage-related phenomena. Typical examples of mechanically induced torque fluctuations during normal operation are imbalances due to assembly errors or the running-in shock in belt drives due to polygon effect [14]. Torque fluctuations as a consequence of damaged mechanical components are not considered in this paper. This issue is addressed under the subject of condition monitoring (cf. [15-17]) including numerous publications in the area of electromechanical axes (e.g. [18, 19]).

What all currently available identification and correction approaches have in common is that they are tailored to exactly one of these categories (cf. [6, 10]). The paper proposes a novel methodology for an automatic identification of periodic disturbance torques independent of their specific cause. In contrast to existing approaches, this leads to a generally applicable correction methodology that is not limited to specific scenarios. First of all, the approach is described in more detail in the following section. Section 3 provides a functional verification on a single-axis test rig. The article concludes with a summary and an outlook.

2. Methodology

The developed methodology is based on the premise that the superposition of all periodic disturbance components can be described as a function of the angular position φ_m (Eq. (2)).

$$\tau_{ld,\Sigma} = \sum_{k=1}^{n} A_k \cdot \sin\left(f_{pos,k} \cdot 2\pi \cdot \varphi_m + \varphi_k\right) \qquad (2)$$

While the spatial frequencies $f_{pos,k}$ are independent of the current motion state, the individual amplitudes A_k and phases φ_k vary depending on the motor speed. This is due to the parameterization and structure of the drive control as well as the mechanical transfer behaviour. In addition, aliasing frequencies $f_{i,a}$ arise due to the limited sampling time of the drive controller f_s and must be reconstructed from the original frequencies f_i by using Eq. (3).

$$f_{i,a} = |i \cdot f_s - f_i|, with\ i = \left\lfloor \frac{\frac{f_s}{2} + f_i}{f_s} \right\rceil \qquad (3)$$

As depicted in Fig. 2, the overall approach is divided in an offline identification routine and an online signal correction. During the identification phase, k speed operation points are approached three times and in both directions with a fixed step width while recording the current motor position and torque values. For the presented test rig, a maximum position range between -3600° and 3600° with a rotational speed increment of 10 min^{-1} up to 500 min^{-1} has proven to be suitable. An additional buffer zone of 1080° each was used to accelerate to the required speed level or decelerate to standstill. In case of linear axes, full utilization of the available travel range is recommended. After averaging the three runs, the signal amplitudes in the spatial frequency range are evaluated to select the most significant oscillation parts. Their number can be freely parameterized, whereby a compromise between the accuracy of the correction and the required computation resources should be aimed for. To reduce the calculation effort, it is recommended to consider only spectral components with an amplitude greater than one per thousand of the maximum motor torque.

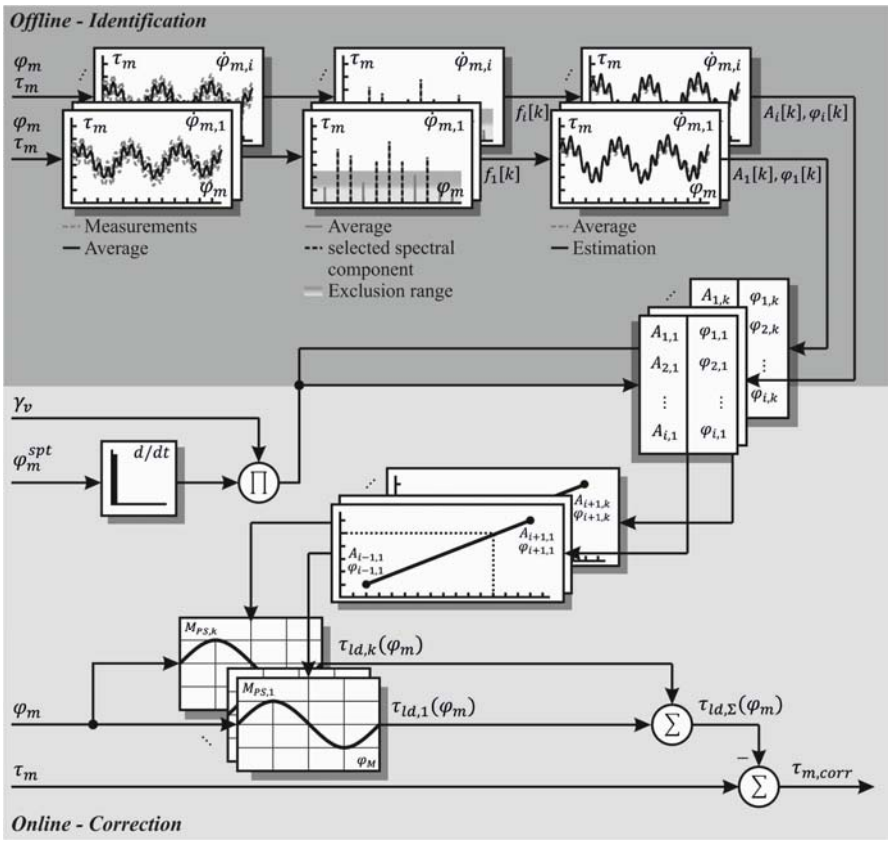


Fig. 2. Approach for the identification and correction of periodic disturbance torques.

Subsequently, the associated amplitudes ($A_i[k]$) and phases ($\varphi_i[k]$) of the k sinusiodal oscillations are estimated for each spatial frequency $f_{pos,k}$ according to Eq. (2) using a least squares algorithm. Eventually, the calculated values are stored in speed and direction dependent lookup tables. During operation, the entries of the lookup tables are selected in each sample step based on the first derivative of the setpoint position φ_m^{spt} and a standstill index γ_v. Using the setpoint value of the motor position prevents major deviations due to measurement noise. The standstill index γ_v is based on the evaluation of the motor speed and is calculated to zero in standstill phases and equal to one during axis motion. Linear interpolation takes place between the individual operating points. The corrected motor torque $\tau_{m,corr}$ results from the sum of the calculated sinoids $\tau_{ld,\Sigma}(\varphi_m)$ in each sampling step. The correction is deactivated if the speed setpoint exceeds the maximum value of the lookup table.

3. Experimental Results

3.1. Offline Identification

The described methodology is validated on the uniaxial rotational drive test rig depicted in Fig. 1. A detailed description of the test axis is provided in [2]. First of all, the offline identification of the sinoids takes place. Therefore, the defined speed operating points between 10 and 500 min^{-1} are approached three times at intervals of 10 min^{-1} while the position and torque values are recorded. Subsequently, the three measurements are averaged and transferred to the spatial frequency domain. Fig. 3 shows the results for both directions of movement for all speed operating points. For the sake of clarity, the spatial frequency axis was limited to 144 periods per mechanical revolution. Although the spectra basically show qualitative similarities, smaller differences occur depending on the direction of rotation. In particular, there are increased torque amplitudes in the upper speed band for negative direction of motion.

Furthermore, the amplitude curve of the oscillation with a spatial frequency of 18 periods per 360° shows significant differences depending on the direction of movement. This is due to deviations in the angular position of the motors on the drive and load side after closing the electrical coupling. Even small differences lead to an offset of the motor poles and thus to a direction-dependent torque amplitude. The singular torque peaks that occur occasionally (e.g. at n_m = 250 min^{-1} and 128 periods per 360°) are caused by small position deviations of the encoder, which are reflected in lower spatial frequency ranges.

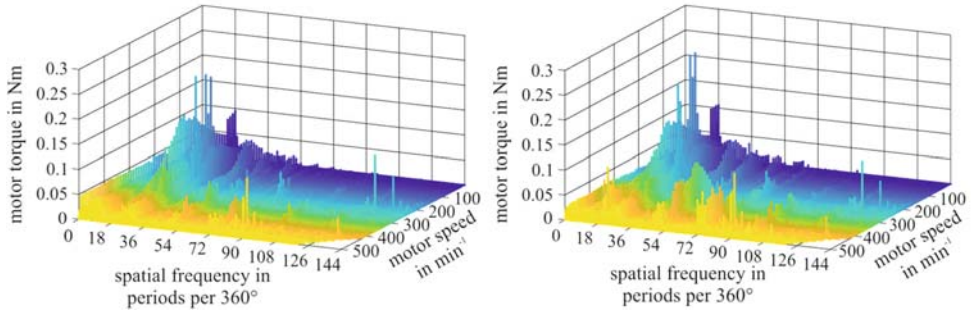


Fig. 3. Spatial frequency spectra of the averaged motor torques for all stored speed operating points with positive (left) and negative (right) direction of rotation

Table 1 shows the identified oscillation components sorted by spatial frequency and their assignment to the allocated drive component as described in Fig 1. Despite the oscillations with f_i = 1 period per 360°, f_i = 9 periods per 360° and f_i = 36 periods per 360°, the identified spatial frequencies are independent of the direction. Although these oscillation components also occur in the opposite direction of rotation, their amplitude does not reach the specified threshold value for all speed operating points considered.

Table 1. Identified spatial frequencies and the allocated drive part (a – servomotor, b – encoder, c – mechanics).

Direction	**Positive**								
f_i in periods per 360°	1	9	18	19	24	26	27	38	48
allocation	c	a	a	a	c	c	a	a, c	a, c
Direction	**Negative**								
f_i in periods per 360°	18	19	24	26	27	36	38	45	48

3.2. Online Correction

In the following, the performance of the proposed approach will be illustrated using motion sequences with constant motor speed. Fig. 4 illustrates the result for two exemplary speed operating points with n_m = 10 min^{-1} (left) and n_m = 250 min^{-1} (right). The diagrams in the upper section show the measured motor torques τ_m (gray), the sum of all periodic correction torques $\tau_{ld,\Sigma}$ (dashed blue) and the torque difference $\tau_{m,corr}$ (orange) over the motor angle φ_m for an angle range from -180° to 180°. The corresponding spatial frequency spectra are shown below. In case of n_m = 10 min^{-1}, an additional zoomed spectrum limited to 48 periods per 360° illustrates the lower frequency range. Constant torque offset due to friction was eliminated beforehand.

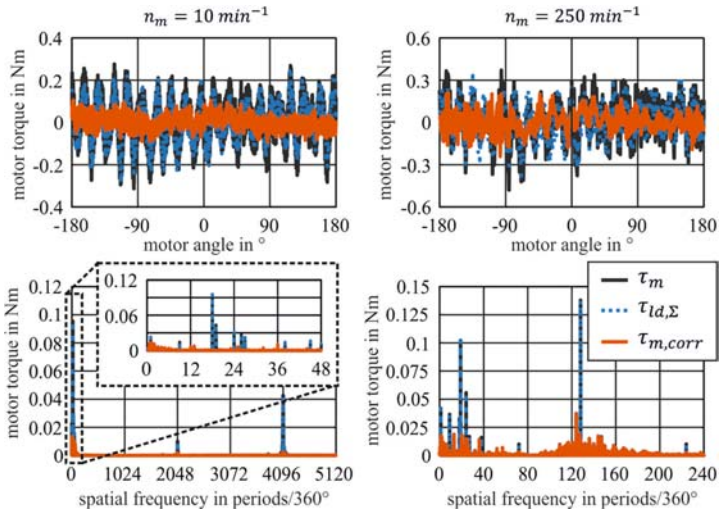


Fig. 4. Results of the correction of position-related disturbance torques for speed operating points n_m = 10 min^{-1} and n_m = 250 min^{-1} with positive direction of rotation.

The spectrum in the lower left part of the figure shows that the spatial frequencies of the disturbance torques resulting from the errors of the encoder do not lead to a violation of Shannon's sampling theorem [20]. For the second speed operating point (n_m = 250 min^{-1}), the fundamental frequency of the encoder and its first harmonic exceed half the sampling frequency of the signal acquisition, which leads to a reflection in a lower spatial frequency range ($f_{pos} \approx$ 128 periods per 360°). Nevertheless, the proposed approach is capable of correctly reproducing the associated oscillations. Note that the value of the aliasing frequency is not constant due to noise in the measured speed signal. This effect was deliberately suppressed for the correction approach due to the inverse frequency characteristic.

Regardless of the specified motor speed, the application of the correction torque leads to a significant reduction in torque ripple. This is confirmed when looking at the integral of absolute error (A_{IAE}) and maximum deviations (e_{max}) for the measured and corrected motor torques across all identified speed operating points (Fig. 5). Compared to an ideal correction ($\tau_m(\varphi_m)$ = 0 Nm), the values of A_{IAE} are reduced by an average of 50 % regardless of the direction. The maximum deviations are reduced by an average of approximately 38 %. Furthermore, it is noticeable that the effectiveness of the correction method decreases with increasing speed.

This is confirmed by the measured spatial frequency spectra (Fig. 3). The modeled spectral parts contribute most to torque ripple in the lower and medium speed range. For higher speeds, oscillations with a comparatively low spatial frequency are more prominent, but do not exceed the specified threshold value of one per thousand of the maximum torque. Hence, depending on the motor speed, oscillations with a comparatively low torque amplitude are corrected without having a significant influence on the quality of the correction. This could be countered by a speed-related selection of the oscillations or reduced threshold value for selecting the frequency components. However, this may also lead to increased calculation effort.

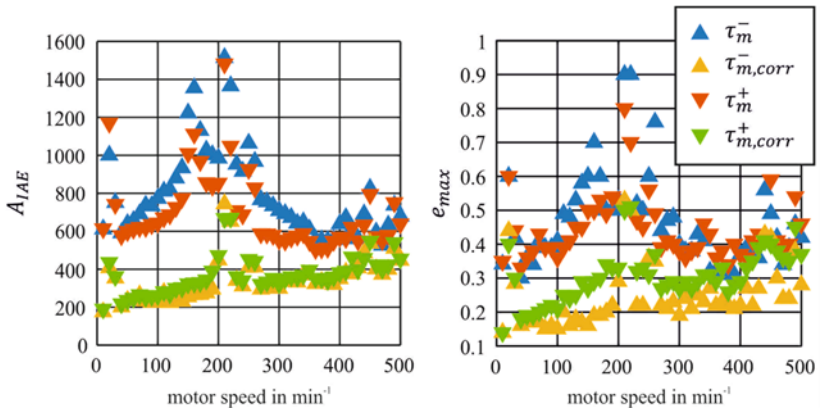


Fig. 5. Integral of absolute error (left) a maximum amplitude deviations (right) of measured and corrected motor torque compared to an ideal correction for all examined speed operating points and both directions of rotation.

So far, only speed operating points have been analyzed, which are stored in the lookup tables. Therefore, no statements on the linear interpolation of the proposed approach are possible. For this reason, further speed operating points were examined which are located between the lower and upper limits of the lookup tables. The calculation of the nominal values listed in Table 2 is based on the Renard R6 series for preferred numbers, descending from the maximum speed 500 min^{-1} [21].

The position range for the movement sequences is limited from φ_m = -3600° to φ_m = 3600°. An additional averaging of three measurements for each speed level and direction reduces position fluctuations on load side. These arise as a result of small deviations when closing the electrical clutch.

Table 2. Speed operating points for validating the online correction of periodic torque fluctuations.

No.	1	2	3	4	5	6	7	8	9	10	11	12
n_m in min^{-1}	500	341	232	158	108	74	50	34	23	16	11	8

For the sake of clarity, Fig. 6 shows the absolute integral of error and the maximum deviations compared to an ideal correction. For better visualization, the x-axis is scaled in logarithmic division. Basically, it is noticeable that the quality of the correction is slightly worse compared to the measurements displayed in Fig. 5. The integral of absolute error compared to an ideal correction is reduced by an average of approximately 40 % and the maximum deviations by an average of approximately 35 %. On the one hand, one reason is the linear interpolation of amplitude and phase values, which only represents an approximation of the actual values between the stored speed operating points.

In particular, the amplitudes of the encoder-related spectral components are set too low for the speed specifications n_m = 8 min^{-1} and n_m = 11 min^{-1}. This could be countered by a finer resolution of the speed operating points in this range. In the case of n_m = 23 min^{-1}, the correction even leads to a slight increase in the resulting deviations, caused by an out-of-phase connection of the basic encoder oscillation. This is caused by the large frequency difference of approximately 100 Hz between the two neighboring entries in the lookup table, which results from the calculation of the frequencies in the drive cycle. As a result, relevant amplitude and phase information may be lost during the subsequent linear interpolation.

On the other hand, deviations arise due to the structural design of the test rig. For example, the electric clutch is always open in the de-energized state and must be actively closed after startup. Although the servomotors on drive and load side are referenced beforehand, there are non-reproducible deviations between the measured position values of the two motors after the electrical clutch is closed and the controller enable is withdrawn on the load side.

As a result, differences in amplitude and phase values already arise during the offline identification phase, particularly for the encoder-related torque fluctuations. However, this effect only occurs on the test rig or electromechanical axes with shiftable clutches, respectively. Nonetheless, the results in Fig. 6 show that the proposed approach is also capable to significantly reduce torque ripple for not stored speed levels.

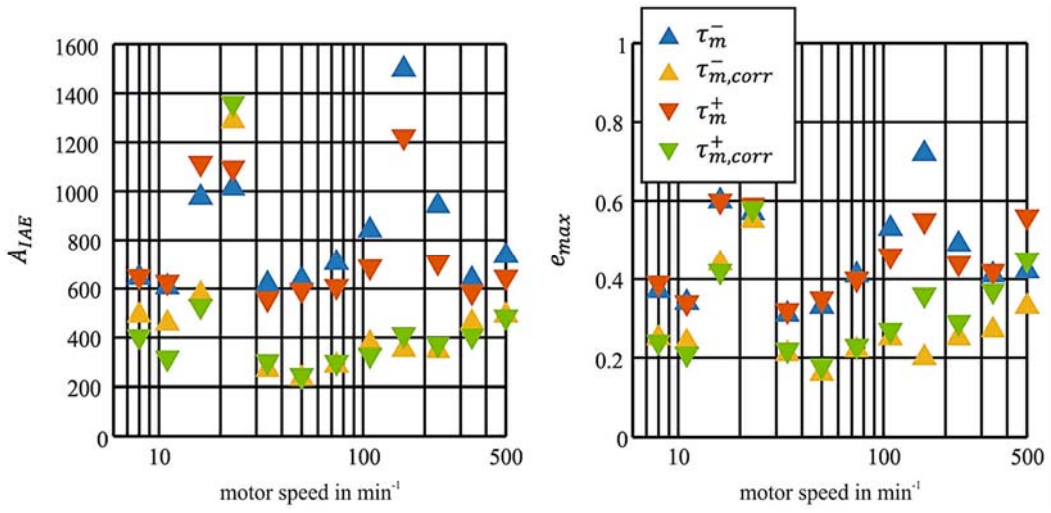


Fig. 6. Integral of absolute error (left) and maximum deviations (right) of measured and corrected motor torque compared to an ideal correction for the speed operating points listed in Table 2 and both directions of rotation.

8. Conclusions

The paper postulates a novel approach for correcting periodic fluctuations in the motor torque signal of electromechanical axes. The proposed solution incorporates an offline identification of the necessary oscillation frequencies including magnitude and phase combined with a subsequent online correction using speed-related lookup tables. In contrast to established methods, the approach is not restricted to specific disturbance sources. The experimental results carried out on a single-axle drive test rig demonstrate the performance of the proposed method. Even in case of operating points between the stored values of the lookup tables, a significant reduction in torque ripple is achieved. Future work should investigate experimental tests on different axis configurations and the performance with continuous speed changes (e.g. in case of circular movements). Furthermore, a limitation of the size of the lookup tables depending on defined motor speed ranges can reduce the required storage and processing resources of the control for clock-synchronous calculation.

Acknowledgements

Supported by:

Federal Ministry for Economic Affairs and Climate Action

ZIM Zentrales Innovationsprogramm Mittelstand

References

[1]. C. Schöberlein, et al., Simulation and disturbance estimation of speed-controlled mechatronic drive systems, *MATEC Web of Conferences*, Vol. 306, 2020, 04001.

[2]. C. Schöberlein, et al., Experimental investigation and comparison of approaches for correcting acceleration phases in motor torque signal of electromechanical axes, in *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics (ICINCO'23)*, 2023, pp. 140-151.

[3]. C. Schöberlein, et al., Identification and correction of periodic fluctuations in the motor torque signal of electromechanical axes, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 107-109.

[4]. F. D. Scholl, Reduzierung der drehmomentwelligkeit von permanenterregten servoantriebssystemen, PhD Thesis, *Universität Stuttgart*, 1995 (in German).

[5]. A. Albrecht, Wärmedehnungskompensierte rekonstruktion von prozesskräften an vorschubantrieben, PhD Thesis, *Technische Universität Carolo-Wilhelmina Braunschweig*, 2009 (in German).

[6]. B. Denkena, et al., Reconstruction of process forces in a five-axis milling center with a LSTM neural network in comparison to a model-based approach, *Journal of Manufacturing and Materials Processing*, Vol. 4, Issue 3, 2020, 62.

[7]. SINAMICS S120 – Drive functions, Function Manual, *Siemens AG*, 2020.

[8]. B. Höscheler, Entwicklung eines prüfstands zur echtzeitsimulation mechanischer systeme unter verwendung einer hochauflösenden lageerfassung, PhD Thesis, *Universität Erlangen-Nürnberg*, 2001 (in German).

[9]. U. Probst, Servoantriebe in der Automatisierungstechnik: Komponenten, Aufbau und Regelverfahren, *Vieweg + Teubner*, 2011.

[10]. A. Bünte, H. Grotstollen, P. Krafka, High-performance speed measurement by suppression of systematic resolver and encoder errors, *IEEE Transactions on Industrial Electronics*, Vol. 51, Issue 1, 2004, pp. 210-217.

[11]. Y. Daaboul, Fehlermodelle und fehlerkorrektur für inkrementalgeber bei servoantrieben, PhD Thesis, *Technische Universität Carolo-Wilhelmina Braunschweig*, 2017 (in German).

[12]. J. Wittmann, Beschleunigungsregelung von servoantrieben basierend auf positionsmesswerten, PhD Thesis, *Technische Universität München*, 2018 (in German).

[13]. Siemens AG, SINAMICS S120/S150 parameter list, Function Manual, Siemens AG, 2014.

[14]. R. Perneder, Handbuch Zahnriementechnik: Grundlagen, Berechnung, Anwendungen, *Springer*, 2009.

[15]. J. Kolerus, Zustandsüberwachung von Maschinen, 7th Ed., *expert Verlag*, 1995 (in German).

[16]. A. Davies (Ed.), Handbook of Condition Monitoring: Techniques and Methodology, 1st Ed., *Chapman & Hall*, 1998.

[17]. R. Isermann, Fault-Diagnosis Applications: Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-tolerant Systems, *Springer*, 2011.

[18]. U. Frieß, Zustandsüberwachung von werkzeugmaschinen durch adaptive kennwertinterpretation auf basis des fuzzy-clustering autonom erkannter systemzustände, PhD Thesis, *Technische Universität Chemnitz*, 2020 (in German).

[19]. C. Trimborn, Signalbasierte zustandsüberwachung von linearantriebskomponenten am beispiel eines pressentransfers, PhD Thesis,, *Technische Universität Chemnitz*, 2021 (in German).

[20]. C. E. Shannon, Communication in the presence of noise, *Proceedings of the IRE*, Vol. 37, Issue 1, 1949, pp. 10-21.

[21]. ISO 3:1973, Preferred numbers – series of preferred numbers, *International Organization for Standardization*, 1973.

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 19-27

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Vehicle Localization Algorithm with Lane Discrimination Based on Inertial and Geomagnetic Sensor Data for GNSS-Denied Environments and Its Evaluation for Different Vehicles and Drivers

Takayoshi YOKOTA
Center for Data Science and Artificial Intelligence Education, Institute of Science Tokyo,
2 Chome-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
Tel.: 090-9003-8168
E-mail: yokota@dsai.isct.ac.jp

Received: 12 Nov. 2024 /Revised:31 March 2025 /Accepted: 21 April 2025/ Published: 30 April 2025

Abstract: We have been developing a method to estimate a moving vehicle's position using MEMS sensor data, including acceleration, gyroscope, and geomagnetic sensors. This algorithm has been evaluated on several fixed courses, achieving good results with a position error of less than 1 meter. However, questions remain regarding whether the algorithm can distinguish between lanes and if it is applicable across different drivers and vehicles than those used to obtain reference data. This paper presents evaluation results of the algorithm conducted on a two-lane road, demonstrating its ability to differentiate between lanes. Furthermore, it confirms that the algorithm can accurately estimate the vehicle's position even when tested with different vehicles and drivers.

Keywords: Localization, Acceleration, Gyroscope, Geomagnetic, MEMS sensor.

1. Introduction

The author has developed a vehicle localization algorithm that reduces reliance on GNSS. Specifically, sensor data from a reference vehicle (Vehicle A) is recorded in advance using MEMS sensors to capture features such as road surface conditions and vehicle heading. Subsequently, sensor data gathered from another vehicle (Vehicle B), which also uses MEMS sensors but lacks GNSS, is cross-correlated with the reference sensor data. By identifying the timestamp in the reference data that produces the highest correlation value, the precise position information from the RTK-GNSS recorded at that timestamp is used to estimate the position of Vehicle B. The outline of the algorithm is shown in Fig. 1. In preliminary experiments conducted in suburban Tottori, Japan, the algorithm achieved a positioning error of 0.37 m in smooth traffic conditions and about 1.6 m in fluctuating traffic conditions [1-4]. To the authors' knowledge, previous studies have focused on fixed road sections with a single lane in each direction [1-16]. To generalize the algorithm for roads with multiple lanes, evaluation tests were conducted in Ibaraki and Tokyo, verifying whether the algorithm can distinguish between adjacent lanes using MEMS sensor data. This paper is the extended version of the conference paper [1].

2. Results of Experiments

2.1. Results of Multiple Lane Distinction on the Shuto Expressway

In Fig. 2, images of the Tokyo Metropolitan Expressway Route 7 (Shuto Expressway) test course

are shown. The course consists of two lanes: a driving lane and a passing lane. Fig. 3 illustrates the overall evaluation test circular course, which includes the Tokyo Metropolitan Expressway Route 7 and is approximately 13 km long.

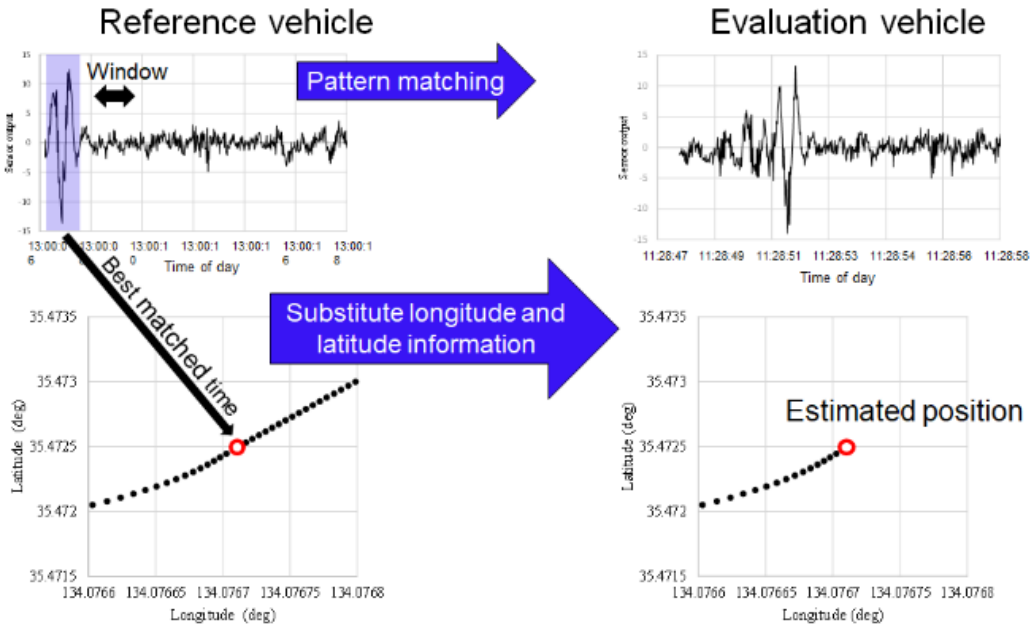


Fig. 1. Outline of the proposed algorithm.

(a) Running in the left lane (b) Running in the right lane

Fig. 2. Metropolitan Expressway Route 7. The test course was 3.9km long.

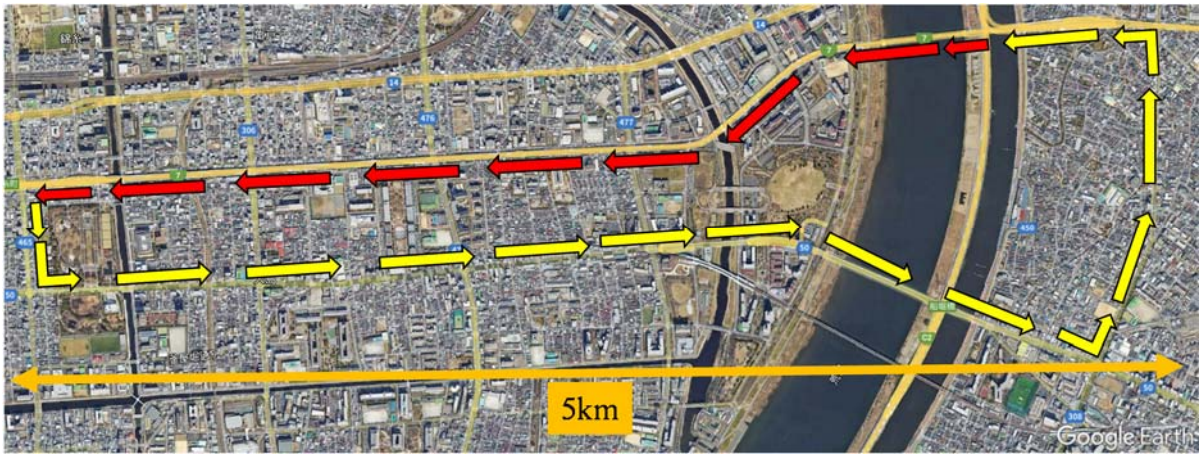
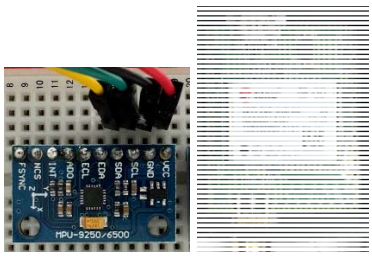


Fig. 3. Road Course for Evaluation Data Acquisition. The total length of the circular course is approximately 13 km. The selected evaluation road section, marked by red arrows, has a length of 3,972 m and is located on Metropolitan Highway Route 7.

Fig. 4 also displays the devices used in the evaluation experiments. Fig. 4(a) shows the MEMS sensor InvenSense MPU-9250[17], which contains a 9-axis sensor, including an accelerometer, gyroscope, and geomagnetic sensor. Fig. 4(b) presents the RTK-GNSS module u-blox F9P [18], which achieves centimeter-level precision under favorable conditions and is used to obtain ground truth data for vehicle positioning.

Fig. 5 shows the Toyota Noah 2-liter, 5-door minivan used for data acquisition during the experiments. In this paper, a specific section of the circular course is selected for detailed examination to address the following purposes. The selected section marked by red arrows, shown in Fig. 3 is 3,972 meters long and corresponds to the Metropolitan Highway Route 7.



(a) MEMS senor (b) RTK-GNSS

Fig. 4. InvenSense's mpu9502 uBlox F9P.

Fig. 5. Vehicle used TOYOTA NOAH 2 litter 5-door minivan.

The objectives of the evaluation experiments are as follows:

1. Can the proposed vehicle localization algorithm identify the specific lane (driving or passing) on which the vehicle is traveling on a multilane road?
2. How accurate is the estimated vehicle position?

To answer these questions, we conducted the following data acquisition runs, summarized in Table 1. The vehicle was driven on the circular test course (Fig. 3) four times. From these runs, we extracted four time-series datasets, each starting and ending at the times specified in Table 1. In Japan, the left lane is referred to as the "driving lane," while the right lane is known as the "passing lane," where vehicles generally travel at higher speeds. This is evident from the travel time data provided in Table 1.

Table 1. Data Acquisition Runs.

TRIP	Running lane	start	end	Travel time(s)
TRIP1	Left	10:26:26	10:29:19	173
TRIP2	Left	10:58:06	11:01:25	195
TRIP3	Right	11:21:30	11:23:52	142
TRIP4	Right	11:46:20	11:48:46	146

2.2. Sensor Data Processing

Pitch rate data obtained during TRIP1 and TRIP2, for example, are shown in Fig. 6. Among the nine types of sensor data, the pitch rate contributes the most significantly [4]. During TRIP1 and TRIP2, the vehicle traveled on the same left lane, and the pitch rate time series are expected to resemble each other. However, since the velocity in TRIP2 is slightly higher than that in TRIP1, the length of the time series differs. Here, the "time of day" in seconds is defined by the following equation:

$$Time\ of\ day\ =\ 3600 \times hour + \\ +60 \times minute + sec \tag{1}$$

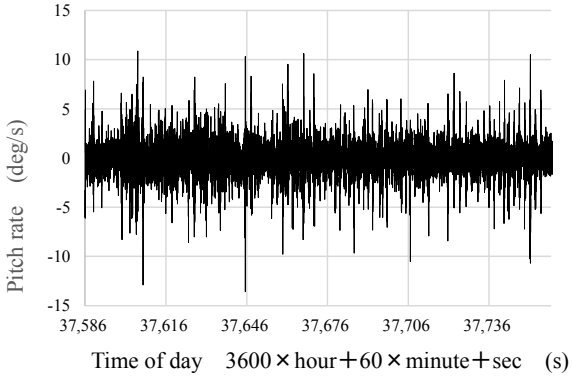


Fig. 6(a). Pitch rate data for TRIP1.

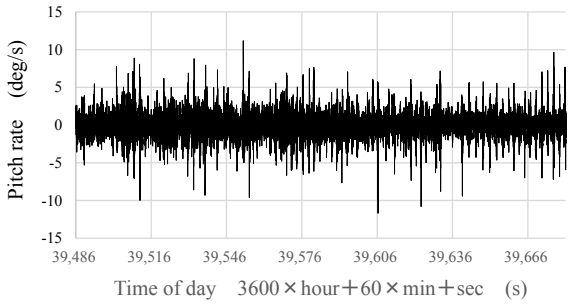


Fig. 6(b). Pitch rate data for TRIP2.

Fig. 7 shows the velocity profiles of TRIP1 and TRIP2. As seen in Fig. 7, the velocity profile is not

constant but varies depending on the driver's characteristics and traffic conditions. The sensor data, in this sense, are velocity-modulated by the velocity profile. The cross-correlation function, calculated from the sensor data under constant velocity conditions, would exhibit a constant time lag between the two sets of vehicle data. In this paper, the first dataset is referred to as the "reference data," while the second dataset is called the ""evaluation data".

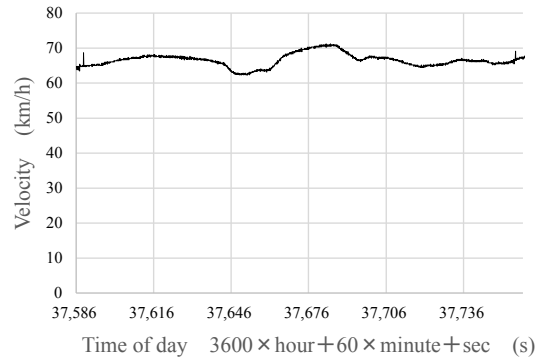


Fig. 7(a). Velocity data for TRIP1.

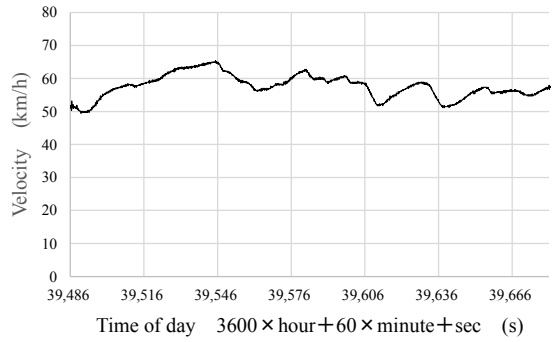


Fig. 7(b). Velocity data for TRIP2.

For this analysis, the pitch rate data from TRIP1 serve as the reference data, and those from TRIP2 serve as the evaluation data.

Examples of cross-correlation calculations for the two sets of MEMS sensor data (TRIP1 and TRIP2, left lane) are shown in Fig. 8. Here, we calculated a weighted sum of the nine cross-correlation functions derived from the 9-axis sensor data (pitch, roll, and yaw rates, as well as acceleration and geomagnetic data for the x, y, and z axes). The optimal weights were determined using the algorithm described in [3]. In Fig. 8, the vertical axis represents the time lag, and the horizontal axis represents the travel time of the evaluation vehicle. The initial time lag was set to 0 seconds.

A bright line is visible in Fig. 8, indicating a strong correlation. However, the bright line descends as travel time progresses due to the non-negligible velocity difference between TRIP1 and TRIP2, as shown in Fig. 7. Consequently, the time lag begins at 0 seconds and ends at approximately -30 seconds. This significant change in time lag complicates the design of a noise filter to accurately estimate the optimal time lag from the cross-correlation results.

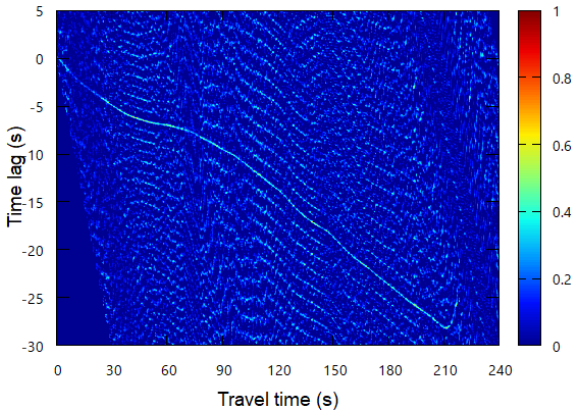


Fig. 8. Cross-correlation of TRIP1-2.

Fig. 9 illustrates the cross-correlation results for TRIP3 and TRIP4 (right lane). In this case, the time lag does not change as significantly as it does for TRIP1 and TRIP2. However, it still presents challenges for designing an effective noise filter.

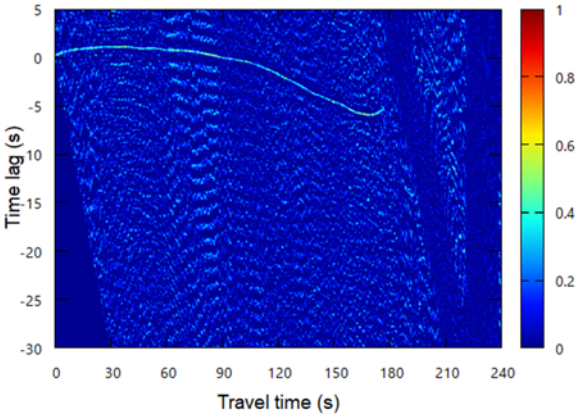


Fig. 9. Cross-correlation of TRIP3-4.

2.3. Velocity Compensation and Cross-Correlation Analysis

Fig. 10 shows the pitch rate sensor data after compensating for them according to the velocity profiles, as if the vehicle were traveling at a constant velocity of 60 km/h. The compensation was applied using the following formula:

$$\varphi(t) = \int_{\tau=0}^{t} \frac{v(\tau)}{v_s} d\tau, \quad (2)$$

where φ(t) represents a hypothetical time axis under the assumption that the vehicle moves at a constant velocity v_s and τ denotes the real-time axis.

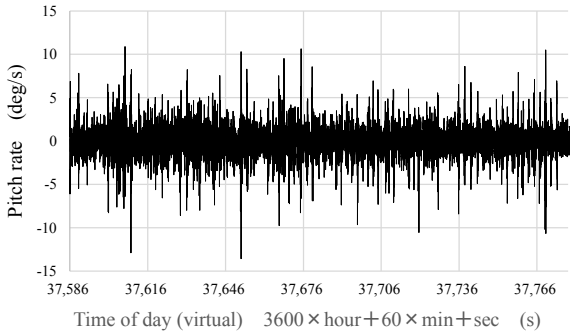


Fig. 10(a). Pitch rate data for TRIP1 after velocity compensation.

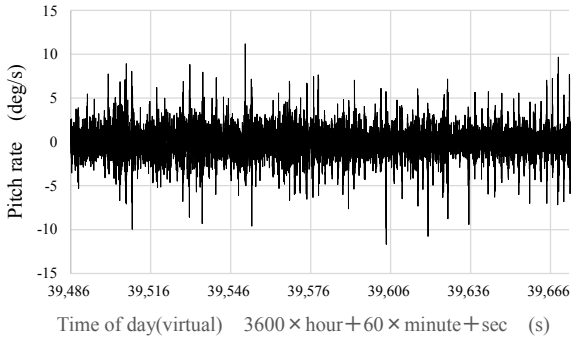


Fig. 10(b). Pitch rate data for TRIP2 after velocity compensation.

After velocity compensation, the travel times of the evaluation vehicle are adjusted to match those of the reference vehicle. Using this velocity-compensated sensor data, the cross-correlation was recalculated, and the results are shown in Fig. 12. In same-lane cases, such as Figs. 11(a) and 11(b), clear horizontal bright lines appear in the correlation functions. However, in the different-lane case shown in Fig. 11(c), where TRIP1 and TRIP3 data were correlated, the correlation values are very poor. Similarly, other different-lane cases, including TRIP1-4, TRIP2-3, and TRIP2-4, also exhibit poor cross-correlation results.

2.4. Vehicle Localization from Cross-Correlation Functions

After obtaining the cross-correlation functions, the algorithm estimates the optimum time lag profile. Once this profile is identified, the algorithm estimates the location of the evaluation vehicle according to the procedure described in Fig. 1. From the results in Fig. 11(a), (b) and (c), the correlation functions for same-lane and different-lane cases show distinct characteristics, enabling lane distinction.

Fig. 12 shows the maximum correlation plots for TRIP3-4 (same-lane case). These plots form an almost horizontal line, though some noisy data points are present.

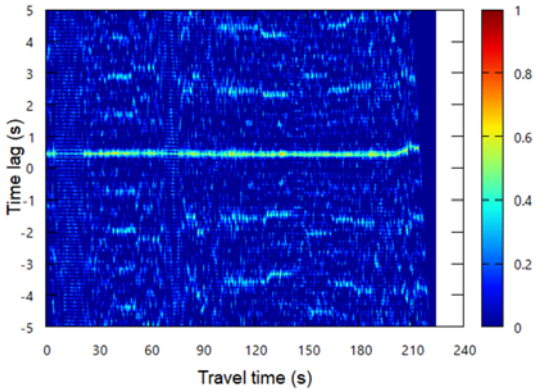


Fig. 11(a). Cross-correlation of TRIP1-2 (driving lane).

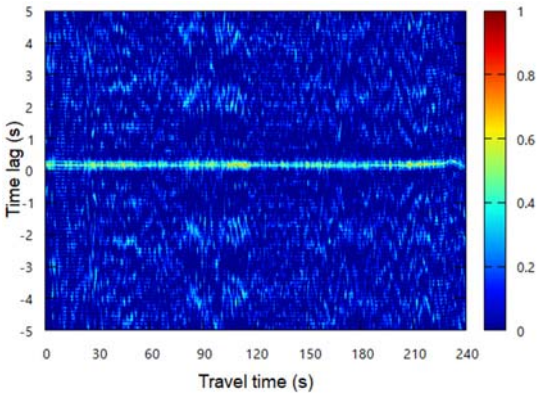


Fig. 11(b). Cross-correlation of TRIP3-4 (passing lane).

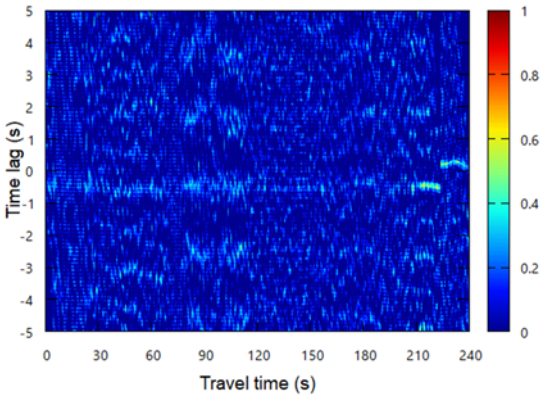


Fig. 11(c). Cross-correlation of TRIP1-3 (different lane).

Fig. 13 displays the result of applying a simple Kalman filter to Fig. 12, yielding a much clearer horizontal line.

Fig. 14 shows the maximum correlation plots for TRIP2-3 (different-lane case). In this case, the plots are highly noisy, and no effective noise filter can produce a horizontal line like that of Fig. 13.

Fig. 15 presents the histograms of maximum correlation plots for TRIP1-3 (different-lane case) and TRIP3-4 (same-lane case). It is evident that the histogram is much broader for different-lane cases compared to same-lane cases. This difference in histogram shapes allows for distinguishing whether the vehicle is running in the same lane as the reference

data. From the time lag estimation, the vehicle's location is derived using the RTK-GNSS location data of the reference vehicle at the corresponding time, as described in Fig. 1.

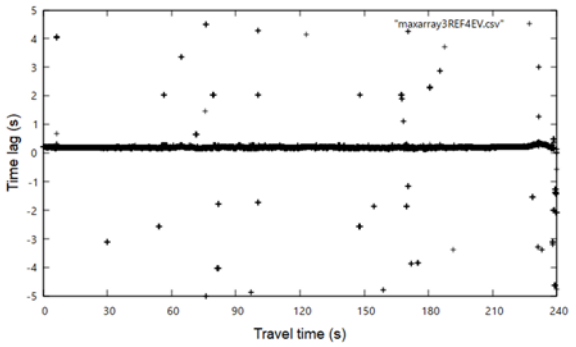


Fig. 12. Maximum correlation plot of TRIP3-4 cross-correlation function.

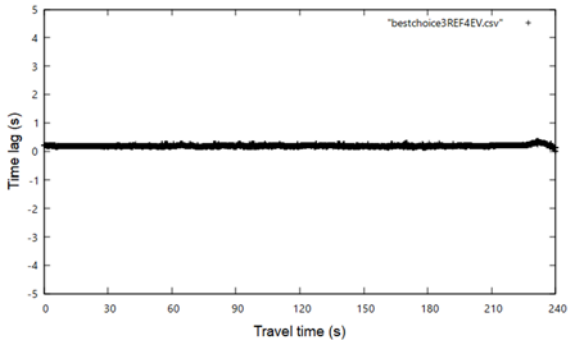


Fig. 13. After applying noise filter to Maximum correlation plot of TRIP3-4 cross-correlation function.

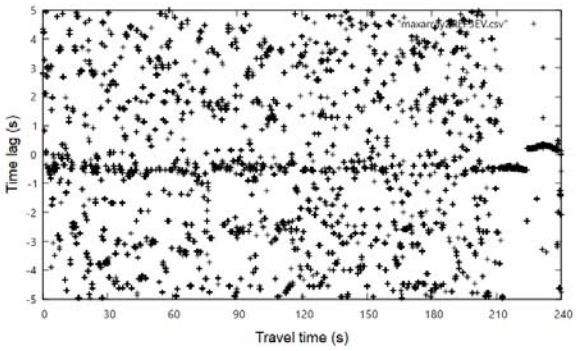


Fig. 14. Maximum correlation plot of TRIP2-3 cross-correlation function. The plot is too noisy to allow the application of a noise filter.

In Fig. 16(a), the position estimation error over time is shown for the evaluation vehicle running in the left lane (driving lane), while Fig. 16(b) illustrates the directional error for the same scenario. Similarly, Fig. 17(a) presents the position estimation error over time for the vehicle running in the right lane (passing lane), with Fig. 17(b) displaying the corresponding directional error. Based on the results from Figs. 16(a) and 17(a), the RMSE errors were 1.40 m for the driving lane and 1.41 m for the passing lane, indicating comparable levels of accuracy in both lanes, despite occasional error spikes. However, when the vehicle is running in a different lane, position estimation becomes challenging due to poor cross-correlation. This limitation, nonetheless, highlights a positive aspect of the algorithm—it effectively identifies the running lane. The directional position error remains confined to the vehicle's heading, provided that lane distinction is properly recognized.

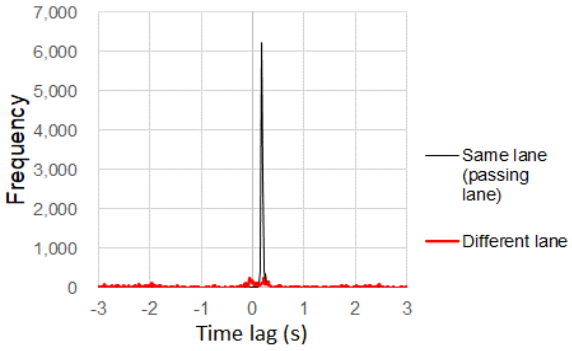


Fig. 15. Histogram of the maxima of the cross cross-correlation function.

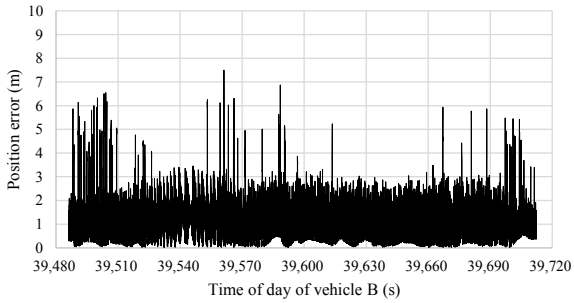


Fig. 16(a). Position estimation error over time when the evaluation vehicle ran in the left lane (driving lane).

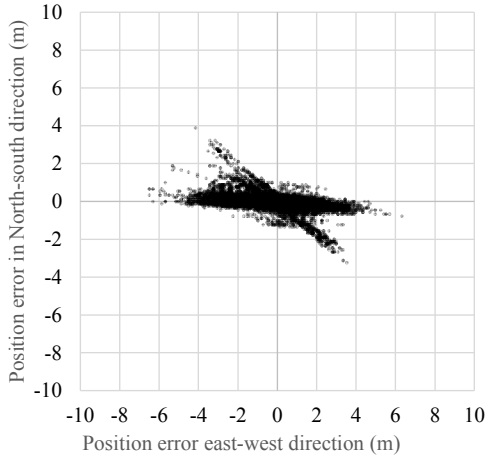


Fig. 16(b). Directional error over time for the same case.

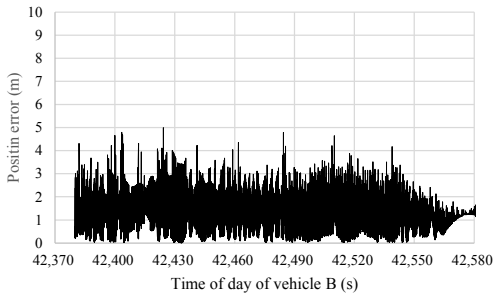


Fig. 17(a). Position estimation error over time when the evaluation vehicle ran in the left lane (passing lane).

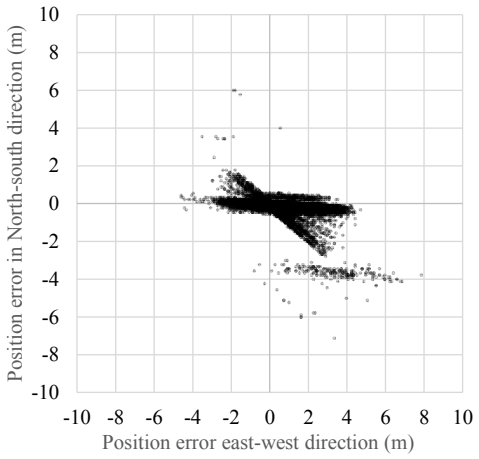


Fig. 17(b). Directional error over time for the same case.

3. Impact of Different Driver and Vehicle on the Proposed Algorithm

Although the lane distinction algorithm performed well on the Metropolitan Expressway, a challenge was encountered on National Route 349 in suburban Ibaraki. The course is shown in Fig. 18. The evaluation section, indicated by the red arrow, is a fairly smooth, two-lane bypass road with a central median running through a rural area.

In this course, we tested the effects of different drivers and different vehicles on the proposed vehicle localization algorithm. Driving data was collected, as shown in Table 2. Eight trips were conducted: the first four trips were driven by a 29-year-old female driver using a SUZUKI SWIFT, while the latter four trips were driven by a 68-year-old male driver using a TOYOTA Yaris Cross.

In Table 3, the two vehicles are compared. The TOYOTA Yaris Cross is larger, heavier, and has a longer wheelbase than the SUZUKI SWIFT. These and the suspension characteristics of the vehicles may influence sensor data to varying degrees.

In Table 4, we summarized the evaluation results by using reference data from each of the eight trips and evaluation data from the same eight trips. The results are presented in a matrix form as shown in Table 4.

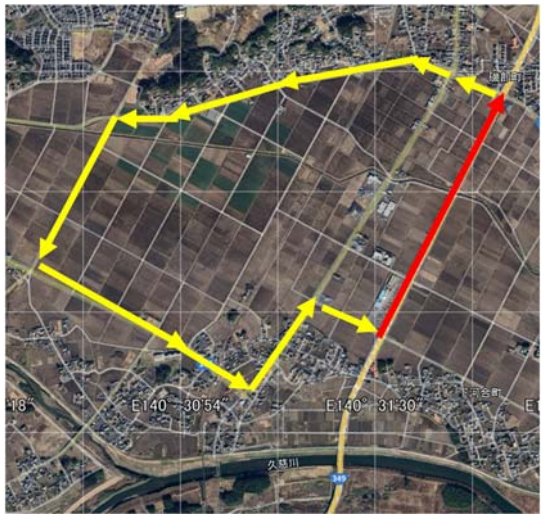


Fig. 18. Road Course for Evaluation Data Acquisition. The total length of the circular course is approximately 5.33 km. The selected evaluation road section, marked by red arrows, has a length of 1.57 km and is located on National route 349 in suburban aria in Hitachi-Ohta city, Ibaraki prefecture.

Fig. 19. Evaluation section view A very flat, two-lane bypass road with a central median running through the rural area.

Table 2. Data Acquisition Runs.

TRIP	Date and time	Driver	Vehilce	Lane	Travel time (s)
0	July 6,2024	29-year-old female	Suzuki SWIFT	Left	80
1				Left	84
2				Right	95
3				Right	82
4	July 7,2024	67-year-old male	TOYOTA Yaris Cross	Left	90
5				Left	95
6				Right	100
7				Right	91

Table 3. Vehicle Comparison.

Vehicle Model	Length (mm)	Width (mm)	Hight (mm)	Wheelbase (mm)	Weight (kg)	Engin Type
SUZUKI SWIFT	3,860	1,695	1,500	2,450	910	Petrol
TYOKOTA Yarsi Cross	4,180	1,765	1,590	2,560	11,160	Hybrid

Table 4. Results of the location errors(m).

			SUZUKI SWIFT				TOYOTA YARIS CROSS			
			TRIP0 lane1	TRIP1 lane1	TRIP2 lane2	TRIP3 lane2	TRIP4 lane1	TRIP5 lane1	TRIP6 lane2	TRIP7 lane2
Ref	SUZUKI SWIFT	TRIP0 lane1		1.44	2.12	3.00	2.09	1.52	10.89	2.78
		TRIP1 lane1	5.35		7.73	8.00	9.51	10.09	11.93	12.82
		TRIP2 lane2	2.17	6.41		1.40	3.08	3.17	1.55	1.51
		TRIP3 lane2	3.27	7.08	1.40		8.21	3.00	1.80	1.61
	TOYOTA YARIS CROSS	TRIP4 lane1	1.55	1.80	3.41	2.55		1.38	5.23	2.54
		TRIP5 lane1	1.58	1.95	3.37	2.37	1.67		4.20	4.03
		TRIP6 lane2	2.07	3.59	11.32	7.36	2.40	2.46		1.91
		TRIP7 lane2	2.06	4.90	2.30	5.28	2.45	2.22	1.90	

To summaries the results shown in Table 4, we can say:

i. The accuracy is generally good when using the same vehicle and driver for both reference and evaluation data;
ii. Even if the vehicle or driver differs between reference and evaluation data, as long as the same lane is driven, no significant increase in error is observed;
iii. Even when the driving lanes differ between reference and evaluation data, positioning is possible due to the correlation. However, errors larger than the lane width may occur;
iv. There are cases where the road surface features are not captured, leading to an increase in errors. (indicated by green areas).

4. Conclusions

This study demonstrated that the MEMS-based vehicle localization algorithm can achieve a position accuracy of approximately 1 meter on the Metropolitan Expressway and effectively distinguish between multiple lanes. However, distinguishing adjacent lanes remains challenging when their sensor data are similar. This issue will be addressed in future research. Improving the results obtained in this experiment may require refining the algorithm to better select the most appropriate reference data. Details of the revised algorithm will be presented in future work. A significant advancement observed is that the algorithm proved to be applicable to different vehicles and drivers, even in limited cases.

Acknowledgements

This work was supported in part by the Japan Road Map Association.

References

[1]. T. Yokota, Vehicle localization algorithm with lane discrimination based on inertial and geomagnetic sensor data for GNSS-denied environments, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics and Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 55-61.

[2]. T. Yokota, Network-wide vehicle localization algorithm based on MEMS sensor data, *Sensors & Transducers*, Vol. 265, Issue 12, 2024, pp. 17-26.

[3]. T. Yokota, Vehicle localization by correlated MEMS sensor data with velocity compensation, in *Proceedings of the IEEE 26th International Conference on Intelligent Transportation Systems (ITSC'23)*, 2023, pp. 1-6.

[4]. T. Yokota, T. Yamagiwa, Vehicle localization by optimally weighted use of MEMS sensor data, in *Proceedings of the 3rd IFSA Winter Conference on Automation, Robotics and Communications for Industry 4.0/5.0 (ARCI'23)*, 2023, pp. 244-249.

[5]. T. Yamagiwa, T. Yokota, Vehicle localization method based on MEMS sensor data comprising pressure, acceleration and angular velocity, in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA'22)*, 2022, pp. 1820-1825.

[6]. T. Yokota, Vehicle localization by altitude data matching in spatial domain and its fusion with dead reckoning, *International Journal of Mechatronics and Automation*, Vol. 8, Issue 4, 2018, pp. 208-216.

[7]. T. Yokota, Vehicle localization by dynamic programming from altitude and yaw rate time series acquired by MEMS sensor, *SICE Journal of Control, Measurement, and System Integration*, Vol. 14, Issue 1, 2021, pp. 78-88.

[8]. T. Yokota, Vehicle localization based on MEMS sensor data, in *Proceedings of the 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE'21)*, 2021, pp. 1468-1473.

[9]. T. Yokota, Localization algorithm based on altitude time series in GNSS-denied environments, in *Proceedings of the SICE Annual Conference (SICE'20)*, 2020, pp. 952-957.

[10]. T. Yokota, M. Okude, T. Sakamoto, R. Kitahara, Fast and robust map-matching algorithm based on a global measure and dynamic programming for sparse probe data, *IET Intelligent Transport Systems*, Vol. 13, Issue 11, 2019, pp. 1613-1623.

[11]. J. Tsurushiro, T. Nagaosa, Vehicle localization using its vibration caused by road surface roughness, in *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES'15)*, 2015, pp. 164-169.

[12]. A. J. Dean, R. D. Martini, S. N. Brennan, Terrain-based road vehicle localization using particle filters, *Vehicle System Dynamics*, Vol. 49, Issue 8, 2011, pp. 1209-1223.

[13]. E. Laftchiev, C. Lagoa, S. Brennan, Terrain-based vehicle localization from real-time data using dynamical models, in *Proceedings of the American Control Conference (ACC'12)*, 2012, pp. 1089-1094.

[14]. J. Gim, C. Ahn, Ground feature-based vehicle positioning, in *Proceedings of the SICE Annual Conference (SICE'20)*, 2020, pp. 983-984.

[15]. J. Gim, C. Ahn, IMU-based virtual road profile sensor for vehicle localization, *Sensors*, Vol. 18, Issue 10, 2018, 3409.

[16]. X. Qu, B. Soheilian, N. Paparoditis, Landmark-based localization in urban environment, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 140, 2018, pp. 90-103.

[17]. InvenSense, MPU-9250 Product Page, https://invensense.tdk.com/products/motion-tracking/9-axis/mpu-9250/

[18]. u-blox, ZED-F9P Product Page, https://www.u-blox.com/en/product/zed-f9p-module

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 28-36

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Robotic Automation for Nuclear Decommissioning: Development of a Tool Carrier System with Milling and Clearance Measurement Capabilities

Siavash KAZEMI
Karlsruhe Institute for technology, Institute of Technology and Management in Construction,
Bldg. 50.33, Am Fasanengarten, D-76131 Karlsruhe, India
Tel.: +49 721 608-48235
E-mail: Siavash.kazemi@kit.edu

Received: 11 Nov. 2024 /Revised:31 March 2025 /Accepted: 28 April 2025/ Published: 30 April 2025

Abstract: As part of the ROBDEKON [1] project, the Karlsruhe Institute of Technology's Institute for Technology and Management in Construction (KIT-TMB) is developing a robotic system designed to position two specialized tools for the decontamination and clearance measurement of concrete walls in nuclear power plants. To achieve this, a Manitou 100-VJR lifting platform has been retrofitted with essential electronic components including controllers, sensors, actuators, communication modules, and safety systems transforming it into a robotic system known as "DekontBot" [2].
In addition to the robotic platform, an automatic milling tool and an automated contamination array have been developed as tools for decontamination and clearance measurement. The system supports operation in both teleoperation and (semi-)autonomous modes. Two graphical user interfaces have been developed, one for the DekontBot and another for the contamination array. These GUIs enable intuitive control and real-time visualization of system operations using RViz2 on the ROS2 platform, allowing users to operate the systems in a user-friendly and accessible manner.

This paper presents a detailed account of the development and implementation of the DekontBot, the milling tool, and the contamination array.

Keywords: Robotic, Automation, Decontamination, PLC, Robot operating system (ROS), Nuclear power plant (NPP), Clearance measurement.

1. Introduction

On April 15, 2023, Germany shut down its last three nuclear power plants Emsland, Neckarwestheim 2, and Isar 2. Currently, there are 25 nuclear power plants and research reactors in Germany at various stages of decommissioning and dismantling [3]. Globally, approximately 297 nuclear power plants are expected to begin the decommissioning process by 2030 [4]. To date, most decommissioning and dismantling activities have been performed manually, making it a high-risk activity not only due to the high levels of radiation but also because of the physical and mental stress imposed on employees (Fig. 1).

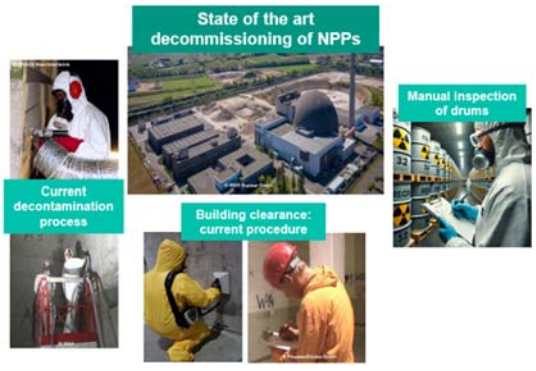


Fig. 1. State of the art in the decommissioning of nuclear power plants.

At the Institute for Technology and Management in Construction (TMB) at the Karlsruhe Institute of Technology (KIT), the "Deconstruction and Decommissioning of Conventional and Nuclear Buildings" department conducts research into topics in the conventional sector such as recycling-friendly dismantling, mechanical demolition methods and the automated separation of hazardous and non-hazardous waste. This field of research is supplemented by the dismantling of nuclear facilities. The focus here is on pilot projects designed to make the dismantling of nuclear power plants safer, more efficient and more economical for everyone involved in the dismantling process. The focus is on developing practical new dismantling technologies (pilot projects) for open problems, including large-scale testing.

However, organizations such as the OECD have noted that current technological capabilities in the dismantling process are not being fully leveraged. Increasing the use of robots and automation technologies offers the potential to reduce both costs and radiation exposure for personnel involved in the decommissioning process. Given the growing number of facilities undergoing dismantling, there is significant potential for digital and automated solutions in this domain [5].

The ROBDEKON project—commissioned by the Federal Ministry of Education and Research (BMBF)—aims to develop autonomous and semi-autonomous robotic solutions to enhance decontamination efforts in hazardous environments [6].

As part of ROBDEKON, KIT-TMB is collaborating with Karlsruhe University of Applied Sciences and Götting KG to develop a comprehensive automation chain for the decommissioning of nuclear power plants [7].

This automation chain comprises several components. One of them is the GammaBot, used for environmental exploration through geometric and radiological measurements. The collected data supports the implementation of Building Information Modeling (BIM) for digitization and detailed planning [8].

Additionally, autonomous decontamination and clearance measurement are carried out using portable robotic tools integrated into an autonomous carrier platform called DekontBot. The system also includes an autonomous forklift truck, developed by Götting KG, to handle the transportation of radioactive waste [9].

A visual representation of this integrated system is shown in Fig. 2.

Fig. 2. Closed automated chain for decommissioning of NPPs.

2. Development of DekontBot

The objective of this research is to develop universal, robot-based solutions for a fully automated decontamination process within nuclear facilities, which can be flexibly adapted to diverse and changing environmental conditions.

The following sections provide a detailed description of the third and fourth steps in the automation chain: automated decontamination and automated clearance measurement, both implemented using the tool-carrier robotic system DekontBot.

2.1. Description of the Robot's Main Task

Once contamination or hotspot locations are identified using the GammaBot, these areas must undergo targeted decontamination. To address this requirement, KIT-TMB has developed an automated, robot-portable milling tool specifically designed for decontaminating concrete walls.

After the decontamination process, the affected surfaces are autonomously measured to ensure they meet the required decontamination standards. For this task, a second tool, known as the "Contamination Array" [10] has been developed.

The primary role of the DekontBot is to accurately position these two tools, operating in either teleoperated or (semi-)autonomous mode, depending on the specific operational needs.

Fig. 3. Actual State of Development of DekobtBot.

2.2. Description of Original Platform

To achieve the described objectives, the "Manitou 100-VJR" platform was further developed and automated to serve as the base vehicle for the robotic system.

The platform measures approximately 1.2 meters in length, 1 meter in width, and 2 meters in height, allowing it to pass through standard doorways and operate in a wide range of indoor environments, from large halls to relatively small rooms.

Equipped with a manipulator featuring three degrees of freedom, the system can reach working heights of up to 8 meters, utilizing a telescopic joint and a robotic arm. Furthermore, it is capable of precisely positioning tools with a maximum weight of 270 kg.

A visual representation of the platform's dimensions and its degrees of freedom is provided in Fig. 4.

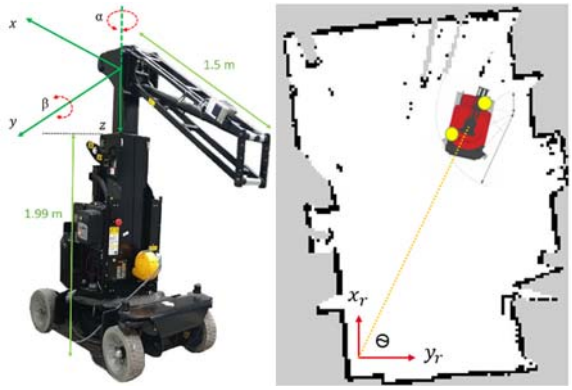


Fig. 4. Dimensions and degrees of freedom of the DekontBot.

2.3. Description of the Integrated Electronics

The foundation of the automated decontamination system is the mobile platform described in Section 2.2. To transform this platform into a fully functional automated system, it was equipped with the necessary components, as illustrated in Fig. 5.

A Rexroth CtrlX3 EtherCAT control unit (PLC) was integrated to serve as the central control system. This PLC connects to the three motor controllers, two for the traction motors and one for the pump motor, via a CAN-EtherCAT gateway. It also interfaces with various sensors to determine the position of the manipulator.

To monitor the manipulator's articulation, two inclination sensors were installed to measure the arm's orientation, and two wire-draw encoders were used to track the position of the telescopic joint and orientation of the tower.

The motor controllers provide real-time feedback to the PLC, including motor current, motor and battery voltage, and motor temperature.

For localization, the platform is equipped with two SICK microScan3 lidar sensors, which continuously calculate the vehicle's position. This positioning data is transmitted to the PLC using UDP communication.

Real-time operations such as motor control, sensor/actuator integration, and automation algorithms are executed on the PLC. Meanwhile, higher-level tasks including visualization, HMI, integration of cameras, and communication with other robotic systems and the control station are handled by a separate PC equipped with an Intel Core i7 CPU.

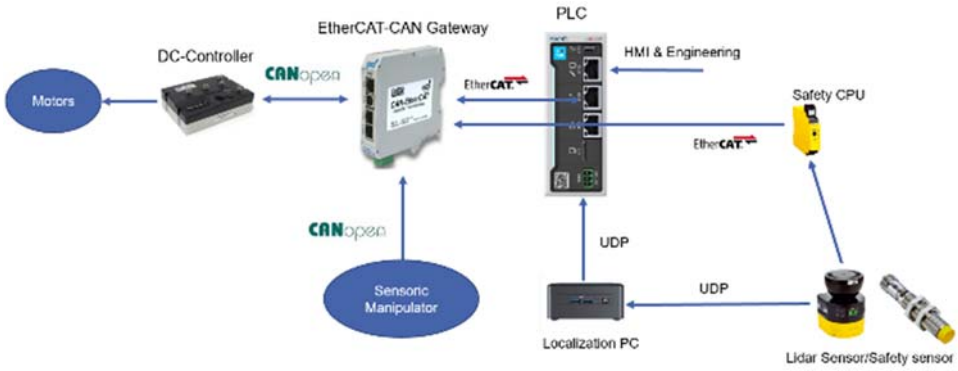


Fig. 5. Integrated electronics on DekontBot [11-14].

2.4. Safety System

System Safety is managed by a SICK-CPU0 safety PLC. The platform is equipped with safety laser scanners that detect people and obstacles in the surrounding area. Based on the detected object's position, the system generates protective or warning signals to prevent collisions or reduce speed.

The safety zones are divided into three color-coded regions, as shown in Fig. 6:

- Red Zone: Indicates a high risk of collision. If an object enters this area in front of the DekontBot, forward motion is immediately stopped;
- Yellow Zone: Indicates a caution zone. When an object is within this area, the platform reduces its maximum driving speed to 50 %;
- Violet Zone: Serves purely as a visual warning for operators. No movement restriction is applied.

The safety system is direction-sensitive: if an object is in front of the machine, only forward motion

is restricted; reverse motion remains possible, and vice versa.

To prevent tipping, additional safety features including tilt measurement system is mounted, which consist of two analog tilt sensors that monitor the inclination of the platform. The first sensor continuously measures the platform's tilt, while the second acts as a redundant monitoring unit. If the measured tilt exceeds ±10°, or if the readings between the two sensors differ beyond a defined tolerance, forward movement is blocked if the condition occurs while moving forward, and reverse movement is blocked if it occurs during backward motion.

The movement of the manipulator joints is restricted to a tilt angle of ±2° to prevent the platform from overturning, particularly when handling tools weighing up to 270 kg at a height of 8 meters. Additionally, five digital switches are installed to verify that the manipulator is in its home position, which is a precondition for allowing platform movement. Other switches monitor whether any manipulator joint is at its end position, preventing unnecessary hydraulic pressure buildup.

The positions of the wheels are measured using analog potentiometers, which serve as both end switches and inputs for visualizing the wheel orientation in a Unified Robot Description Format (URDF) model. All safety-critical signals are routed through a dedicated gateway to the main control unit. In hazardous situations, such as potential collisions or risks of overturning, the system can halt all movement immediately, wait until conditions return to a safe state, or autonomously recalculate and execute an alternative path to reach its target.

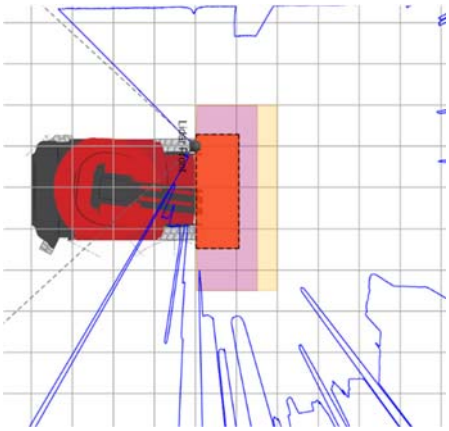


Fig. 6. Safety Zones of DekontBot: Front Safety Scanner.

3. GUI, HMI, Visualisation and Data Transfer

To enable intuitive user interaction, a user-friendly interface has been developed, allowing operation of the system without requiring in-depth technical knowledge. The interface supports language switching between German and English. (Fig. 7).

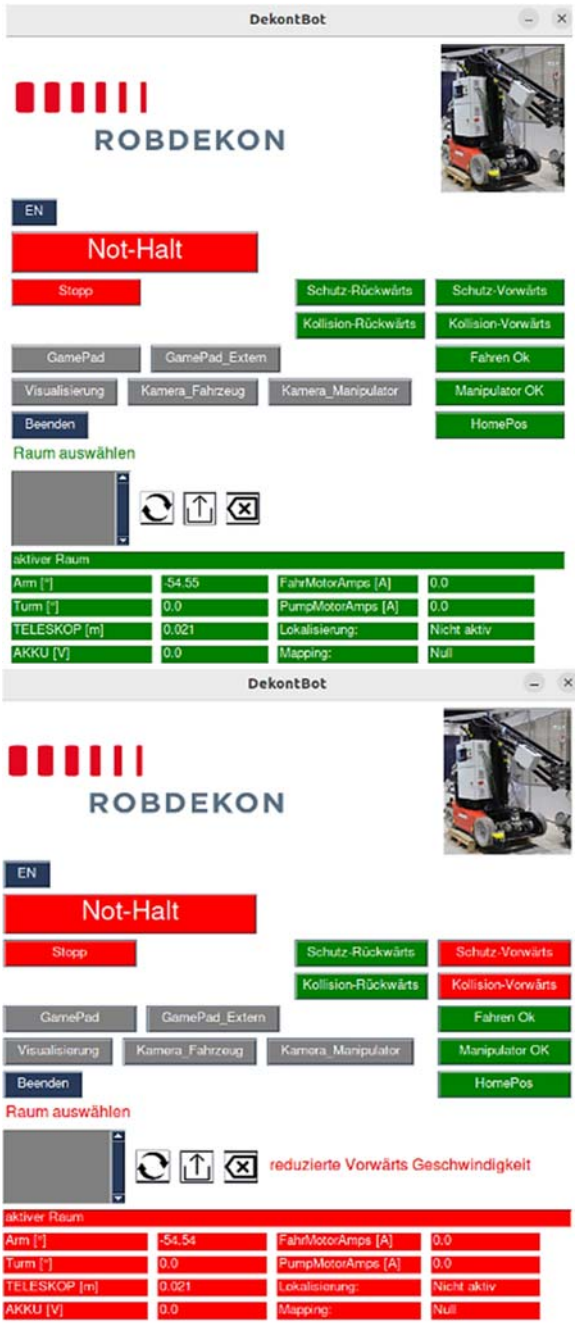


Fig. 7. Top: No Safety Warning Signal, Down: Collision Signal.

The decontamination process is based on a contamination map generated in advance by the GammaBot. This map includes a two-dimensional layout for localizing the DekontBot, as well as a three-dimensional point cloud enriched with radiological information. Using this data, the DekontBot can identify contaminated areas, calculate the optimal trajectory from the starting point to the contamination sites, and position the required tools for decontamination or clearance measurement accordingly. All relevant information is transmitted automatically from the GammaBot to the DekontBot via a TCP server/client architecture using the ROBDEKON VPN.

The 2D map will be converted, processed, and the localization information will be transmitted from the PC to the PLC (Fig. 8).

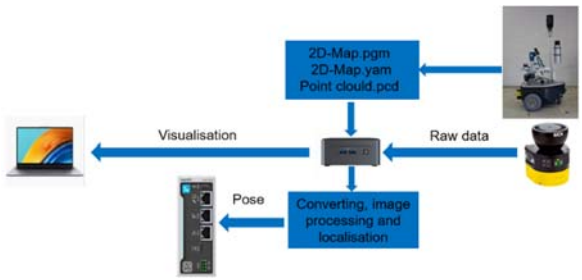


Fig. 8. Information Processing: From Transport to Localization and Visualization [11].

Through the graphical user interface (Fig. 7), users can select or delete existing contamination maps, start or stop system operation, and switch between teleoperation and automatic mode. The interface also displays key safety and system parameters, including collision risks, manipulator joint positions, battery voltage, motor currents, and localization status. Changes in system status are visualized through color transitions, from green to yellow or red along with corresponding warning messages.

To enhance spatial awareness, the selected contamination map is visualized in RViz2 using ROS 2. This includes both the 2D map and the 3D point cloud of the building, overlaid with a dynamic URDF model of the DekontBot (Fig. 9), providing real-time visualization of the robot's position and configuration.

For additional situational awareness, the platform is equipped with two USB 3.0 cameras: one for monitoring the vehicle's surroundings and another for observing the manipulator during operation.

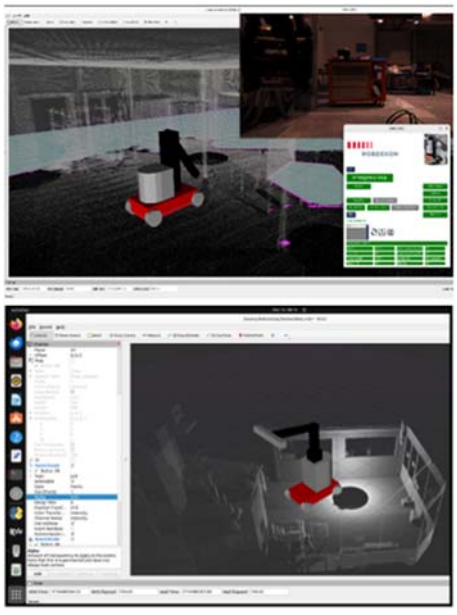


Fig. 9. Visualisation of point cloud and URDF model.

4. Developed Tools

As previously mentioned, the primary task of the DekontBot is to position two end-effectors: a milling tool for decontamination and a contamination array for clearance measurements, at the designated target positions. The following section provides a detailed description of the goals and development of these tools.

4.1. Milling Tool

The structure of the milling tool is shown in the Fig. 10.

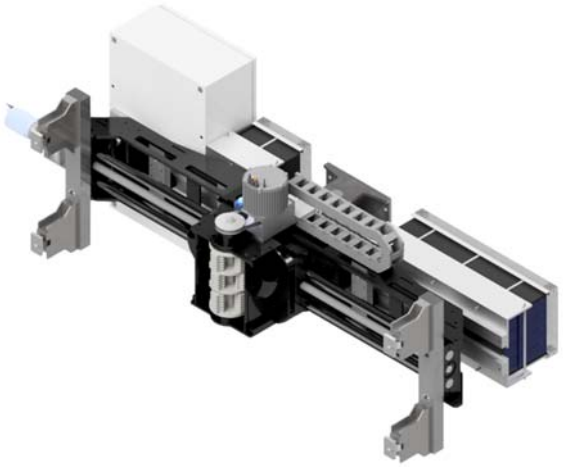


Fig. 10. The structure of the milling tool.

The mechanical components of the milling tool were developed by Contec, GmbH [15], while the electrical cabinet (Fig. 11), battery, motors, sensors, and corresponding software development were handled by KIT-TMB.

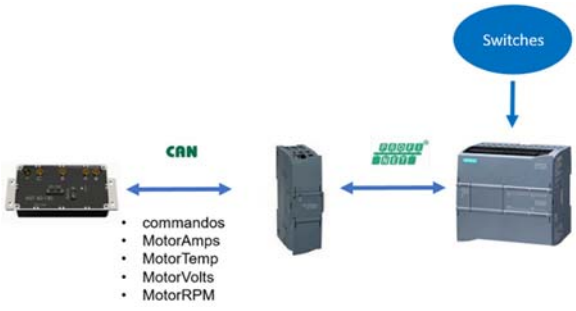


Fig. 11. Integrated electronics on milling tool [16-18].

The milling tool is equipped with 2 degrees of freedom, facilitated by two motors. The first motor, with a power of 2.2 kW, drives the milling head and executes the cutting process. The second motor, with a power of 0.4 kW, is responsible for the lateral motion of the milling head.

Four digital switches are mounted at the front of the machine. Once the milling tool makes contact with the surface, all four switches are activated, and the machine begins operation. This mechanism is primarily for safety and to reduce power consumption, preventing the tool from operating when it is not in contact with the surface.

Additionally, two sensors are mounted on the sides of the milling tool. When the milling head reaches the end of its trajectory, these sensors are activated, prompting the lateral motion of the milling head to shift in the opposite direction.

To control the lateral velocity of the milling head and ensure optimal milling performance, a current sensor is installed to measure the current of the milling motor. When the motor's current is low (i.e., during free-run operation), the lateral velocity is high, indicating that there is little to no resistance and the head is not cutting material. If the current remains low, the tool will stop cutting.

Conversely, when the milling motor's current increases, the lateral velocity decreases. This increase in current indicates that the milling head is encountering material to cut, leading to higher torque and a reduction in lateral speed to optimize the cutting process.

The milling tool is equipped with a Siemens CPU, which communicates with the motor controllers via CAN protocol using a Profinet/CAN gateway. The entire milling tool is powered by a 48 V, 120 ampere-hour battery.

Communication between the milling tool and the DekontBot is established through UDP and TCP protocols. This integration allows the milling tool to be monitored and controlled via the GUI. Sensor data, such as motor speed, motor currents, and battery voltage, is transmitted via UDP, while commands such as Release or Stop are sent using TCP.

Additionally, the milling tool is equipped with a suction unit for efficient dust and debris collection during the milling process.

Fig. 12 shows the milling result of the tool on a concrete wall.

Fig. 12. Milling result.

4.2. Contamination Array

To achieve the step of autonomous clearance measurement, a robot-portable, battery-operated tool known as the ‘contamination array’ was developed at KIT-TMB.

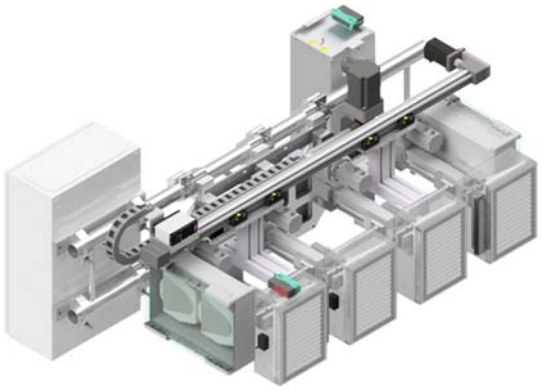


Fig. 13. The structure of the contamination array.

As shown in Fig. 13, the contamination array consists of four contamination detectors mounted in parallel. It is crucial to prevent the detectors from making contact with surfaces, as this could damage them or cause them to become self-contaminated, leading to inaccurate measurements.

To address this, the first step in the measurement process involves scanning the surfaces using a laser scanner mounted on a linear drive to detect any obstacles. Fig. 14 illustrates the scanning area and the scan profile.

Using this information, each contamination detector will be positioned accurately using separate linear drives. This ensures that the detectors not only measure surface activity from a pre-defined distance but also avoid collisions with detected obstacles. Once the detectors are properly positioned, the radiological measurements will begin, lasting approximately 10 seconds.

In the next step, the entire system will move sideways to cover the areas between the four contamination detectors, as shown in Fig. 13, and to position the detectors in areas between any obstacles.

Additionally, external laser distance sensors are mounted on the tool to measure the positions of the machine within its own coordinate system. The final output of the contamination array will be a digital documentation file containing all necessary information, such as radiological measurements, positions, time, and date.

Similar to the DekontBot, a GUI has also been developed for the contamination array, as illustrated in Fig. 15.

Using this GUI, the user can start or stop the measurements, monitor the radiological measurement values, track the positions, and view the scan profile of the surfaces.

Additionally, users can initiate zero-effect measurements to assess background radiation for calibrating the detectors. Similar to the milling tool, the system also includes an interface to transfer

information between the contamination array and the DekontBot.

5. Results and Further Developments

The mechanical and electrical expansion of the DekontBot and its two tool has been successfully completed, and the platform is now fully operational for teleoperation. All degrees of freedom of the original platform are controllable through the integrated electronic components. The required system operation values are measurable via selected sensors, and the platform's safety hardware is fully integrated.

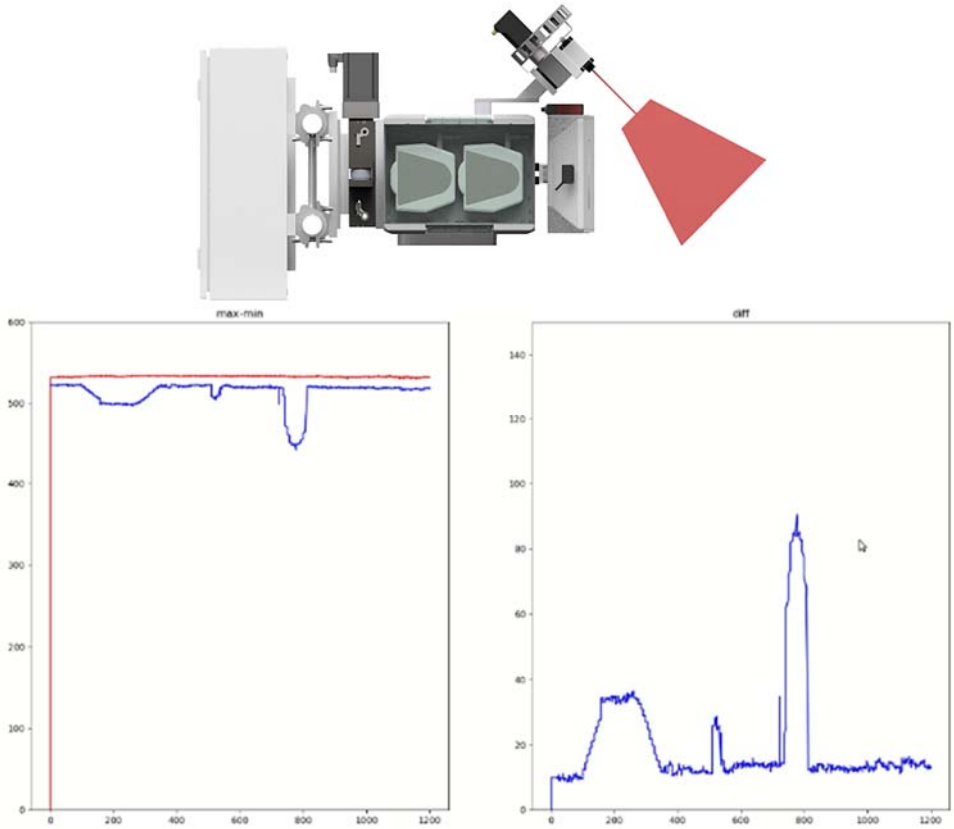


Fig. 14. Scanning area and the scan profile.

Fig. 15. GUI of contamination array.

With the developed GUI and visualization tools, the systems can be operated without prior knowledge of robotics or programming, making it highly user-friendly for its intended users.

The next phase involves integrating path planning and control algorithms to automate the systems further. Additionally, to ensure more precise tool positioning, particularly for the decontamination and clearance measurement of inclined walls, floors, and ceilings, further degrees of freedom will be added to the end effector, as shown in Fig. 16.

To achieve this, an additional joint will be designed and built at KIT-TMB, enabling both end-effectors to rotate ±90° around their axis. Two parallel-mounted analog laser distance sensors will be used to measure the inclination of the surfaces. The 0.24 kW motor will be controlled via a motor controller, which will be connected to the central PLC using the CANOpen protocol.

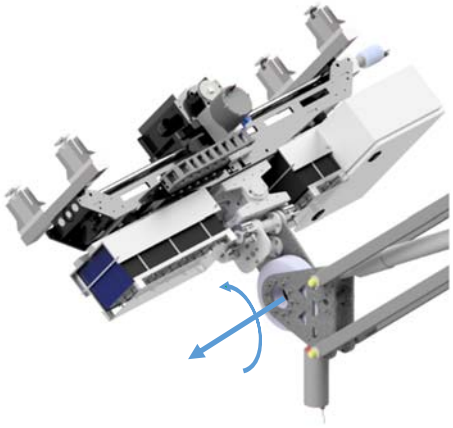


Fig. 16. Swivel module for expanding the degrees of freedom of the end effector.

The safety system also requires enhancements to meet the TÜV (Technical Inspection Association) testing and verification standards. Additional optimizations are necessary, such as improving the quality of point clouds, refining details in the URDF model, and ensuring robust communication between the DekontBot and the control center. These improvements will ensure that the DekontBot and its tools are fully prepared for initial testing within a nuclear facility.

6. Conclusions

It has been demonstrated that automation and robot-based solutions hold significant potential in the decontamination processes of nuclear facilities. These solutions not only alleviate the physical burden on personnel, who typically carry heavy tools and wear uncomfortable protective suits, but also significantly reduce the radiation exposure to their bodies and minimize the generation of secondary waste.

As a result, the DekontBot and its two end-effectors were developed specifically for decontamination and clearance measurement of concrete walls in nuclear power plants. In collaboration with partners like Karlsruhe University of Applied Sciences and Götting, who are developing the GammaBot and the automated forklift, the automated decontamination chain will soon be completed and made available to users.

With the integration of this automated chain, the effort and associated risks of these tasks can be substantially reduced. There is considerable interest from operators of nuclear power plants in adopting autonomous solutions or robotic systems for decontamination operations.

Furthermore, the DekontBot can also serve as a tool carrier for a variety of tools weighing up to 270 kg. After discussions with relevant personnel, it was determined that one potential application for the DekontBot would be in conventional decontamination work (non-nuclear), such as chemical decontamination removal.

References

[1]. J. Petereit, M. Hoffmann, H. Wüstenberg, M. Pfingsthorn, et al., ROBDEKON: robotic systems for decontamination in hazardous environments, in *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR'19)*, 2019, pp. 110-116.

[2]. S. Kazemi, S. Gentes, Development of a robot system as a tool-carrier machine for the decontamination process of nuclear power plants, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 47-52.

[3]. KernD.de, Kernkraftwerke außer Betrieb: Leistung und erzeugter Strom, https://www.kernd.de/kernd/themen/strom/Zahlen-und-Fakten/01_index.php#anchor_27a91b5f_Accordion-Kernkraftwerke-ausser-Betrieb

[4]. A. D. Little / Statista, Rückbau von Kernkraftwerken weltweit bis zum Jahr 2030, https://de.statista.com/statistik/daten/studie/201684/umfrage/rueckbau-von-kernkraftwerken-weltweit/

[5]. R. McGrath, R. Reid, P. Tran, Guidance in the use of robotics and automation for decommissioning nuclear power plants, EPRI Report 3002011738, Electric Power Research Institute (EPRI), 2017.

[6]. P. Woock, et al., ROBDEKON – competence center for decontamination robotics, *at – Automatisierungstechnik*, Vol. 70, Issue 10, 2022, pp. 827-837.

[7]. S. Gentes, et al., From environmental exploration to clearance measurement – developing mobile robot systems for decommissioning of nuclear power plants, *at – Automatisierungstechnik*, Vol. 70, Issue 10, 2022, pp. 850-862.

[8]. Z. Chen, S. Gentes, From point cloud to as-built BIM: semi-automated wall reconstruction for dismantling of nuclear power plants, in *Proceedings of the 32. Forum Bauinformatik*, 2021, pp. 151-158.

[9]. Göetting KG, News, https://www.goetting.de/news/

[10]. A. Wernke, Automatisierung der kontaminationsmessung im rückbauprozess kerntechnischer anlagen, *atw – International Journal for Nuclear Power*, Vol. 67, Issue 5, 2022, pp. 277-281 (in German).
[11]. Bosch Rexroth, Application & Tools Portal, https://apps.boschrexroth.com
[12]. SICK AG, Austria Homepage, https://www.sick.com/at/de
[13]. Esd electronics GMBH, Homepage, https://esd.eu/
[14]. Roboteq, https://www.roboteq.com/index.php
[15]. CONTEC GmbH, Deutschland Homepage, https://www.contecgmbh.com/de
[16]. Siemens, Deutschland Homepage, https://www.siemens.com/de
[17]. HMS Networks, https://www.hms-networks.com
[18]. Plettenberg Motors, https://plettenbergmotors.com

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 37-44

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

A COTS-based Lightweight, Low-power and Versatile Companion Computer for Nano UAVs

Rintaro OKUDERA, Yixiao LI, Yutaka MATSUBARA and Hiroaki TAKADA

Graduate School of Informatics, Nagoya University, Nagoya 464-8601, Japan
E-mail: r_okudera@ertl.jp

Received:29 Jan. 2025 /Revised: 14 March 2025 /Accepted:21 April 2025 /Published:30 April 2025

Abstract: Recent advances in Unmanned Aerial Vehicles (UAVs) have driven a rapid expansion in their application domains, ranging from logistics and surveillance to disaster relief and entertainment. In particular, Nano UAVs (weight: under 100 g) offer significant advantages in cost, safety, and regulatory compliance. However, their limited payload capacity and the high-power consumption of conventional control systems severely constrain flight time. Traditional companion computers designed for computationally intensive tasks, such as image processing, often require 2-5 times more power and nearly double the weight of the onboard flight controller, further exacerbating these limitations. In this study, we present a novel companion computer based on Sony's Spresense platform that is lightweight, low-power, and highly versatile. Weighing only 7 grams and featuring a POSIX□compliant RTOS alongside a multi□core architecture, the proposed solution is optimally tailored for Nano UAV applications. Extensive Hardware□in□the□Loop (HITL) evaluations demonstrate that our Spresense□based system consumes only 1/18 the power of the Raspberry Pi 4 Model B while delivering comparable functionality. Moreover, real Nano UAV tests using the Nano Mind 110 (weight: 36 g) confirm that integrating our companion computer results in only a modest increase in overall power consumption, thus preserving flight time. Compared to current state□of□the□art approaches, our design effectively addresses challenges in availability, extensibility, and ease of development, offering a cost□effective and practical alternative for Nano UAV systems. Future work will extend these findings through further real□world validations, including advanced navigation, obstacle avoidance, and ROS□based applications, to confirm the robustness and scalability of our approach.

Keywords: Nano UAVs, Low-power, Companion computer, Versatile UAV applications, POSIX-compliant RTOS.

1. Introduction

In recent years, Unmanned Aerial Vehicles (UAVs) have rapidly expanded in their application scenarios. UAVs are increasingly being considered a novel means of transportation for both people and goods, and are now widely employed in logistics, disaster relief, and entertainment. In particular, Nano UAVs – those weighing under 100 grams – excel in terms of cost, ease of maintenance, regulatory compliance, and safety, making them suitable for both indoor and outdoor applications [1]. However, when deploying Nano UAVs in professional settings, several challenges remain.

(1) Flight Time

Large UAVs can extend their flight time by incorporating high-capacity batteries and engine power [2]. However, such solutions are not feasible for Nano UAVs due to their strict size and weight restrictions, since every gram allocated to onboard electronics or sensors directly competes with the available capacity for battery storage.

(2) Power Consumption

UAVs inherently require high power consumption for their propulsion systems – especially for the motors – which results in short continuous flight time. Research efforts have focused on lightweight designs, improved aerodynamic performance, motor

enhancements, and refined control algorithms to address this issue [3]. Moreover, as the range of UAV applications expands, the computational power required is also increasing. For example, consider image processing: while early UAVs (e.g., around 2010 with the AR.Drone) typically used VGA-resolution cameras (approximately 300000 pixels), it is now common in 2025 to mount cameras with FHD or even higher resolutions. When applying a linear filter, the computational load for processing an 8K image (approximately 30000000 pixels) is roughly 100 times that for a VGA image. Although this example involves relatively simple processing, more complex tasks can demand even greater computational resources. Consequently, the computational performance required by UAVs is growing exponentially.

(3) System Configuration

Fig. 1 illustrates a simplified system configuration for typical UAVs, which generally consists of two main components:

- **Flight Controller:** It handles basic tasks by computing the UAV's attitude from sensors (e.g., IMU, GNSS, pressure sensor, etc.) and controlling the motors that drive the propellers. Its applications typically run on RTOS (Real-time Operating System) based firmware such as PX4 from the open□source Dronecode ecosystem [4];
- **Companion Computer:** It manages advanced tasks (image processing, SLAM, path planning, etc.) using data from cameras and depth sensors. Most of the existing solutions are based on GPOS (General-purpose Operating System) like Linux.

Typically, flight controllers are built around Cortex□M microcontrollers, whereas companion computers use platforms such as Raspberry Pi 4 or NVIDIA Jetson. Compared to the flight controller, the companion computer generally consumes 2–5 times more power and generates significant heat. In addition, it is roughly twice as heavy, and the need for a heat sink further increases its overall weight. These factors combine to reduce flight time.

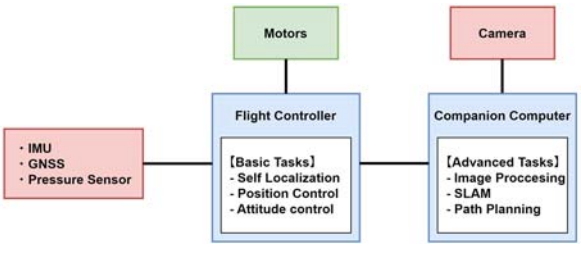


Fig. 1. System Configuration of Typical UAVs.

(4) Diversification of Functional Requirements

Initially, companion computers were primarily used for processing camera images. With the expansion of UAV applications, their roles have diversified to include autonomous flight control, AI processing, and swarm control. Since the choice of companion computers must align with the required functions, there is no one-size-fits-all solution. Moreover, more advanced functionalities typically come with increased power consumption and weight, further reducing flight time. For additional examples, please refer to the survey by Foisal et al. [5].

From the above discussion, reducing power consumption and enhancing the versatility of companion computers emerge as critical challenges. In this study, we present a new companion computer suitable for Nano UAVs built on PX4. We have implemented the necessary communication architecture and utilized multi-core task processing to achieve both energy efficiency and high performance. The effectiveness is evaluated by comparing our proposed approach with Raspberry Pi 4. This paper is an extension of work originally presented in DAUS' (International Conference on Drones and Unmanned Systems) 2025 [6]. In this paper, we present new evaluation results including experiments on real Nano UAV and more literature reviews to further validate our approach.

This section has provided an overview of our study and its contributions. Section 2 reviews related studies and products, Section 3 details the proposed Spresense-based companion computer and its implementation, Section 4 presents our HITL evaluation experiments, Section 5 presents our real-world evaluation experiments, and Section 6 positions our work within the context of existing studies before concluding in Section 7.

2. Related Work

This section describes state-of-the-art studies and products related to companion computers for Nano UAVs.

2.1. FPGA-based Approaches

Cheng et al. [7] developed a UAV control system using an FPGA as a companion computer for Bitcraze's Crazyflie 2.1 [8]. The system utilizes AMD's XC6SLX9 FPGA [9], which features 5720 slice LUTs, 11440 flip-flops, and 32 blocks of 18 Kbit Block RAM. The primary advantage of using an FPGA is its extremely low power consumption – over 100 times lower than that of a comparable embedded GPU. However, due to inherent limitations in implementing complex arithmetic circuits on FPGAs, these devices are less versatile for executing complex computations.

2.2. SoC-based Approaches

GAP8 [10] is a commercial product from GreenWaves Technologies that complies with the

Crazyflie-AIdeck [11] standard and is designed to offload mission control tasks from Crazyflie. GAP8 integrates a Ri5cy core as its host CPU, is equipped with 1.5 MB of on-chip SRAM, and features a parallel-programmable cluster of eight additional Ri5cy cores. The Ri5cy is a 4-stage pipeline core based on RV32 and compliant with the custom XpulpV2 RISC-V ISA, which includes extensions for DSP and machine learning applications (supporting 16/8-bit SIMD operations and hardware loops). In GAP8, only the Fabric Controller Core can access peripheral devices, while the remaining cores serve solely as accelerators for computation. This limitation restricts scalability and versatility for complex IoT applications. Moreover, although these systems support an RTOS, they do not support Linux [12], making it difficult to leverage existing Linux-based software assets.

In contrast, Shaheen [13] supports both RTOS and Linux, enabling the use of extensive software stacks such as ROS. It integrates a fully programmable parallel 8-core RV32 cluster accelerator. The RV32 cores in Shaheen are derived from Ri5cyNN cores and support mixed precision, achieving up to 8.5 times faster performance than Kraken. Additionally, it incorporates an RV64 core featuring advanced virtualization and security functions, along with up to 512 MB of main memory. However, Shaheen is only a prototype and has not been released commercially. Furthermore, while the SoC's power consumption is discussed, the overall board power consumption and scalability remain unclear, leaving its effectiveness as a companion computer for Nano UAVs uncertain.

3. Proposed Method

3.1. Overview

Companion computers always face trade-offs among versatility, power efficiency, and weight. The main objective of our approach is to develop a companion computer that enhances the performance of Nano UAVs by effectively balancing these three factors. Our proposed companion computer is based on Sony's Spresense, a commercial off-the-shelf (COTS) microcontroller board developed by Sony [14] (see Fig. 2 for an overview of the board's design and key components). Its compact and lightweight design makes it particularly well-suited for Nano UAV applications compared to larger solutions (e.g., Raspberry Pi 4: 85 × 56 mm, 46 g).

Spresense is equipped with a variety of useful peripherals, including an integrated Global Navigation Satellite System (GNSS) module for accurate navigation and localization. This high degree of extensibility enables users to develop a wide range of feature-rich UAV applications. Moreover, many official and third-party add-on boards are available to provide additional functionalities, such as high-resolution cameras, extra sensors, Wi-Fi, and LTE connectivity.

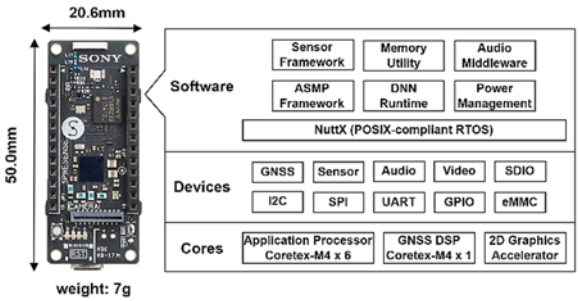


Fig. 2. Overview of Sony Spresense Board.

The application processor on Spresense features six Arm Cortex□M4 cores, offering robust parallel processing capabilities for demanding tasks such as image processing and Deep Neural Network (DNN) inference. A dedicated Arm Cortex□M4 DSP handles GNSS processing, further enhancing overall performance. The board's SDK is based on NuttX, a POSIX□compliant RTOS that provides both high real□time performance and efficient development productivity.

The unique combination of performance, power efficiency, and extensibility provided by Spresense has proven its value in space robotics applications, including lunar exploration robots [15]. In this study, we implement functionalities that enable Spresense to serve as a companion computer within the Dronecode open□source UAV development ecosystem [4]. This approach not only facilitates efficient and rapid prototyping of Nano UAVs but also benefits from a supportive community that continuously contributes new features and improvements.

3.2. Comparison with Candidate Boards

We have also investigated other popular COTS boards with size small enough to be equipped on a Nano UAV. Table 1 shows a comparison of the Raspberry Pi Zero WH [16], Arduino Nano 3.0 [17] and Spresense on power consumption, performance and capabilities. For Raspberry Pi Zero WH, the power usage is measured with Wi-Fi and Bluetooth turned off. As the table indicates, Spresense provides computational performance comparable to the Raspberry Pi Zero WH while consuming only 1/16 of its power and less than 1/3 of the power of the Arduino Nano 3.0. Thus, we determined that Spresense was the most suitable COTS board to serve as a companion computer for Nano UAVs.

3.3. Communication Architecture Implementation

PX4 is the firmware for flight controllers in Dronecode ecosystem. Fig. 3 depicts the communication architecture between the flight controller and our companion computer. By default,

PX4 and Spresense share no common protocol. Thus, we ported the MAVLink protocol [18] to Spresense due to its lightweight and reliable messaging capabilities. MAVLink enables seamless data exchange between PX4 and Spresense through a serial interface. This also allows users or developers to bridge existing ROS2 applications on Spresense using micro-ROS via MAVLink [19].

PX4 applications cannot directly use MAVLink for communication. Instead, they are required to use the publisher-subscriber middleware uORB provided by PX4. To fully leverage the functionalities of Spresense, we have also extended the PX4 firmware to enable the translation between custom MAVLink messages and uORB messages. This enables developers to offload flight controller functions – such as GNSS data acquisition – to Spresense.

Existing high-level APIs such as MAVSDK [20] offer powerful UAV control features but are not optimized for microcontrollers like Spresense [21]. We addressed this by developing custom UAV control APIs based on the MAVLink protocol. Through analysis of MAVLink message flows in existing systems, we tailored them to Spresense, creating lightweight APIs optimized for resource-limited environments while maintaining compatibility with PX4.

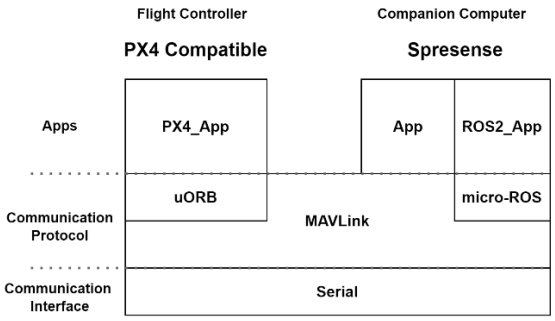


Fig. 3. Communication Architecture between the Flight Controller and the Companion Computer.

Table 1. Comparison of COTS boards with tiny form factor.

	Raspberry Pi Zero WH [16]	**Arduino Nano 3.0 [17]**	**Spresense [13]**
Power (No load) [mW]	500	100	30
DMIPS	1250	20	1200
Size	65.0 mm × 30.0 mm	43.2 mm × 18.5 mm	50.0 mm × 20.6 mm
Weight	7 grams	9 grams	7 grams
Capabilities	Wi-Fi/Bluetooth, Display output, Camera interface	-	GNSS receiver, Hi Resolution Audio I/O, Camera interface

3.4. Multi-core Task Processing

Spresense operates at 156 MHz per core, which is significantly lower than the 1.8 GHz found in Raspberry Pi 4. This lower clock speed contributes to its ultra□low power consumption. To overcome this limitation, we exploit Spresense's multiple cores to concurrently manage several tasks, such as GNSS positioning and image processing, within the confines of its limited computational resources. The Spresense SDK supports both Symmetric Multi□Processing (SMP) and Asymmetric Multi□Processing (AMP); in our implementation, we employed SMP to efficiently distribute tasks across available cores. For example, one core is dedicated to flight controller communication while others handle GNSS processing or image processing. As all cores share the same operating system, inter□core communication is straightforward. Fig. 4 demonstrates the task distribution strategy, including how unused cores are put into sleep mode to conserve power without sacrificing performance.

4. HITL Evaluation

We evaluated the performance, power efficiency, and versatility of our proposed companion computer using a Hardware-in-the-Loop (HITL) setup. All tests employed Pixhawk 6× [22] as a flight controller.

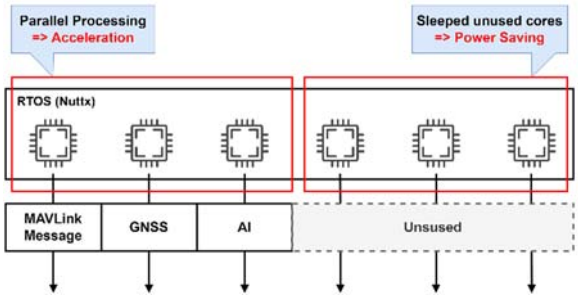


Fig. 4. Task Distribution for Multi-Core Utilization.

4.1. Communication Latency

We measured the Round-Trip Time (RTT) to assess the communication latency among Spresense, the Raspberry Pi 4 Model B, and a PC. The PC utilized a USB serial converter (DSD Tech SH-U09C5), whereas the other platforms used direct serial communication.

Fig. 5 illustrates the sequence diagram used for measuring RTT. In this test, RTT is defined as the time

elapsed from sending data from the companion computer to the flight controller and receiving the returned data from the flight controller. All UARTs were configured at 921600 baud, 8N1, with no flow control. The flight controller transmitted acknowledge (ACK) messages at 50 Hz, while each companion computer sent synchronous (SYN) messages at frequencies of 1 Hz, 10 Hz, and 100 Hz and processed ACKs at 1000 Hz.

Fig. 6(A) shows the RTT distributions for the PC, the Raspberry Pi 4 Model B, and Spresense. At 10 Hz, the PC exhibited an average RTT of 35.05 ms (standard deviation (SD) = 3.97 ms), the Raspberry Pi 4 Model B 30.86 ms (SD = 1.72 ms), and Spresense 20.85 ms (SD = 0.21 ms). These results indicate that Spresense achieves a 32 % reduction in average RTT and a markedly lower variability compared to the Raspberry Pi 4 Model B, while consuming only 1/18 of its power. Based on these findings, we conclude that Spresense offers a more stable and power-efficient performance for time-critical UAV operations. Fig. 6(B) shows the RTT distributions on Spresense when varying the baud rate. Four baud rates were tested: 57600, 460800, 921600, and 2000000, while keeping all other conditions constant. The results indicate that for SYN frequencies of 1 Hz and 10 Hz, no RTT losses occur regardless of the baud rate. However, at SYN frequencies of 100 Hz and 1000 Hz, RTT losses begin to emerge. At a baud rate of 57600, communication fails when the SYN frequency is set to 100 Hz or 1000 Hz. Moreover, ACK reception losses are observed at these higher SYN frequencies regardless of the baud rate. Therefore, it is recommended that the baud rate be set between 460800 and 2000000 for reliable communication.

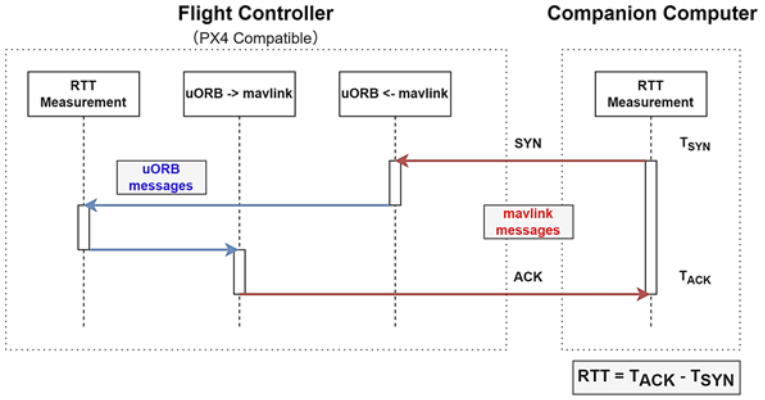


Fig. 5. Sequence Diagram of RTT (Round Trip Time).

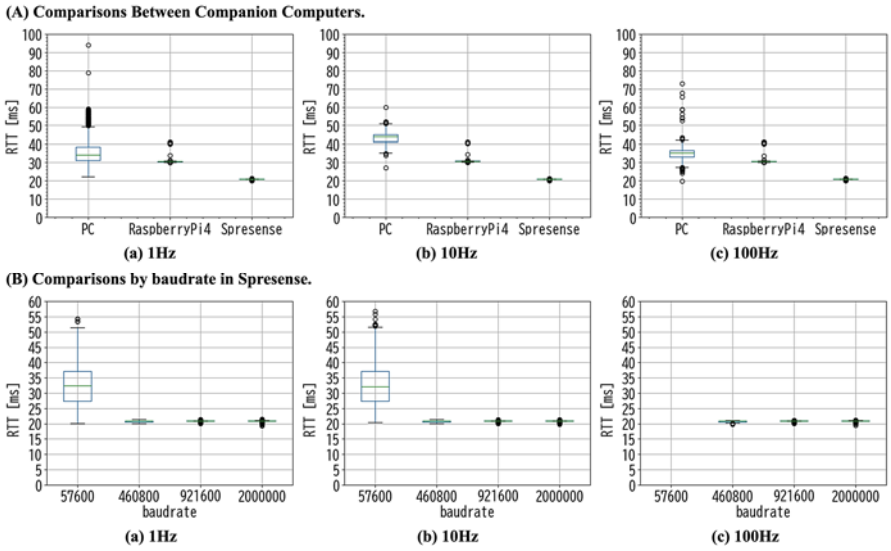


Fig. 6. RTT Results (A: Device Comparison, B: Baud Rate Variation).

4.2. Versatility and Power Consumption

We evaluated the versatility and power consumption of the developed companion computer by creating three applications – an Operation App, a GNSS App, and an AI App – on both Spresense and the Raspberry Pi 4 Model B.

Operation App: sends basic flight commands (takeoff, altitude hold, landing).

GNSS App: acquires real-time location data and sends it to the flight controller via MAVLink.

AI App: performs real-time image recognition by processing camera inputs with a pre-trained machine learning model.

Table 2 shows the measured power consumption in our HITL environment. Spresense consistently consumed about 1/18 the power of the Raspberry Pi 4 Model B while retaining comparable functionality.

Table 2. Power Consumption Comparison.

Application	Raspberry Pi 4 Model B [mW]	Spresense[mW]
Standby	3088	40
Operation App	3178	50
GNSS App	3219	88
AI App	3549	510

5. Real UAV Evaluation and Analysis

We validated the effectiveness of our proposed method using a real Nano UAV, the Nano Mind 110 [23]. Nano Mind 110 is a Nano UAV featuring a 110 mm diagonal body weight of 36 g, and a maximum take-off weight of 80 g. It is equipped with the PX4-compatible flight controller, MindRacer [24].

5.1. Experimental Environment

Fig. 7 illustrates the experimental setup for real□world testing. Instead of a 3.7 V lithium polymer battery, we used a DC stabilized power supply set to 4.2 V to emulate a fully charged battery. A weighing scale was employed to estimate thrust by monitoring changes in its readings. Additionally, to minimize the effects of ground effect [25], all experiments were conducted in an environment free of obstacles beneath the UAV.

We evaluated the relationship between power consumption and thrust for the Nano Mind 110. Fig. 8 shows the measurement results. In our experiment using a DC power supply, we were only able to measure thrust up to 25 g.

5.2. Power Consumption Evaluation

However, given that the torque and power consumption of a DC motor are linearly related, we extrapolated the linear approximation curve to estimate the relationship up to the maximum thrust of 80 g. The derived linear fit is given by:

$$\text{Power}[\text{mW}] = 386 \cdot \text{Thrust}[\text{g}] + 1811 \quad (1)$$

To verify the validity of this approximation, note that the Nano Mind 110 has a maximum payload of 80 g. Assuming it is equipped with four motors, and referencing the specifications of motors of the same size [26], the maximum power consumption is calculated as:

$$35360[\text{mW}] = 4 \cdot 3.4\,[\text{V}] \cdot 2600[\text{mA}] \quad (2)$$

From the approximation curve, the estimated power consumption at a thrust of 80 g is

$$32691[\text{mW}] = 386 \cdot 80[\text{g}] + 1811 \quad (3)$$

Since these two values are close, the estimation of the relationship between thrust and power consumption is considered valid. Moreover, since the drive system consumes almost no power at 0 g thrust, the baseline power consumption of 1811 mW can be attributed to the flight controller.

5.3. Power Consumption Analysis

Based on Equation (1), we estimated the power consumption of the Nano Mind 110 in a hovering state. With a body weight of 36 g, the hover power consumption is calculated as:

$$15707[\text{mW}] = 386 \cdot 36 + 1811 \quad (4)$$

Next, consider equipping the Nano Mind 110 with Spresense as a companion computer while running an AI application. In this configuration, Spresense consumes 510 mW and adds 7 g to the weight, resulting in a total weight of 43 g. The hover power consumption is then estimated as:

$$18919[\text{mW}] = 386 \cdot 43[\text{g}] + 1811 + 510 \quad (5)$$

Thus, integrating Spresense increases power consumption by approximately 20 %. In contrast, when equipping the Nano Mind 110 with the Raspberry Pi 4 Model B running a samapplication, the Raspberry Pi 4 Model B consumes 3549 mW and weighs about 65 g, yielding a total weight of 104 g. The estimated hover power consumption becomes:

$$45504[\text{mW}] = 386 \cdot 104[\text{g}] + 1811 + 3549 \quad (6)$$

This represents an increase of approximately 190 % relative to the base Nano Mind 110. Moreover, the total weight of 104 g exceeds the maximum take-off weight of 80 g, rendering the Raspberry Pi 4 Model B unsuitable for this application.

Given that the flight time of the Nano Mind 110 is approximately 7 minutes, the hover flight time is estimated to be around 5.8 minutes with Spresense and only about 2.4 minutes with Raspberry Pi 4 Model B. These results indicate that integrating Spresense results in a much smaller increase in power consumption – and consequently a less severe impact

on flight time – compared to using Raspberry Pi 4 Model B.

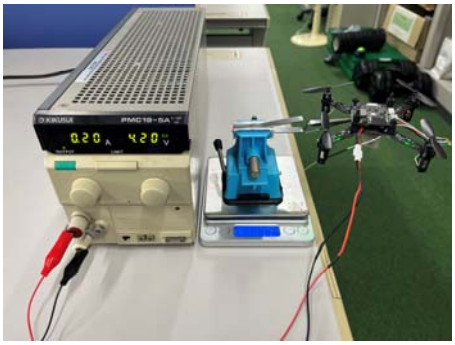


Fig. 7. Experimental Setup for Real□world Testing.

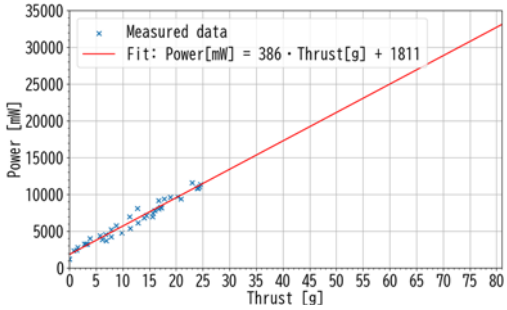


Fig. 8. Relationship between Thrust and Power Consumption.

6. Comparison with State-of-the-art

Table 3 shows a comparison between our proposed method and three state-of-the-art related studies or products.

J. Cheng et al. [7]: Propose an FPGA-based companion computer notable for its low power consumption. However, due to inherent circuit constraints in FPGAs, performing complex computations is challenging, and the system remains in the research stage, making practical deployment difficult.

AI-deck [9]: COTS microcontroller board developed by Bitcraze. AI-deck adapts GAP8 processer by GreenWaves Technologies. GAP8 itself supports the POSIX-compatible RTOS NuttX; however, AI-deck is limited to PULP OS and FreeRTOS, complicating the reuse of Linux-based development assets. Furthermore, in GAP8, only one core – the Fabric Controller Core – can access peripheral devices, while the other cores function solely as accelerators for computation. This design significantly limits its extensibility and versatility for complex IoT applications.

Shaheen [11]: Supports Linux in addition to an RTOS. Nonetheless, Shaheen is only a prototype and has not been released as a commercial product. Furthermore, while the S oC's power consumption is mentioned, the overall board power consumption and extensibility remain unclear, making its effectiveness as a companion computer for Nano UAVs uncertain.

Table 3. Comparison with State-of-the-art Companion Computers.

	J. Cheng et al. [7]	AI-deck [9]	Shaheen [11]	Spresense (Ours) [13]
Availability	Prototype	Product	Prototype	Product
Category	FPGA	SoC	SoC	SoC
Application Processor	-	RI5CY	CVA6	6× Cortex-M4
Accelerator	XC6SLX9	8× RI5CY	8× FLEX-V	Cortex-M4 2D Graphic Accelerator
Max Freq. [MHz]	-	250	500 – 600	156
Memory	-	1.5 MB SRAM + 8-64 MB HyperBUS	1 MB SRAM + 32-512 MB HyperBUS	1.5 MB SRAM + 8 MB Flash
OS	-	RTOS	RTOS + GPOS	RTOS
Power [mW]	≦ 50	60	200	40

Our Companion Computer: Overcomes all the issues found in these studies. Since Spresense is marketed by Sony, it is readily available. It supports NuttX thus allowing reuse of Linux-based development assets. Moreover, official and third-party add-on boards (e.g., for cameras, Wi-Fi, and LTE, etc.) are available to extend its functionality. In terms of power consumption, it achieves levels comparable to J. Cheng et al.'s FPGA-based companion computer, which boasts the lowest power consumption among the compared systems. Additionally, Spresense offers a significant cost advantage, priced at $65.00 compared to AI-deck's $240.00, making it an ideal choice for budget-sensitive applications without sacrificing performance or extensibility.

7. Conclusion

In this study, we developed a lightweight, low-power, and versatile companion computer for Nano UAVs based on the Spresense platform, offering

balanced performance, straightforward development with NuttX, and robust extensibility. HITL evaluations show that, compared to Raspberry Pi 4 Model B, Spresense achieves approximately 30 % reduction in round-trip time (with the standard deviation reduced to 1/10) while consuming only 1/18 of the power.

Moreover, the real□UAV evaluations confirmed that integrating Spresense into the Nano Mind 110 results in only a modest increase in power consumption and hover power requirements – ensuring a relatively minor impact on flight time. In contrast, employing Raspberry Pi 4 Model B would dramatically increase power consumption and weight, rendering it impractical for Nano UAV applications.

Notably, our proposed method demonstrates a unique integration of availability, extensibility, and ease of development – features that previous studies could not simultaneously achieve. Future work will focus on real-world validation – indoor navigation, outdoor obstacle avoidance, and ROS 2–based applications using micro-ROS – to confirm its robustness and scalability.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 24KK0184.

References

[1]. A. Kushleyev, D. Mellinger, C. Powers, V. Kumar, Towards a swarm of agile micro quadrotors, *Autonomous Robots*, Vol. 35, Issue 4, 2013, pp. 287-300.

[2]. Yamaha Motor, Hybrid Drone Engine, https://global.yamaha-motor.com/np/hybrid-drone/

[3]. P. K. Chittoor, B. Chokkalingam, L. Mihet-Popa, A review on UAV wireless charging: fundamentals, applications, charging techniques and standards, *IEEE Access*, Vol. 9, 2021, pp. 69235-69266.

[4]. Dronecode Project, https://dronecode.org/

[5]. F. Ahmed, M. Jenihhin, A survey on UAV computing platforms: a hardware reliability perspective, *Sensors*, Vol. 22, Issue 16, 2022, 6286.

[6]. R. Okudera, Y. Li, Y. Matsubara, H. Takada, A lightweight, low-power and versatile companion computer for nano UAVs, in *Proceedings of the 1st International Conference on Drones and Unmanned Systems (DAUS'25)*, 2025, pp. 292-295.

[7]. J. Cheng, Y. Yamaguchi, A study of FPGA-based micro-drone controller, https://www.ieice.org/publications/conference-FIT-DVDs/FIT2019/data/pdf/CC-003.pdf (in Japanese)

[8]. Bitcraze, Crazyflie 2.1 Product Page, https://www.bitcraze.io/products/old-products/crazyflie-2-1/

[9]. AMD, Spartan-6 Tech Specs, https://www.amd.com/en/products/adaptive-socs-and-fpgas/fpga/spartan-6.html

[10]. GreenWaves Technologies, GAP8 Low Power Processor, https://greenwaves-technologies.com/low-power-processor/

[11]. Bitcraze, AI-Deck Product Page, https://www.bitcraze.io/products/ai-deck/

[12]. PULP Platform, https://pulp-platform.org/

[13]. L. Valente, A. Nadalini, A. H. C. Veeran, M. Sinigaglia, et al., A heterogeneous RISC-V based SoC for secure nano-UAV navigation, *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 71, Issue 5, 2024, pp. 2266-2279.

[14]. Sony Developer World, Spresense, https://developer.sony.com/spresense

[15]. Takara Tomy, SORA-Q Product Page, https://www.takaratomy.co.jp/products/sora-q/

[16]. Raspberry Pi Zero W Product Page, https://www.raspberrypi.com/products/raspberry-pi-zero-w/

[17]. Arduino Docs, Nano Hardware Page, https://docs.arduino.cc/hardware/nano/

[18]. MAVLink Developer Guide, https://mavlink.io/en/

[19]. Micro-ROS, https://micro.ros.org/

[20]. MAVSDK Documentation, https://mavsdk.mavlink.io/main/en/index.html

[21]. MAVSDK Documentation, FAQ, https://mavsdk.mavlink.io/main/en/faq.html

[22]. PX4 Autopilot User Guide, Holybro Pixhawk 6X, https://docs.px4.io/main/en/flight_controller/pixhawk6x.html

[23]. PX4 Autopilot User Guide, Nano Mind 110, https://docs.px4.io/main/en/complete_vehicles_mc/NanoMind110.html

[24]. PX4 Autopilot User Guide, MindRacer, https://docs.px4.io/main/en/flight_controller/mindracer.html#mindracer-hardware

[25]. A. Matus-Vargas, G. Rodriguez-Gomez, J. Martinez-Carranza, Ground effect on rotorcraft unmanned aerial vehicles: a review, *Intelligent Service Robotics*, Vol. 14, Issue 1, 2021, pp. 99-118.

[26]. Bitcraze Store, 4×7×20 mm DC motor pack, https://store.bitcraze.io/products/4-x-7x20-mm-dc-motor-pack-for-crazyflie-2-x-thrust-upgrade

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 45-58

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

A Hybrid Machine Learning and Physics-based Approach for Accurate Energy Consumption Modeling of Electric Buses in Public Transport

[1,2,3 *] Lucas ADAM, [1,3,4] Robert PELLERIN, and [1,2,3] Bruno AGARD

[1] Polytechnique Montréal - Department of Mathematical and Industrial Engineering, Canada
[2] LID, Data Intelligence Laboratory, Polytechnique Montréal, Canada
[3] CIRRELT - Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, Canada
[4] Jarislowsky/AtkinsRéalis Research Chair in the Management of International Projects, Canada
[*] E-mail: lucas.adam@polymtl.ca

Received: 12 Nov. 2025 /Revised:31March 2025 /Accepted:15 April 2025 /Published:30 April 2025

Abstract: Public transport organizations are increasingly concerned about reducing air pollution, leading many to transition their fleets into electric vehicles (EVs). In this context, limited battery range and charging times remain significant hurdles. Precise modeling of electric bus energy consumption is crucial. Still, existing methods often face difficulties due to the complexities of real-world conditions, such as diverse driving patterns and external factors. To tackle this, the study proposes a hybrid model combining physical principles and machine learning using real-world data from 30 buses across 130 routes over one year. Key variables like passenger load, weather, and route characteristics are incorporated. Several machine learning models, including MLP, KAN, and XGBoost, are compared using Mean Absolute Percentage Error (MAPE). The hybrid model outperforms others, achieving a low MAPE of 5.59 % on test data and 5.79 % on validation data with a low Standard Deviation. Additionally, models incorporating operational factors, such as bus lines and time of day, enhance prediction accuracy. The study concludes that integrating physical laws with machine learning offers a more accurate and stable approach to energy consumption modeling, providing a promising framework for fleet management and energy efficiency in public transport systems.

Keywords: Public transport, Electric bus, Telemetry, Energy consumption, Big data, Artificial intelligence.

1. Introduction

As consciousness arises about the role of road transport in air pollution, the global trend points towards the use of more energy-efficient vehicles and the reduction of greenhouse gas emissions. According to Krawiec et al. [2], this affects not only personal vehicles but also public transport vehicles.

The share of electric-powered buses operated by public transport organizations is expected to reach 100 % by 2050. Despite the significant advantages associated with electric vehicles, such as zero tailpipe emissions [3] and reduced noise pollution, which contributes to improved overall life quality in urban areas [4], these innovative solutions come with their challenges [3]. Some of the most critical issues are their limited battery range and longer charging times [5]. This highlights the crucial need for accurately modeling discharge patterns in electric vehicles [1].

While many researchers have proposed approaches based on simulated data [5], these methods often fall short in real-world applications, where "energy consumption prediction is challenging due to complicated driving cycles and a wide range of influential factors" [6]. Modeling electric energy consumption remains a complex challenge [7].

Numerous approaches have been proposed in the scientific literature, each with distinct advantages and limitations. Rule-based models use fundamental physical principles to simulate vehicle-environment interactions, incorporating factors such as rolling resistance, aerodynamic drag, road gradient, and acceleration forces. These models offer a mechanistic understanding of energy consumption. In contrast, data-driven models uncover hidden relationships between variables using historical data, offering a more empirical approach [6].

This article proposes a different approach by combining physical and machine learning approaches to predict the required energy consumption of electric buses per trip. As such, the paper's main objective is to provide a method to incorporate telemetry and planning data into prediction models that can make energy predictions at fleet scale, considering the specifics of multiple bus lines and vehicles.

The remainder of the paper is structured as follows. First, the following section presents the literature review. The proposed models and the data used in this study are then presented. The performance of the proposed models is then evaluated, followed by an analysis of variable influences. Finally, limitations and future works are discussed in the discussion and conclusion sections.

2. Literature Review

The literature identifies a broad range of parameters influencing energy consumption in electric vehicles. Physical factors such as speed, acceleration, weather conditions, ambient temperature, vehicle mass, and road slope are frequently cited as key determinants [7]. Data-driven models, however, expand upon these by incorporating additional variables, including stop frequency, traffic conditions, trip timing, and in-vehicle climate control usage, offering a more nuanced understanding of real-world energy consumption [7]. These models can be categorized into three main types [8].

The first category includes statistical models, such as linear regression and ordinary least squares. While effective for small datasets, they assume linear relationships that may not fully capture complex dependencies. The second category consists of machine learning models, including support vector regression, XGBoost, and Random Forest. These models handle intricate, non-linear relationships and automatically identify relevant features from large datasets. However, they require extensive feature engineering and struggle with highly unstructured data. The third category comprises neural network models, such as deep neural networks, Long Short-Term Memory (LSTM) networks, and Transformers. These excel at modeling time-series data and capturing long-term dependencies, but they demand significant computational resources and large datasets for effective training.

Several studies advocate for hybrid models combining physical and data-driven approaches to leverage the strengths of both paradigms [8]. By integrating physical laws with machine learning insights, such models enhance accuracy and generalization. Despite the growing interest in electric vehicle energy modeling, relatively few studies focus specifically on electric buses [2]. Most recent work, particularly those utilizing real-world datasets, centers on passenger cars [9]. While valuable, these approaches often fail to account for the unique challenges of public transport modeling, such as cyclical usage patterns and significant variations in vehicle mass due to passenger loading [10].

Beckers et al. [11] proposed a physical modeling approach to quantify the impact of friction forces due to cornering on energy consumption. They found that cornering losses account for approximately 3 % of propulsion system energy consumption. While this study offers valuable insights into specific aspects of energy loss, such as friction during turns, it does not address broader operational variables like passenger load and weather conditions, which can also significantly affect energy consumption.

In contrast, Fiori et al. [12] adopt a more comprehensive decomposition of power consumption, separating it into two components: motion-related and auxiliary consumption. The motion-related component, modeled using physical principles and calibrated with data from 435 trips, is validated with an impressive RMSPE of 2.56 %. In addition, the authors also consider auxiliary consumption, such as HVAC systems, by modeling it as a polynomial function of outdoor temperature. While this approach is more holistic, accounting for factors like HVAC energy consumption, it is limited by the need for separate calibration per vehicle and possibly per line, which makes it impractical for large-scale, real-time operational use.

Furthermore, Beckers et al. [13] take a similar physics-based approach to model energy consumption, focusing on factors like motor efficiency and rolling resistance based on road surface conditions. This study also emphasizes calibration through real-world testing campaigns. However, it struggles with incorporating regenerative braking effects, a critical aspect of modern electric vehicles that could influence energy recovery and overall efficiency. This limitation further highlights a common challenge across physics-based models. While they provide valuable insights into specific components of energy use, they often fail to capture the full complexity of vehicle operation, particularly in terms of energy recovery mechanisms like regenerative braking.

Blades et al. [14] addressed the challenge of generating realistic drive cycles, which is often limited by the cost and complexity of real-world data collection. By developing a generic Battery Electric Bus (BEB) model and applying machine learning techniques such as linear regression, LASSO, ridge regression, Random Forest, and Extra Trees, they

identified the average road gradient and speed as the most influential factors affecting energy consumption. The study reported MAPE errors ranging from 2.1 % to 10.67 % and recommended including additional factors such as passenger load and weather conditions – factors that are explicitly considered in our own approach. This aligns with the growing recognition that contextual factors beyond the road and vehicle characteristics play a crucial role in predicting energy consumption.

Similarly, Basso et al. [15] also utilized machine learning techniques to predict energy consumption, focusing on multiple buses and a range of time aggregation levels (5, 10, 15, and 30 minutes). Their use of 30 variables, including a physics-based consumption estimate, and their application of models like linear regression, Random Forest, SVR, and neural networks resulted in a best-performing model (linear kernel SVR) with a MAPE of 22 %, a significant improvement over the physics-based model's MAPE of 43\%. The study underscored the importance of long-term telemetry data and multi-route contextual variables, which are central to our research. While their approach was more data-intensive, it highlights the necessity of incorporating diverse contextual factors to improve prediction accuracy.

Meanwhile, Pamula and Pamula [16] took a deeper dive into deep learning approaches, specifically comparing LSTM, Deep Learning with Autoencoder (DLNA), and Multilayer Perceptron (MLP) models on two bus routes. They found that DLNA outperformed the other models with a MAPE of 6.2 %, followed closely by MLP (6.7 %) and LSTM (7.2 %). Despite deep learning's promise in sequential energy prediction, their study was constrained by a limited sample size and a narrow range of routes.

Xu et al. [17] partly tackled this issue by analyzing a larger dataset of a year's worth of data from ten buses operating on a single route. They compared Elastic Net, GPR, and LightGBM models, with the best-performing models achieving MAPEs of 9.17 %, 7.54 %, and 7.71 %, respectively. Elastic Net was noted for its generalizability, and the study emphasized the significant impact of external temperature, driver behavior, vehicle characteristics, and road attributes on energy consumption. These findings are consistent with the notion that operational factors beyond the vehicle itself – such as weather conditions and driver behavior – are essential to understanding and predicting energy use. However, the study also pointed out the frequent omission of passenger load variation.

Finally, Sennefelder et al. [18] analyzed driver behavior across 30 diesel buses and extrapolated the findings to electric buses. After filtering the data, they were left with 149 trips from 24 vehicles, highlighting the inherent challenges of real-world data collection. Using the Neighborhood Component Feature Selection (NCFS) algorithm, they classified energy consumption into three levels and found that Random Forest achieved the highest accuracy at 83 %. This study reinforced the complexities of real-world data acquisition and the importance of robust filtering techniques when dealing with large and noisy datasets.

By comparing these studies, we can see the evolution of data-driven approaches in predicting energy consumption for electric buses. Early studies, such as Pamula and Pamula [16], focused more on identifying key influencing factors. As the field progressed, studies like Basso et al. [15] and Blades et al. [14] incorporated more complex models and larger datasets, addressing some of the limitations of earlier approaches.

Table 1 summarizes the key variables used in different studies related to bus energy consumption prediction. It highlights the specific factors considered by the key studies we have identified. By examining these variables, we can better understand the scope and limitations of previous research and refine our approach to predicting bus energy consumption. Table 2 summarizes the key characteristics of the highlighted studies on energy consumption prediction for electric buses. This table highlights the context variables such as the number of vehicles, model types, temperature ranges, data types, and study durations. By comparing these attributes, we gain insight into the diverse approaches and setups used in previous work and identify potential gaps or areas for further exploration in our research.

The limited focus on electric buses is often attributed to challenges in data availability, as proprietary datasets or small sample sizes (such as [12, 13, 15]) hinder generalizable conclusions [19]. Many studies analyze a restricted number of bus routes [12, 15, 16] or small fleets [11-15], limiting the applicability of findings. This limitation in route diversity is a challenge we also aim to address in our work, as broader route diversity can significantly affect model generalization and robustness.

To the best of our knowledge, existing models in the literature do not directly incorporate variables such as vehicle number, route number, and direction. We believe that including these variables could significantly enhance machine learning models by helping to identify cyclical patterns or consumption behaviors that vary by bus, ultimately improving predictions. Furthermore, these pieces of information are easily accessible from scheduling software, allowing for seamless integration by transport companies.

3. Proposed Model

To overcome the limitations mentioned, we seek to leverage an extensive real-world dataset from 30 buses operating on over 130 routes over an entire year, enabling us to identify seasonal influences on energy consumption predictions. Our approach incorporates a diverse range of influential variables – including, speed, acceleration, passenger load, temperature, vehicle identification, and route attributes – offering a robust and practical framework for energy

consumption prediction in electric bus fleets. Additionally, our industrial partnership provides access to buses equipped with advanced telemetry systems, ensuring reliable data acquisition. This collaboration allows us to address key gaps in the literature and refine our understanding of electric bus energy consumption.

Table 1. Summary of Variables Used in Previous Studies.

	Physical Model			**Hybrid**	**Machine Learning**			
Article	[13]	[11]	[12]	[15]	[14]	[16]	[17]	[18]
Speed	×	×	×	×	×		×	×
Acceleration	×	×	×	×	×		×	×
Passenger Load		×	×	×	×			
Temperature			×	×			×	×
Slope	×			×	×			
Weather Conditions						×		
Distance			×	×	×	×		
Travel Time					×	×		
Number of Stops					×			
Elevation					×	×	×	
Time of the Day					×		×	
Trip Number			×		×			
Day of Week							×	
Cornering Angle	×							

Table 2. Summary of Study General Context. Phy.: Physical model; ML: Machine Learning model; Aux.: Auxiliary model; Sim.: Simulated.

Article	[13]	[11]	[12]	[15]	[14]	[16]	[17]	[18]
Number of vehicles	1	1	1	6	1		10	30
Number of lines			1	1		2	1	
Model Type	Phy.	Phy.	Phy. and Aux.	Hybrid	ML.	ML.	ML.	ML.
Max Temperature (°C)			17	17	20	24		
Min Temperature (°C)			-15	11	1	-5		
Data Type	Real	Real	Real	Real	Sim.	Real	Real	Sim.
Study Duration	1 trip	62 hours	2 months	1 day		9 months	1 year	11 days
City			Howald	Santiago		Jaworzno	Beijing	Seville
Country			Luxembourg	Chile		Poland	China	Spain

We first introduce the physical model found in the literature. Then we detail how some forces will be grouped to create consumption variables for the hybrid model. Finally, we detail which machine learning approaches will be used to create the hybrid models that will be compared in the next sections.

The literature categorizes as "Rule-based" [19] the model used to assess energy consumption defined by equations (1) to (6):

$$P_{\text{total}} = \frac{kv_t}{\eta_m \eta_t}\left(F_{rr} + F_d + F_g + F_a\right), \quad (1)$$

$$F_{rr} = C_r mg \cos \alpha_t, \quad (2)$$

$$F_d = \frac{\rho_a}{2} C_d A_f v_t^2, \quad (3)$$

$$F_g = mg \sin \alpha_t, \quad (4)$$

$$F_a = m\delta a_t, \quad (5)$$

$$E_{\text{battery}} = \int_t^T P_{\text{total}}, \quad (6)$$

where P_{total} is the power provided by the battery at time t. F_{rr}, F_d, F_g and F_a are respectively forces due to rolling resistance, drag, road gradient, and acceleration. v_t, a_t and α_t are the velocity, acceleration, and road slope at time t. The constants used are C_r and C_d, rolling resistance and drag coefficients, ρ_a air density, A_f the bus front surface, m vehicle mass, g gravitational acceleration, rotational inertia, η_m and η_t motor and transmission efficiencies, whose values are derived from [15]. The regeneration factor k is varying from 0 to values close to 1 based on the technology used. Integrating those elements for a trip leads to $E_{battery}$, the total energy consumption during the considered trip.

This model has several possible variations, depending on the available data. The details of these physical model variants are summarized in Table 3, while the architecture of each model is illustrated in Fig. 1. Model A is used when the data of speed, load, slope, Heating, Ventilation, and Air Conditioning (HVAC) power are available and also when the driver utilizes regenerative braking. Model B applies when

HVAC data is unavailable. Model C is used when energy generation is deactivated due to unsafe driving conditions, such as slippery road surfaces during freezing or snowy weather. Model D is employed when only the speed profile is available.

Table 3. Summary of the different models and their associated parameters.

Model	Available Data	Load	Slope
A	Speed, Acceleration, Slope, Load, HVAC	Variable	Variable
B	Speed, Acceleration, Slope, Load	Variable	Variable
C	Speed, Acceleration, Slope, Load	Variable	Variable
D	Speed, Acceleration	Empty Weight	0

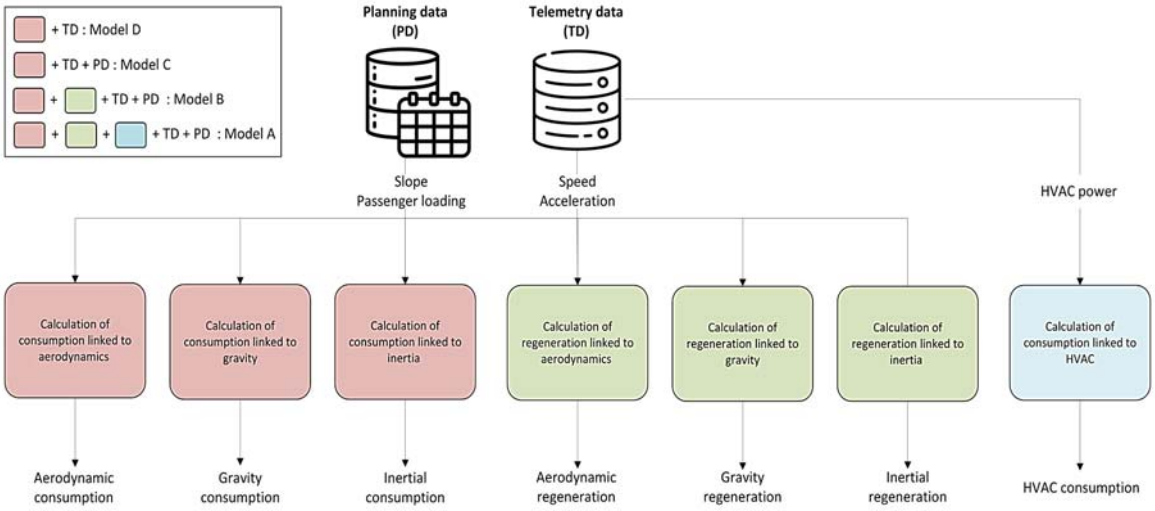


Fig. 1. Architecture of physical models (A, B, C, D).

We propose addressing the limitation of conventional physical models with a physical model coupled with a machine learning model. The physical model provides the consumption linked to aerodynamics, gravity, and inertia as well as their regeneration contributions. Instead of adding the energy of every force, we split them into three power groups described by equations 7 to 9. To calculate the energy consumption of a contribution (gravity, aerodynamic or inertia) during the trip, equation 10 is used. If the battery power is positive, then k = 1 and consumption contributions are calculated, if the battery power is negative, k = 0.3 and regenerative contributions are calculated.

$$P_{gravity} = \frac{kv_t}{\eta_m \eta_t}(C_r mg\cos\alpha_t + \\ +mg\ \sin\alpha_t), \quad (7)$$

$$P_{aerodynamic} = \frac{kv_t}{\eta_m \eta_t}(\frac{\rho_a}{2} C_d A_f v_t^2), \quad (8)$$

$$P_{inertia} = \frac{kv_t}{\eta_m \eta_t}(m\delta a_t), \quad (9)$$

$$E_{\text{contribution}} = \int_t^T P_{\text{contribution}} \quad (10)$$

From a physical point of view, the rolling resistance and aerodynamic drag cannot regenerate energy. In our case, "aerodynamic regeneration" or "gravity regeneration" is their consumption during the regeneration phase. After multiple tests, constructing these variables allows the following machine learning to better predict the energy consumption than if each force and its related consumption or regeneration is considered independently.

Based on the literature review, we identified that the Multi-Layer Perceptron (MLP) is a common machine learning approach, so we chose to compare its performance. Since its release in April 2024, the article by Liu et al. [21] has sparked the interest of many machine learning researchers due to its potential to replace architectures using dense layers from MLPs. Also, ensemble methods like XGBoost appeared promising for their ability to represent complex relationships between data [8].

In the first model, M1, the energy contributions are used directly as inputs to a Multilayer Perceptron (MLP). In the second model M2, the vehicle identification number, bus line number, direction, day of the week, time of the day, and trip distance are added to provide for each prediction and adjust the forecast accordingly. The M3 model uses the variables

used by M2 with two new variables, average energy and average speed. The M4 Model uses a Kolmogorov-Arnold Network (KAN) while the M5 model uses an XGBoost model with the same variables as M3. The details of those hybrid models are given in Table 4. The overall process is illustrated in Fig. 2. We will investigate the influence of the data available on the accuracy of prediction using these model variants. Less data available is expected to lead to worse performance of the associated predictions.

Table 4. Summary of the different models and their associated parameters.

Model	Model Architecture	Context
A, B, C, D	Physical	None
M1	Physical followed by MLP	None
M2	Physical followed by MLP	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance
M3	Physical followed by MLP	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed
M4	Physical followed by KAN	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed
M5	Physical followed by XGBoost	Vehicle identification number, Bus line number, Direction, Day of the week, Time of the day, Trip distance, Mean Temperature, Mean Speed

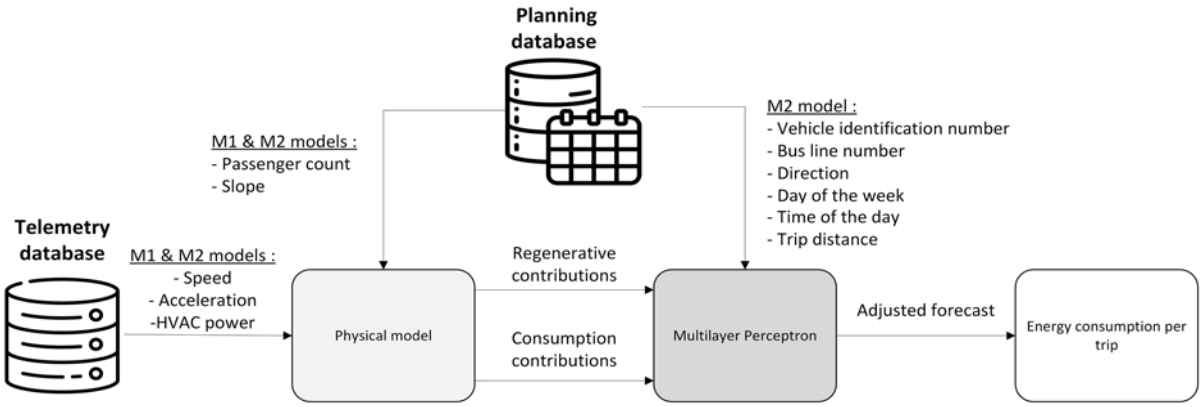


Fig. 2. Architecture of the proposed models (M1 and M2).

4. Data Description

As previously mentioned, the data used for this study comes from the operational databases of a transport company. These data are not derived from simulations or controlled test campaigns but from the real-world use of these vehicles. This case study encompasses nearly 100000 trips across 30 electric buses, covering over 25000 work cycles on nearly 130 different bus routes over one year. Each work cycle consists of multiple trips on various lines.

The telemetry system implemented is designed to monitor vital signs such as speed, heating power, and the number of active battery packs. It is primarily used to send alerts and error messages for maintenance purposes. Additionally, variables such as acceleration, slope, or passenger count are not directly measured in vehicles but must be reconstructed by merging data from different sources. A detailed list of the variables considered is provided in Table 5.

We have chosen to conduct an initial exploratory study to highlight the specific characteristics of this application case. The driving speed varies from 0 to 90 km/h, which is quite common in a large city, with the higher end occurring when the bus drives on interurban portions. The mean acceleration of 0.19 m/s² corresponds to a very gentle acceleration, which maximizes passenger comfort. The higher value of 0.77 m/s² is more noticeable for passengers but remains safe for carrying standing passengers. The slope mean is approximately 0, highlighting the cyclical behaviors of most trips. When a bus faces a positive slope during its trip, it also faces a downward slope later. The maximum and mean values of 1.5 and -1.47 rad, respectively, are unrealistic, representing slopes close to 90 degrees. After further investigation of the values, we found that the altitude gathered from GPS positioning is inaccurate for public uses, as it does not have high spatial resolution. This is the kind of noise our models must correct to provide the best possible forecasts.

This distance varies from 1 km to 39 km, with a mean value of 9.7 km. This highlights the significant variability of the routes we are trying to forecast energy consumption. It shows that when real operational data is used, it is often far from the simulations and idealized cases used in previous literature studies, justifying further studies on

real-world datasets. The temperature varies from -14 °C to 44 °C, with a mean of 15 °C, highlighting the various environmental conditions these buses must face. It also covers a broader range of temperatures than most literature studies. It is worth noting that these values are subject to inconsistencies as they are measured directly by bus sensors. These measurements might be higher in winter than the actual exterior temperature because the vehicle temperature heats the sensor due to heat propagation. In the summer, if exposed directly to sunlight, the rays might heat the sensor, again disrupting its measurements. We have chosen to use bus temperature sensors rather than those of meteorological stations because they could better represent micro-climates. The average duration of a trip is 32 minutes, which is relatively short for public transport. However, the maximum value of 240 minutes (4 hours) indicates the presence of exceptional situations, such as breakdowns. During these trips, it is likely that the vehicle's energy consumption is not truly representative of nominal usage.

Table 5. Description of Variables, Sources, Types, and Sampling Methods.

Variable	Source	Type	Sampling
Speed	Telemetry system	Float	Triggered
Acceleration	Calculated from speed	Float	Per second
Slope	Calculated from GPS altitudes and distance between stops	Float	Mean between planned stops
Distance	Calculated from speed	Float	Per second
Temperature	Telemetry system	Float	Triggered
Duration	Planning database	Float	Trip
Passenger count	Planning database	Integer	Between planned stops
Energy consumption	Telemetry system	Float	Triggered
Vehicle identification number	Planning database	Categorical	Trip
Bus line number	Planning database	Categorical	Trip
Direction	Planning database	Categorical	Trip
Day of the week	Planning database	Categorical	Trip
Time of the day	Planning database	Categorical	Trip

The average number of passengers is 21, which is reasonable for an urban bus. The maximum value of 190 passengers describes exceptionally crowded trips, such as in high-density areas or during city events attracting large crowds.

The average energy consumption is 13.7 kWh, which seems reasonable for an urban bus. However, the maximum value of 73 kWh indicates long or particularly energy-consuming trips during strong accelerations, significant fluctuations in the number of passengers, or prolonged use of air conditioning and heating systems. The negative value indicates that the buses do not consume energy on specific trips but rather generate it. While this observation may initially seem impossible, it is caused by a particular route covering only a short trip with a steep slope. When the bus goes downhill with passengers, it generates energy. When it goes uphill, it consumes a significant amount of energy to overcome gravity. This particular route is not cyclical. To observe the downhill and uphill slopes, two different directions of the route need to be considered. Therefore, in our scheduling database, we will have two different trips.

We have chosen not to work on full working days but rather on individual trips so that we can later optimize the scheduling of working days based on forecasts of individual trips. The results of this preliminary study are presented in Table 6.

Table 6. Descriptive Statistics of Variables.

Variable	Min	Max	Mean
Speed (km/h)	0	90	18
Acceleration (m/s²)	0	0.77	0.19
Slope (rad)	-1.47	1.5	0
Distance (km)	1	39	9.7
Temperature (°C)	-14	44	15
Duration (minutes)	0	240	32
Passenger count	0	190	21
Energy consumption (kWh)	-0.8	73	13.7

5. Preprocessing and Model Evaluation

To use these models, the data from the telemetry system must first be preprocessed (see Fig. 3) Only trips with a quality index higher than the threshold required by our partner are considered. The quality index, expressed as a percentage, measures the proportion of scheduled stops not served by the bus. The lower the index, the fewer planned stops were adhered to.

We then use each trip's start and end times, provided by the planning data, to extract the corresponding speed profile, battery power, and HVAC cooling power from the telemetry database.

The vehicle's empty weight and the number of passengers determine the variable load. As mentioned earlier, the slope is not directly measured; it is

calculated based on the altitude obtained from GPS coordinates and the distance between bus stops. To do this, we used GPS data tiles available from the Consortium for Spatial Information (CGIAR-CSI) [20]. These data tiles have a spatial resolution of approximately 30 meters. Since the stop position data comes from consumer-grade GPS services, they have a resolution on the order of ten meters. The altitude returned is accurate to the meter. Given that we work with horizontal precision on the order of a meter and distances between stops on the order of a hundred meters, the relative error on the slope is considered acceptable in relation to the precision of the telemetry system's measurements.

Regarding the vehicle's mass, we calculate it as the sum of the empty weight and the number of passengers on board, with an average body mass of 80 kg per passenger. The number of passengers is calculated from the boarding and disembarking at each stop. The data available indicates the number of passengers on the bus on arrival and departure from each planned stop. If passengers board or alight outside these planned stops, an algorithm assigns them to the nearest stop.

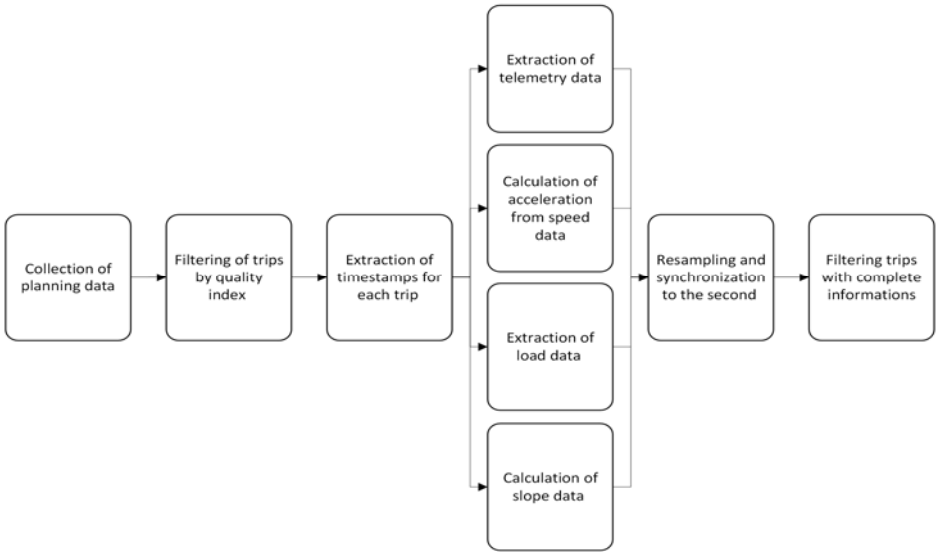


Fig. 3. Data processing from the telemetry system.

The data is then resampled to a per-second basis using forward filling. The timestamps for each measurement are not natively synchronized because each vehicle sends its telemetry data based on triggered sampling to save bandwidth. Even with these considerations, the data stored is on the order of terabytes per year for a fleet of a thousand vehicles.

During the preprocessing phase, less than 5 % of the measured trips do not meet the required quality index and are therefore removed from the analysis. This means the trips do not follow the planned path closely. For the remaining trips, 52 % allow for the complete calculation of the model (Model A) because of missing variable values. To further process data and make it useable by the hybrid models, categorical data such as vehicle identification number, line number, direction, bus number, day of the week, and time of the day have been processed using One Hot Encoding, whereas numerical values such as HVAC consumption, aerodynamic, gravity, inertia-linked consumptions, and regeneration have been standardized.

The performance of the models is evaluated by calculating the Mean Absolute Percentage Error (MAPE) between the actual and predicted energy consumption for each trip.

$$MAPE = 100 * \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i},$$

where n is the number of observations, y_i the true value, and $\hat{y}_i$ the forecasted value. We have chosen this evaluation metric because it does not depend on the observation scale, is easy to understand as a percentage, and because the absolute calculation prevents positive and negative values from canceling each other.

6. Results

This section will expose the results of our study after describing how we set up the training and evaluation framework.

6.1. Experimental Setup

This case study data is used with the models described in Table 1. To set the value of k for the physical model, we compared the total energy consumed with and without regeneration, ultimately concluding that $k = 0.3$ for our case study. We used the Optuna framework to tune the hyperparameters of

the MLP, KAN, and XGBoost models. Their hyperparameters are essential for optimizing their performance in predicting energy consumption. The error is minimized over the training set, and the best model is the one that minimizes the error over the test set. Predictions of energy consumption are finally done on the validation set to validate the generalization of the model. Our complete dataset is split between three sub-datasets: training set 70 %, test set 20 %, and validation set 10 %.

For the MLP model, the first and second hidden layers have 46 and 32 neurons, respectively, with a learning rate of 0.001. The optimizer used is Adam, and the model trains for 2500 epochs with L1Loss as the loss function.

For the KAN model, the table lists the values for the number of neurons in the first and second hidden layers (hidden size1 and hidden size2), set at 29 and 20, respectively. Other hyperparameters include a grid search parameter (grid) of 3, a k-fold cross-validation value (k) of 2, a learning rate of 0.010, the use of the Adam optimizer, 250 epochs for training, and the use of L1Loss as the loss function.

The XGBoost model parameters include a learning rate of 0.068, 210 estimators (trees), and a maximum depth of 12 for the decision trees. The model also uses a subsample rate of 0.635 and a colsample bytree value of 0.679 to control the features considered in each tree. The gamma parameter is set to 0.001 to control the regularization of the model, and the minimum child weight is set to 8 to prevent overfitting.

These model architectures are shown in Table 11 from the Appendix. The output layer of each neural network model contains a single neuron. Multiple loss functions were compared to train the model, including L1 Loss, MSE, and MAPE. These results are presented in Table 7.

Table 7. Performance of Multiple Loss Functions during Training.

Loss Function	MAPE (%)	Standard Deviation
MAPE	9.6	76.4
MSE	5.6	28.7
L1 Loss	4.8	21.5

Depending on the loss function used during the training phase, the MAPE of the training phase varies greatly. One could think that the best way to minimize the MAPE is to use it during the training phase. However, this error introduces noise, leading to convergence issues. The Mean Squared Error (MSE) penalizes the prediction errors far from the mean more significantly, proving less effective than L1 Loss during our training phase. It scores a MAPE of 4.8 %, followed by MSE and MAPE. Their standard deviations reinforce our observation with the same ranking. For the rest of our neural network models, L1 Loss will be used.

Adam is the optimizer used to train the neural network models because it allows fast convergence due to its adaptive learning rate and low memory requirements [22]. The maximum number of epochs has been set to 2500 for the MLP and 250 for KAN because they allow the models to converge based on our dataset. For physical models, it is not necessary to separate the presentation of the results since there is no training phase, but we have chosen to evaluate the performance on the same data as the machine learning models.

6.2. Models Performances

The results from the different models are presented in Table 8 for the training dataset, Table 9 for the test dataset, and Table 10 for the validation dataset.

Table 8. Model Performance Comparison on Training Dataset.

Variable	MAPE (%)	Standard Deviation
A	21.15	42.25
B	17.47	32.00
C	28.02	90.68
D	31.89	83.84
M1	12.72	48.14
M2	5.99	21.63
M3	4.78	18.13
M4	7.05	33.37
M5	3.00	17.67

The results of the models tested on the training, test, and validation datasets show a clear difference in terms of performance and stability, with M5 (Physical followed by XGBoost) dominating. This model achieves the best results across all datasets, with a very low MAPE of 3.00 % on the training set, 5.59 % on the test set, and 5.79 % on the validation set, accompanied by low standard deviations, indicating an excellent ability to generalize and provide stable predictions. This remarkable performance can be attributed to the approach combining physical data and using XGBoost. This gradient-boosting decision tree model is particularly effective for this type of task. M3 (Physical followed by MLP with additional features, including temperature and mean speed) closely follows, with a MAPE of 4.78 % on training, 6.05 % on test, and 6.30 % on validation, but its standard deviation is higher for both training and validation datasets, indicating some instability in its predictions. The M4 model, built with a KAN instead of MLP, scores a MAPE of 7.05 % on training, 7.58 % on the test, and 7.79 % on validation. It is, therefore, not more performant than a conventional MLP and requires substantially more training time than the MLP. M2 (Physical followed by MLP with basic features) presents similar results to M3, with a MAPE of 5.99 % on training and 7.78 % on test, but its standard deviation remains reasonably controlled, suggesting

good stability. However, it indicates that using the Mean speed and Mean Temperature as predictors helps improve the prediction. In contrast, A and B models show poor performance, with high MAPEs (21.15 % and 17.47 % on training, respectively) and significant standard deviations, reflecting their inability to capture complex relationships between different variables and provide reliable predictions. It is worth noting that Model B (Without HVAC consumption) provides better results than Model A (With HVAC consumption). This seems counterintuitive since one would think a more complete model would yield better results. This can be explained by the overall tendency of physical models to over-predict the power consumption to put the vehicle in motion. When HVAC consumption is added, the prediction quality increases, thus worsening the error. Model C is the second least performing model, with exceptionally high MAPEs (up to 28.02 % on training and 27.11 % on test), and very high standard deviations. Finally, the worst-performing model is D, with a MAPE of up to 33.34 % on the validation dataset and a standard deviation of 137.57. This indicates that for public transport vehicles and, more widely, vehicles with high variations in mass, it is essential to consider the energy that can be gained from regenerative braking and its loading. In conclusion, M5 clearly stands out as the model to prioritize for predicting the energy consumption of electric buses, thanks to its combination of physical data and the XGBoost algorithm. In contrast, models based solely on physical features or simpler architectures such as A, B, and C should be considered less suitable for this task and be used only when few variables are available.

6.3. Variables Contributions

To assess the influence of the various variables used in our study on the prediction output, we used SHAP (SHapley Additive exPlanations) [23]. SHAP is a widely used method that provides interpretable machine learning models by attributing the output of a model to each feature based on cooperative game theory. It calculates Shapley values, which represent the contribution of each feature to the prediction, allowing us to understand the impact of each variable on the model's output. This technique is valuable for identifying which factors most influence our predictions and for providing better transparency in the model's decision-making process. This method will be used to explain the performance of the best performing model, M5. As it uses a XGBoost model, TreeExplainer proposed by Lundberg et al. [24] is used to perform the SHAP analysis. The results, presented in a summary plot based on the validation dataset (Fig. 4) highlight the feature importance for the model prediction. Features are ranked by their impact on the prediction, from top to bottom. The color gradient, from blue to red, indicates the range of feature values from Low to High in the dataset. Each point represents a single observation, and the horizontal positioning of these points reflects their contribution to the model output.

Table 9. Model Performance Comparison on Test Dataset.

Variable	MAPE (%)	Standard Deviation
A	20.64	22.87
B	16.91	17.38
C	27.11	40.46
D	32.99	88.38
M1	12.28	16.88
M2	7.78	17.38
M3	6.05	11.84
M4	7.58	19.49
M5	5.59	8.38

Table 10. Model Performance Comparison on Validation Dataset.

Variable	MAPE (%)	Standard Deviation
A	20.98	24.52
B	13.13	16.73
C	26.65	31.77
D	33.34	137.57
M1	12.36	19.62
M2	8.06	20.27
M3	6.30	15.69
M4	7.79	16.82
M5	5.79	9.01

The cumulative distance is the most important feature. It represents a clear trend of consuming more energy as the distance increases. This relationship is intuitive as longer distances usually mean more time spent driving and consequently using energy. Inertia represents the energy used to put the vehicle in motion, the more the bus must stop, the more energy it uses to restart. Aerodynamic and Gravitational energy show similar patterns as increased values improve the overall consumption. Counterintuitively, the model associates higher inertia and gravitational regeneration values with increased consumption. This contradicts physical principles, as these types of energy should reduce total consumption. The explanation for this might lie in their strong correlation with consumption variables. When a vehicle regenerates energy during braking, it still requires additional energy to get back in motion. Additionally, the vehicle typically experiences uphill and downhill transitions, as bus routes are cyclic, with few exceptions. The large variations in regeneration values are partly due to the method used to differentiate between motor-driven and regenerative power, as engine power data is not available. Since the power is negative during regeneration, an offset related to constant vehicle consumption (e.g., lighting, battery management) distorts the interpretation. Even during stops, we could not clearly identify the "base consumption" of the vehicle. This means that consumption or regeneration could be attributed to the wrong variable during

transition phases, which creates strong noises the model must deal with. The XGBoost might then misunderstand the contribution of inertia, gravitational, and aerodynamic regeneration to the overall consumption.

An increased Heating energy is also strongly linked with higher consumption, with extreme SHAP values that rival those of distance. The average speed seems to be inversely proportional to total energy consumption. This might appear counterintuitive as well, but it is in line with the expert observations from the company: lower speeds result in longer travel times, leading to higher heating energy use.

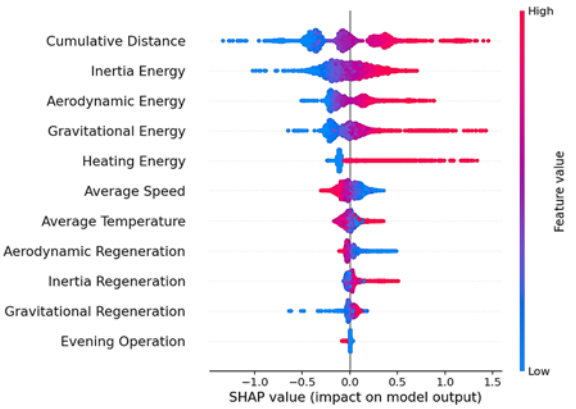


Fig. 4. Main SHAP values on the validation dataset.

Lower average speed is also linked to increased stops and higher inertia consumption. The average temperature shows no clear trend in energy consumption, with blue and red points merging near zero. This might be caused by specific heating thresholds between full diesel, full electric, or both heating modes simultaneously when the weather is particularly cold. A previous study explained that cold weather reduces the battery capacity of vehicles. Our model does not consider the available energy from the battery but rather the expected consumption. In further work, we would like to use charging data to understand the impact of temperature and state of health on the real available energy. This also highlights that averaging temperature over a trip masks important nuances and is not a well-suited feature for the model. For instance, a bus starting its day from a heated or cooled depot requires time to adjust to ambient temperatures, introducing bias in the mean temperature. On the other hand, temperature measurements from meteorological stations do not reflect what the vehicle faces in its microclimate, such as when a bus is exposed to the sun on a summer day. Investigating the impact of both temperature measurements on the model would be a valuable future direction. Lastly, the time of day has a relatively low impact on energy consumption. Still, energy consumption decreases during evening operations when traffic is lighter and fewer passengers are on board.

Although their SHAP contributions are negligible compared to the previously analyzed features, examining the remaining SHAP values can provide valuable insights into specific operational patterns. Fig. 5 highlights that certain bus lines are associated with improved energy efficiency, while others exhibit the opposite trend. Similarly, some vehicles, such as Vehicle 4, consistently show higher energy consumption than others. This could be related to a vehicle in less good working condition or with low-pressure tires and could suggest maintenance.

While Evening Operation in Fig. 4 was linked to reduced energy consumption, Afternoon Peak Operations and, more notably, Midday Operation display increased consumption patterns. This is primarily influenced by passenger load and traffic density. Such insights could help transit planners adopt a more strategic approach when assigning vehicles to specific routes, adjusting for expected service conditions to optimize energy efficiency.

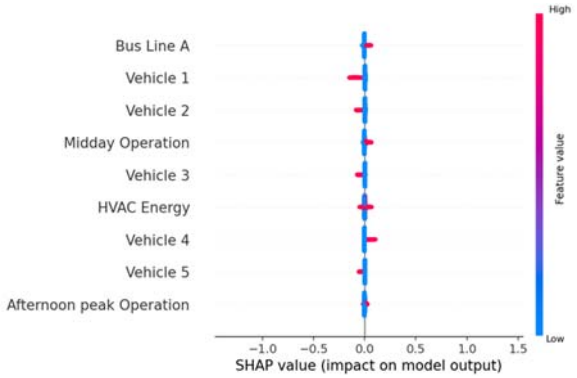


Fig. 5. Following SHAP values.

7. Discussion

In the long run, our models are designed with the ambition to support operational teams in optimizing the dimensioning of electric vehicle batteries for future bus fleets. By accurately predicting the energy needs of these vehicles, we can better plan battery sizes that will meet operational demands throughout the life cycle of the fleet, ensuring that they can handle various driving conditions and route profiles. This will improve fleet management efficiency and reduce unnecessary battery oversizing, contributing to both economic savings and sustainability.

In the short term, our consumption prediction model presents an exciting opportunity to refine the daily operational planning of electric buses. We have observed that the remaining charge at the end of the workday often remains above 50 %, suggesting that bus operations are managed conservatively to avoid service interruptions. This conservative approach, while ensuring reliability, leads to an under-utilization of the battery's capacity. Our forecasting tools allow for a more precise estimate of the energy required for each trip, enabling the operational teams to make more

informed decisions on battery use. As a result, we can safely utilize a higher proportion of the battery's available energy without compromising service reliability, contributing to the overall cost-efficiency of electric bus operations.

Furthermore, while buses are often labeled as "100 % electric", they are still equipped with diesel-powered heating systems to supplement electric heating during extreme weather conditions. These thresholds for switching to diesel heating are often set arbitrarily, and using our model to assess their impact on energy consumption could provide important insights. By analyzing the effect of these heating systems on total energy consumption during different routes, we could determine whether it would be feasible to rely solely on electric heating for specific trips, thereby reducing the environmental impact of bus operations. This would directly contribute to the broader goal of decarbonizing public transport.

Additionally, our model could be integrated with real-time adjustments. For example, energy consumption forecasts would be used to plan daily routes, and adjustments could be made dynamically based on real-time telemetry data. If energy consumption up to a given point exceeds the forecast, diesel heating could be used to conserve energy. Conversely, if energy consumption is lower than anticipated, electric heating could be prioritized to reduce emissions. This flexible approach could also lead to a new strategy of opportunistic charging, where buses would remain parked and charge for a few minutes between successive trips, ensuring they have enough energy to complete their service without interruptions.

These real-time adjustments could have additional benefits beyond day-to-day operations. Simulations using this model could be invaluable for better dimensioning charging infrastructure at different stages of fleet deployment. For instance, as the fleet transitions from diesel to electric buses, it will be crucial to ensure that charging stations are appropriately scaled to meet the increased demand. This includes forecasting the energy needs of buses as the fleet grows and ensuring that charging infrastructure can keep pace with the new fleet composition. The ability to predict energy consumption at various phases of the project (from the introduction of electric buses to the full electrification of the fleet) will allow city planners and fleet operators to make data-driven decisions regarding where and when to deploy new charging stations.

Moreover, our model could help optimize the operational patterns of the fleet by minimizing the number of charging stations required while maintaining an equivalent level of service. By ensuring that buses are appropriately charged before they go on their assigned routes, we can reduce charging downtime, improve the fleet's overall efficiency, and thus, reduce costs.

Like other researchers in this field, our model's significant advantage lies in its adaptability to different datasets. We are currently gathering telemetry data from buses operating in different Canadian cities, each with distinct topologies, climates, and travel behaviors. By applying our model to this diverse range of data, we can evaluate its robustness and accuracy across different operational contexts.

Moreover, our model has potential beyond the public transportation sector. The same principles that drive the consumption patterns of electric buses can be applied to other industries where consumption cycles are well-defined. For example, sectors like aviation, where aircraft undergo similar energy usage patterns, and mining, where heavy vehicles operate in harsh conditions, could benefit from integrating our forecasting model. Testing the model in other industries would allow us to validate its generalizability and refine its capabilities, expanding its scope and impact.

Another promising avenue for future work would be the integration of our model within an optimization framework for vehicle assignment and charging scheduling. While traditional optimization models often rely on probabilistic demand forecasts, they tend to oversimplify the complexities of real-world operations. An approach combining forecasts with real-time adjustment could provide a more dynamic and accurate solution. For instance, rather than relying on static predictions, real-time adjustments based on energy consumption data could be made to adapt charging schedules and vehicle assignments on the fly. This would improve the charging infrastructure's efficiency and the fleet’s operational planning, ensuring that vehicles are always charged when needed without over-committing charging resources.

Another important consideration, often overlooked in existing literature, is the energy consumption associated with off-route movements and non-passenger operations. These activities consume significant energy, including repositioning buses between depots, pre-heating the vehicle for passenger comfort, or repositioning buses between different routes. Experts estimate that these activities could add up to 10\% to the energy consumption forecasts for each trip. Our future work will focus on including these "off-trip" consumption factors in the model, ensuring that we provide a more realistic estimate of energy consumption that reflects the full operational cycle of the buses.

The impact of operational practices on battery lifespan is a key consideration in the electrification of buses. For instance, while effective in maximizing autonomy, opportunity charging may accelerate battery degradation due to more frequent and rapid charge cycles. Similarly, using electric heating, although more environmentally friendly than diesel heating, can cause faster battery discharge, thus reducing its lifespan. Optimizing these practices based on battery charge levels and thermal impact would be crucial to minimize premature wear while maximizing energy efficiency and battery longevity.

Finally, we recognize the growing complexity of machine learning models in the field, particularly with

the rise of large-scale artificial intelligence models and big data. While simpler models often require fewer computational resources, the recent advances in large language models (LLMs) and foundation models have introduced new challenges. These models, which offer great promise in terms of generalizability and flexibility, are often very large and computationally expensive. If these models are used for optimizing operations and supporting sustainability goals, their high training and inference costs could offset the energy savings from fleet electrification. It is crucial, therefore, to consider the energy cost of these models in the broader context of sustainability. We must balance the benefits of advanced AI models and the environmental costs associated with their deployment. Future work could focus on evaluating the energy consumption of these models and integrating that assessment into decisions about their use in optimizing electric vehicle fleets.

8. Conclusion

Our case study first highlighted the complexity of operational data from public transport companies compared to simulated data. We demonstrated the limitations of physical models in a real operational context. To address these challenges, we combined these physical models with various machine-learning approaches, allowing us to fine-tune predictions based on the operational context. The hybrid model combining physical principles with XGBoost delivered the best performance, with an error of 5.59 % on the test set and 5.79 % on the validation set, with respective standard deviations of 8.38 and 9.01.

These results demonstrate the model's ability to adapt to noisy data while providing effective energy consumption predictions across different routes.

References

[1]. L. Adam, R. Pellerin, B. Agard, Analysis of rule-based models for predicting energy consumption in electric bus fleets: a case study with telemetry data, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 68-73.

[2]. S. Krawiec, B. Łazarz, S. Markusik, G. Karoń, et al., Urban public transport with the use of electric buses – development tendencies, *Transport Problems*, Vol. 11, 2016, pp. 127-137.

[3]. S. Corbet, C. Larkin, J. McCluskey, The influence of inclement weather on electric bus efficiency: evidence from a developed European network, *Case Studies on Transport Policy*, Vol. 12, 2023, 100971.

[4]. B. Zhang, N. Niu, H. Li, Z. Wang, et al., Could fast battery charging effectively mitigate range anxiety in electric vehicle usage?, *Transportation Research Part D: Transport and Environment*, Vol. 95, 2021, 102840.

[5]. A. Jahic, M. Eskander, E. Avdevicius, D. Schulz, Energy consumption of battery-electric buses: review of influential parameters and modelling approaches, *B&H Electrical Engineering*, Vol. 17, Issue 1, 2023, pp. 7-17.

[6]. Y. Chen, Y. Zhang, R. Sun, Data-driven estimation of energy consumption for electric bus under real-world driving conditions, *Transportation Research Part D: Transport and Environment*, Vol. 98, 2021, 102969.

[7]. L. K. Lim, Z. A. Muis, W. S. Ho, H. Hashim, et al., Review of the energy forecasting and scheduling model for electric buses, *Energy*, Vol. 263, 2022, 125773.

[8]. X. Zhang, Z. Zhang, Y. Liu, Z. Xu, et al., A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation, *Renewable Energy*, Vol. 234, 2024, 121243.

[9]. P. Iora, L. Tribioli, Effect of ambient temperature on electric vehicles' energy consumption and range: model definition and sensitivity analysis based on Nissan Leaf data, *World Electric Vehicle Journal*, Vol. 10, Issue 2, 2019, 21.

[10]. Y. Chen, G. Wu, R. Sun, A. Dubey, et al., A review and outlook of energy consumption estimation models for electric vehicles, *arXiv preprint*, 2020, arXiv:2001.07429.

[11]. C. J. J. Beckers, I. J. M. Besselink, H. Nijmeijer, Assessing the impact of cornering losses on the energy consumption of electric city buses, *Transportation Research Part D: Transport and Environment*, Vol. 86, 2020, 102360.

[12]. C. Fiori, M. Montanino, S. Nielsen, M. Seredynski, et al., Microscopic energy consumption modelling of electric buses: model development, calibration, and validation, *Transportation Research Part D: Transport and Environment*, Vol. 98, 2021, 102978.

[13]. C. J. J. Beckers, I. J. M. Besselink, J. J. M. Frints, H. Nijmeijer, Energy consumption prediction for electric city buses, in *Proceedings of the 13th ITS European Congress*, 2019.

[14]. L. A. W. Blades, T. Matthews, T. E. McGrath, J. Early, et al., Predicting energy consumption of zero emission buses using route feature selection methods, *Transportation Research Part D: Transport and Environment*, Vol. 130, 2024, 104158.

[15]. F. Basso, F. Feijoo, R. Pezoa, M. Varas, et al., The impact of electromobility in public transport: an estimation of energy consumption using disaggregated data in Santiago, Chile, *Energy*, Vol. 286, 2023, 129550.

[16]. T. Pamuła, D. Pamuła, Prediction of electric buses energy consumption from trip parameters using deep learning, *Energies*, Vol. 15, Issue 5, 2022, 1747.

[17]. Z. Xu, J. Wang, P. D. Lund, Y. Zhang, Analysis of energy consumption for electric buses based on low-frequency real-world data, *Transportation Research Part D: Transport and Environment*, Vol. 122, 2023, 103857.

[18]. R. Sennefelder, R. Martín-Clemente, R. González-Carvajal, Data driven characterization and predictive classification of energy economy for public transport, in *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI')*, 2021, pp. 1254-1262.

[19]. R. M. Sennefelder, P. Micek, R. Martín-Clemente, J. C. Risquez, et al., Driving cycle synthesis, aiming for realness, by extending real-world driving databases, *IEEE Access*, Vol. 10, 2022, pp. 54123-54135.

[20]. CGIAR-CSI SRTM, SRTM Data, https://srtm.csi.cgiar.org/srtmdata/

[21]. Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, et al., KAN: Kolmogorov-Arnold networks, *arXiv preprint*, 2024, arXiv:2404.19756.
[22]. D. P. Kingma, J. L. Ba, Adam: a method for stochastic optimization, *arXiv preprint*, 2014, arXiv:1412.6980.
[23]. S. M. Lundberg, S. Lee, A unified approach to interpreting model predictions, *arXiv preprint*, 2017, arXiv:1705.07874.
[24]. S. M. Lundberg, G. Erion, H. Chen, et al., From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*, Vol. 2, Issue 1, 2020, pp. 56-67.

Appendix

Table 11. Hyperparameters of the neural network models and XGBoost.

Architecture	Parameter	Value
KAN	hidden_size1	29
	hidden_size2	20
	grid	3
	k	2
	learning_rate	0.010
	optimizer	Adam
	epoch	250
	Loss	L1Loss
XGBoost	learning_rate	0.0299
	n_estimators	975
	max_depth	48
	subsample	0.642
	colsample_bytree	0.614
	gamma	1.476e-5
	min_child_weight	56
MLP	hidden_size1	46
	hidden_size2	32
	learning_rate	0.001
	optimizer	Adam
	epoch	2500
	Loss	L1Loss

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 59-65

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Adoption of Post-Quantum Cryptography and FIPS Standards in Existing as well as Emerging Communication Technologies

[1, *] Vaghawan OJHA, [2] Sumit CHAUHAN, [1] Shanita YARAHMADHIAN and [2] David CARVALHO

[1] Mississippi State University, Bagley College of Engineering, Department of Mathematics & Statistics, 410 Allen Hall, 39762 Starkville, USA [2] Naoris Tech Inc., 251 Little Falls Drive, 54321, Wilmington, New Castal County, USA

[*] E-mail: vpo4@msstate.edu, sumit@naoris.com

Received: 30 Oct. 2024 /Revised:22 April 2025 /Accepted: 25 April 2025 /Published:30 April 2025

Abstract: With emerging advancements on Quantum computing and active research on effectiveness of classical cryptographic algorithms, it is imperative that the coming era of Quantum computation is going to have heavy effect on all technologies which requires some form of communication primarily due to the security of encrypted data. National Institute of Standards and Technology (NIST) has recently standardized Crystal-Dilithium for Digital Signature on the 3rd round of NIST Post-Quantum Cryptography (PQC) open competition. There are other algorithms taken to the 4th round for further evaluation. Recently on August 24th, 2024, NIST announced Federal Information Processing Standards (FIPS) for all federal services which includes several PQC cryptosystems, including Crystal-Dilithium, SPHINCH+ for Digital Signature and Crystal-Kyber for Key Encapsulation Mechanism. In the same vein, much speculation and experiments have been done of PQC adoption in existing systems which rely heavily on classical cryptographic algorithms. Hence, in this paper we would like to present the possible adoption framework (both hybrid and fresh) of PQC cryptosystem in existing as well as new distributed and communication technology that may emerge.

Keywords: Post-quantum cryptography, PQC, Distributed systems, Industry 4.0.

1. Introduction

After the invention of Shor's algorithm [4], which directly affects the Prime factorization in which the classical cryptographic algorithms such as RSA, DSA and ECDSA rely on, researchers started to explore alternative cryptographic algorithms which are not directly affected by the quantum computers or can safely surpass the threats. In a similar vein, we previously explored some of the key problems, challenges, and possible solutions in our previous study [2]. NIST announced the first PQC competition in 2016, and that lead to the development and validation of different post quantum algorithms which can be primary choice and secondary alternatives based on the need. These post-quantum algorithms consist of the possible replacement of Digital Signatures and Key-Encapsulation mechanisms, which are used across all digital communication. In the previous study [2], we explored how classical digital signatures can be broken with quantum attacks. Hence, the adoption of PQC is unquestionable and inevitable. In fact, the whole Web 2.0 technologies get affected once the quantum computers with a decent qubit are realized. The most common types of quantum attacks include intercept-and-resend, teleportation, man-in-the-middle, participant, and implementation attacks, each with distinct mechanisms and

implications for security. While some applications such as OpenSSH and Google Chrome have started to implement PQC, the adoption rate is very low at around 0.029 % even in the National Supercomputing clusters [12].

To avoid any potential losses, which may be approaching us, NIST has finalized sets of algorithms which can outstand the known attacks, and so far, have been proven to be quantum resistant.

1.1. FIPS Standards

The US government announced three new Federal Information Processing Standards (FIPS) [5] on August 14, 2024. These standards were well-studied and decided to mitigate the future quantum computing risks. These new standards focus on making key establishment and digital signatures quantum resistant.

The first one, FIPS 203, introduces an algorithm called Module-Lattice Based Key Encapsulation Mechanism ML-KEM (based on Crystal-Kyber) for key encapsulation. Then another standard FIPS 204, which brings in Module-Lattice-Based Digital Signature Algorithm (ML-DSA) for digital signatures - that one comes from Crystal-Dilithium. [10] Both use Module-Lattice as their foundation. The last one, FIPS 205, takes a different approach with a hash-based digital signature scheme that builds on SPHINCS+ [9]. These standardized algorithms brace themselves heavily on the mathematical and computationally difficult problems, often categorized as hard problems in theoretical computer science.

1.2. Finalized PQC Algorithms

In our previous study, we comprehensively reviewed some of the finalized PQC algorithms from NIST round 3 and some under the consideration for round 4 [1]. These algorithms are primarily of two types, one is Key-Encapsulation Mechanism, and the other is Digital Signatures.

To produce quantum-safe asymmetric key pairs, PQC attacks the problems from different fronts. These include lattice-based cryptography, which addresses well-known lattice problems like the shortest vector problem; code-based cryptography, which takes advantage of the complexity of decoding generic linear codes; and multivariate cryptography, which expands on multivariate polynomials over a finite field [1].

Key Encapsulation Mechanism (KEM) is a secure way to wrap up and send secret keys between two parties. It is like putting a key in a special unbreakable box that only the intended parties can open. Whenever we are doing secure messaging, any other form of communication or online banking, KEMs help create those secure connections by safely exchanging the secret keys needed for encryption. Digital Signatures are like digital footprints of our handwritten signature, but cryptographically secure. When one party digitally sign something, the concerned party is creating a unique mathematical stamp that proves two things: 1) That he/she is really the one who sent the message, and 2) That the message hasn't been tampered with since the party signed it which is also known as integrity property of Digital Signature. With this approach, nothing inside the communication can be tempered or forged. Both KEM and Digital Signature are the building blocks of modern cybersecurity.

FIPS 203, introduces the PQC algorithm called Module Lattice based Key Encapsulation Mechanism (ML-KEM). The underlying theory behind this is the Module Lattice and hardness property of Module Lattice [5], and the Digital Signature called Dilithium which also relies on the hardness of Module Lattice. Lattice based cryptographic algorithms gained traction after being consistently proven to be effective under various attacks, both quantum and non-quantum ones. In fact, 7 out of 15 submitted candidate algorithms in NIST round 3 competition were Lattice based algorithms [1]. Giving an alternative to Lattice, NIST, in FIPS 205 also standardized a hash-based Digital Signature called SPHINCS+. Using different security parameters, these algorithms have passed different security level test designed by NIST, such as security level 1-5 which pre-defines number of security aspects in each security level, level 5 being the highest [11].

1.3. Role of FIPS in Developing Secure Communication

Modern internet communication relies on protocols such as HTTP and HTTPS within the SSL/TLS stack, which currently utilize classical cryptographic algorithms – RSA, Diffie-Hellman (DH), Elliptic-curve Diffie-Hellman (ECDH), ECDSA, and the Digital Signature Algorithm (DSA) – that lack quantum resistance [1, 3]. As the realization of sufficiently powerful quantum computers draws closer, these conventional mechanisms face critical vulnerabilities, including “store-now-decrypt-later” attacks that exploit data once quantum decryption becomes feasible.

To address this looming risk, Federal Information Processing Standards (FIPS) – developed by the National Institute of Standards and Technology (NIST) – serve as a guideline for guiding secure communication. FIPS provides standardized best practices for integrating post-quantum cryptographic (PQC) algorithms into both legacy and new infrastructures. It also provides the security parameters needed for the adoption of these algorithms. Alongside NIST’s efforts, organizations such as the Post-Quantum Cryptography Alliance (PQCA) [7], Open Quantum Safe (OQS) [6], PQCrypto [8], and the Linux Foundation are creating standardized frameworks and tools to facilitate PQC adoption industry wide.

2. Proposed Methods for Adoption of FIPS

We propose two major methods of adoption of PQC algorithms suits standardized in FIPS, one focusing on the existing technologies that are already in operation and need the quantum safety nest, and the other, which is a fresh-future systems, which can be designed to be quantum-safe from the beginning. We will also list down the possible solutions that organizations such as PQCA, Open Quantum Safe and Linux Foundations are working on. We discuss both fresh and hybrid adoption methods which industries can adopt based on their needs.

2.1. Key Considerations

There are some key considerations that organizations need to make before they move towards the adoption of PQC algorithms in their existing infrastructure.

- Existing cryptanalysis of existing algorithms in place;
- Feasibility of quantum attacks in question;
- Existence of store-now-decrypt-later attacks;
- Cost of adoption (both performance, time, and money);
- Availability of Software suites/packages to use.

Before organizations transition to post-quantum cryptographic (PQC) algorithms, a careful evaluation of multiple technical and strategic dimensions is important. These include the current robustness of classical cryptographic algorithms being used against cryptanalysis, the practical feasibility of quantum attacks, and the risk posed by threats such as store-now-decrypt-later threats. In addition to these security concerns, implementation costs – spanning performance trade-offs, time for migration, and financial investment – must be considered. Furthermore, the availability and maturity of supporting software suites play a pivotal role in enabling seamless integration. These considerations are summarized in Fig. 1, which shows the interdependencies that organizations must navigate in their decision-making process.

2.2. Adoption Layers

There are two critical aspects that need to be discussed. The first is how to change the most basic communication layers, such as SSL, port communication, and the traditional validation algorithms associated with them. These foundational protocols must adapt to remain secure in a post-quantum era, where traditional cryptographic methods are at risk of being compromised by quantum computing advancements. The second aspect is the higher communication layer, encompassing signatures and the encryption of the data used and shared, which is an integral part of the communication. This layer is equally critical, as it ensures the integrity and confidentiality of data exchanged over networks.

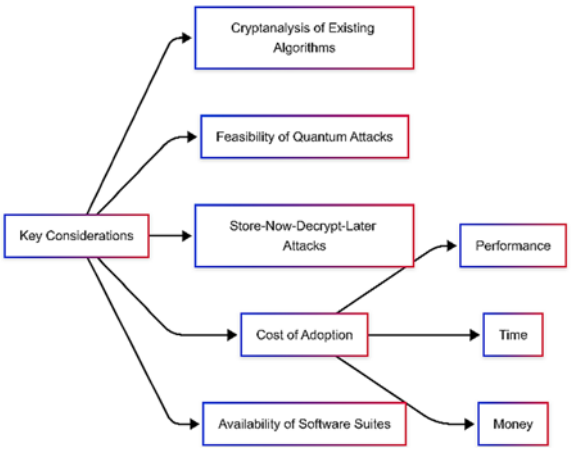


Fig. 1. Key Considerations for Adoption of PQC.

2.2.1. Basic Communication Layer Integration

For foundational protocols like SSL and port communication, adopting hybrid cryptography ensures secure data transmission in real-time while providing interoperability and compatibility measures become important once we introduce the hybrid adoption of the PQC which goes hand in hand with the classical security that is already in place. Specifically, we need to ensure the following:

Backward Compatibility: we need to ensure that the hybrid framework supports backward compatibility to allow legacy systems and devices to continue operating. This minimizes user friction and ensures adoption by both new and existing systems.

Cross-Platform Testing: Conducting compatibility tests across devices, networks, and platforms to confirm consistent encryption/decryption operations with the hybrid model is important. It is critical to test hybrid systems in varied environments, including IoT devices, cloud, and mobile platforms.

Dynamic Configuration: adaptive mechanisms that switch between classical and PQC methods based on network conditions or client capabilities can be helpful especially when the environment on which the system operates varies across different platforms and hardware's. For example, devices with older processors could default to classical cryptography, while newer hardware supports PQC.

Basic communication protocols can implement hybrid key encapsulation mechanisms (KEMs) where a classical encryption key secures ongoing communication, while a post-quantum key ensures long-term data security. For example, a secure TLS handshake can simultaneously use both RSA/ECC and PQC algorithms like Kyber for key exchange. A layered architecture can dynamically adapt encryption schemes depending on the capabilities of the client or server.

2.2.2. Higher Communication Layer Integration

For higher layers, such as digital signatures and encryption, hybrid adoption offers a path to ensure authentication, integrity, and confidentiality.

Signatures: By implementing module-lattice-based digital signatures (e.g., CRYSTALS-Dilithium), systems can validate messages with quantum-safe authentication while maintaining classical fallback options for compatibility.

Encryption: Data transmitted at the higher communication layer can utilize hybrid encryption schemes, ensuring sensitive information is protected both during transit and for long-term storage.

2.3. Adoption Methods

In this section we discuss different adoption methods which can be used whilst considering various aspects of institutional needs and the urgency of PQC adoption to meet the FIPS standards.

2.3.1. Hybrid Adoption Method

Implementing hybrid cryptography enables systems to operate in both classical and quantum-safe modes, offering a buffer as organizations and users adapt. This method allows gradual transition without a single cutover date, which can be beneficial for critical systems requiring high availability.

Dual Cryptographic Layer Design:

- Set up a dual-mode architecture where classical and PQC algorithms can operate in tandem. For instance, we use hybrid key encapsulation where a secure classical encryption key is used for current security, while the PQC key ensures future proofing;
- Flexible Protocol Layering: Define a layering structure for cryptographic protocols that allows for adaptable encryption schemes depending on device or network capabilities.

This can involve:

- Primary Layer (Classical Algorithm): Continues to secure current operations;
- Secondary Layer (PQC Algorithm): Adds a quantum-safe layer to protect data for long-term security.

The hybrid adoption method aligns seamlessly with the two critical communication layers discussed earlier, basic communication layers such as SSL, port communication and higher communication layers such as signatures and encryption. By integrating a dual cryptographic approach, systems can maintain current operational security while preparing for quantum-safe cryptography.

2.3.2. Fresh Adoption

Fresh adoption of the FIPS standards involves a complete overhaul of cryptographic systems, transitioning fully from classical algorithms to quantum-safe solutions or building from ground-up. This approach is more suitable for new systems or environments where backward compatibility with legacy systems is not a constraint. By integrating post-quantum algorithms directly into both the basic and higher communication layers, organizations can ensure a robust foundation for future security.

2.3.2.1. Basic Communication Layer Integration

For the basic communication layer, such as SSL, port communication, and foundational protocol operations, fresh adoption replaces classical cryptographic elements with FIPS-certified PQC algorithms.

Direct Integration of Quantum-Safe KEMs:

In this approach, key encapsulation mechanisms like Kyber are implemented as the sole method for secure key exchange in protocols like TLS. This eliminates the reliance on RSA or ECC and ensures resistance to quantum attacks.

Redesign of Protocols:

Protocols such as SSL and TLS are redesigned to rely exclusively on post-quantum cryptographic standards. This streamlines operations by removing hybrid complexities and focusing solely on quantum-resistant algorithms.

Simplification of Validation Processes:

Traditional validation algorithms are replaced with those optimized for post-quantum operations to ensure security.

2.3.2.2. Higher Communication Layer Integration

At the higher communication layer, focusing on authentication, signatures, and data encryption, fresh adoption ensures that all cryptographic mechanisms are inherently quantum-safe.

Signatures: Module-lattice-based digital signature schemes (e.g., CRYSTALS-Dilithium) are adopted as the primary method for message authentication. These algorithms replace classical signatures entirely, offering a quantum-resistant solution that aligns with FIPS PQC standards.

Encryption: Data encryption protocols are redesigned to use post-quantum encryption algorithms, ensuring the confidentiality and integrity of sensitive communications, even when stored long-term.

Authentication Systems: Fresh adoption involves implementing quantum-safe authentication systems across the network, such as identity verification processes or certificate authorities that solely use PQC algorithms.

3. Available Frameworks

3.1. Open Quantum Safe (QSS)

Open Quantum Safe (OQS) [6] is an open-source initiative under Linux Foundation, dedicated to developing and integrating quantum-resistant cryptographic algorithms into widely used frameworks. The project aims to facilitate research, experimentation, and adoption of post-quantum cryptography (PQC) in both academic and industry settings. It primarily revolves around two key components:

liboqs: A C library that provides implementations of various quantum-safe key encapsulation mechanisms (KEMs) and digital signature algorithms.

Integrations with Popular Cryptographic Libraries: The OQS team maintains branches of well-known libraries like OpenSSL and BoringSSL augmented with PQC algorithms available in liboqs.

From a scientific standpoint, OQS serves as a practical testbed for the performance, security, and interoperability of PQC schemes identified by the NIST PQC standardization process. It provides standardized APIs and reference implementations that enable systematic benchmarking of key generation, encryption/decryption throughput, signature creation/verification times, and ciphertext/signature sizes across various security levels. By integrating with TLS libraries like OpenSSL, OQS also enables interoperability testing under real-world network conditions, allowing researchers and developers to evaluate handshake integrity, latency overhead, and protocol compliance.

Moreover, OQS plays a pivotal role in guiding early industry adoption, particularly for cloud providers and secure communication platforms, by providing drop-in compatibility with widely used security protocols like TLS. Its modular, well-documented architecture and language bindings (e.g., Python) enable rapid prototyping and lower the barrier for developers to explore PQC integration. Additionally, as an open-source testbed, OQS helps surface integration challenges, performance bottlenecks, and protocol-level edge cases, which are critical for refining upstream cryptographic stacks and contributing to the robustness of PQC algorithms prior to their standardization.

3.2. Manual Integration

Manual integration involves directly modifying cryptographic libraries – such as OpenSSL, BoringSSL, or libcrypto – to embed institution-specific post-quantum cryptographic (PQC) algorithms or to enforce custom protocol behaviors. This approach enables **fine-grained control** over algorithm selection, parameter tuning, and handshake customization, aligning closely with internal security policies, performance constraints, or compliance needs. It is especially valuable in niche industries like defense, critical infrastructure, or national security systems, where off-the-shelf solutions may not meet operational or regulatory standards. However, such integrations are resource-intensive – requiring deep knowledge of cryptographic primitives, constant tracking of upstream library updates, and high maintenance overhead.

In contrast, modular integration – as powered by projects like Open Quantum Safe (OQS) – provides extensibility without altering core cryptographic libraries by offering prebuilt wrappers and extension APIs. These wrappers (e.g., OQS-OpenSSL, OQS-SSH) allow organizations to experiment with NIST finalist algorithms (such as Kyber and Dilithium) with minimal changes to their existing infrastructure. Similarly, hybrid TLS modes, where classical and PQC algorithms are combined in the handshake, offer a practical bridge between legacy and future-proof security.

While manual integration offers unmatched flexibility, for many organizations – especially those lacking in-house cryptographic expertise – a wrapper-based or plug-and-play model (like that of liboqs or Microsoft's PQCrypto-VPN prototype) may be more sustainable and future-ready.

4. Case Studies from Naoris

4.1. PQC-Enabled OpenVPN

At Naoris, we leveraged the Naoris core-based Dilithium algorithm, a post-quantum cryptographic signature scheme, into OpenVPN systems to enhance security against quantum computing threats. We also compared the Naoris implementation of Dilithium with traditional cryptographic methods and other Dilithium implementations, focusing on performance metrics as latency, throughput, and computational overhead. The Naoris core is specifically optimized for VPN use cases, aiming to provide robust security while minimizing performance impact. The integration also involved a comparative analysis using diverse hardware configurations and simulated network conditions, and the OpenVPN configuration was modified to incorporate the Naoris core-based Dilithium implementation for key exchange and digital signatures, replacing the default RSA or ECC-based methods. We also assessed the implementation's resistance to side-channel and cryptanalytic attacks, and the security of the Naoris implementation is analyzed against classical and quantum attacks, specifically assessing its resistance to timing, power analysis, and cache-based attacks. Although this involved manual integration of the PQC algorithms on all communication layers, this was proven to be the most effective approach in terms of computational complexity, efficiency, and performance while not trading off the security aspects.

4.2. PQC Enabled Blockchain Network

At Naoris, we've embedded FIPS security into the foundation of our blockchain – "Naoris Protocol". As part of the next generation of blockchain innovation, it stepped into the future with post-quantum (PQ) capabilities at the core of the protocol. The Naoris Protocol is designed to proactively address the growing concerns surrounding quantum computing threats. By successfully integrating post-quantum cryptographic algorithms at its core, it ensure that all transactions, data, and smart contracts are not only secure but are also future-proofed against the potential risks posed by quantum advancements. This enhanced security framework enables the chain to validate every transaction as Post-Quantum enabled, embedding it directly into robust security layer. The protocol is setting a new standard for blockchain security in the quantum era.

5. Discussion

The growing capabilities of quantum computing pose significant risks to classical cryptographic systems [1-3, 5, 12-14], driving the urgent need for Post-Quantum Cryptography (PQC). In this context, FIPS standards play an important role in standardizing the algorithms and certifying them pushing the industries to adopt PQC in their new and legacy systems [2, 12-14]. By offering standardized approaches to algorithm selection and implementation, FIPS [5] gives a baseline for security, interoperability, and regulatory compliance across different industry sectors.

PQC algorithms particularly those based on lattice, code, and hash-based are designed to withstand quantum attacks that threaten classical methods like RSA and ECC [1, 13, 14]. The NIST PQC standardization process has finalized leading candidates CRYSTALS-Kyber, CRYSTALS-Dilithium, ML-KEM that organizations can adopt with reasonable confidence in their security and performance [5]. Integrating these algorithms into real-world protocols and applications remains a complex task [13, 14], yet it is critical for safeguarding data in both the near- and long-term future.

When considering adoption methods, two primary choices emerge. The hybrid method combines classical and PQC algorithms, enabling organizations to maintain backward compatibility for legacy systems [12, 14] while gradually introducing quantum-safe mechanisms. The advantage of this approach is the reduced disruption to existing infrastructure, but it carries additional overhead in maintaining dual systems and protocols. In contrast, a fresh adoption eliminates classical algorithms entirely in favor of post-quantum schemes. This method streamlines cryptographic architectures and ensures maximal quantum resistance, but it requires significant infrastructure overhaul, and any incompatibility with legacy systems can be costly in terms of time and resources.

Despite the clear advantages of PQC, future challenges persist [14]. First, algorithm maturity and real-world resilience remain topics of active research; while many PQC algorithms have undergone intensive cryptanalysis, no solution is completely immune to future cryptanalytic discoveries. Second, performance overhead especially with larger key sizes and signatures can be problematic in constrained environments such as IoT devices. Third, global interoperability demands careful coordination between standards bodies and industry consortia to ensure consistent implementations.

Looking ahead, future directions will likely involve continued optimization of PQC algorithms, including hardware acceleration and specialized libraries to mitigate performance bottlenecks. As the NIST standards progress, more refined guidance on parameter sets and security levels will emerge, helping to unify deployments. Additionally, an evolving ecosystem of open-source libraries, such as Open Quantum Safe, will drive widespread testing and collaboration, further accelerating PQC readiness. This slow pace of adoption is concerning, especially considering the "store now, decrypt later" threat, where encrypted data can be collected now and decrypted by quantum computers in the future [12]. Many organizations are already late in planning their transition to PQC, and this migration is more complex than previous transitions [13]. Hybrid approaches, combining classical and PQC algorithms, are being explored to ensure security during the transition process [13, 14].

To increase adoption across industries, a concerted effort is needed, encompassing broad education, policy incentives, and technical assistance. Organizations must be made aware of the "harvest now, decrypt later" threat model and motivated to modernize their cryptographic infrastructures before quantum computing becomes mainstream. Government agencies and compliance frameworks can expedite this process by mandating or strongly recommending FIPS integration, while academic-industry partnerships can increase the research and tooling necessary for seamless transitions. Ultimately, early action and thorough planning will be critical in safeguarding the ever-prevailing digital communications against emerging quantum threats and standards like FIPS will tremendously help streamline the adoption process.

Acknowledgements

We thank the Naoris team for the continuous guidance and technical support while writing of this paper.

References

[1]. V. Ojha, S. Chauhan, S. Yarahmadian, D. Carvalho, Unfolding post-quantum cryptosystem: review of CRYSTALS-Dilithium, McEliece, BIKE and HQC (unpublished, under review).

[2]. S. Chauhan, V. P. Ojha, S. Yarahmadian, D. Carvalho, Towards building quantum resistant blockchain, in *Proceedings of the International Conference on Electrical, Computer and Energy Technologies (ICECET'23)*, 2023, pp. 1-9.

[3]. Bughunters by Google, Google's threat model for post-quantum cryptography, https://bughunters.google.com/blog/5108747984306176/google-s-threat-model-for-post-quantum-cryptography

[4]. P. Shor, The early days of quantum computation, *arXiv preprint*, 2022, arXiv:2208.09964.

[5]. Federal Register, Announcing Issuance of Federal Information Processing Standards (FIPS) FIPS 203, Module-Lattice-Based Key-Encapsulation Mechanism Standard, FIPS 204, Module-Lattice-Based Digital Signature Standard, and FIPS 205, Stateless Hash-Based Digital Signature Standard, https://www.federalregister.gov/documents/2024/08/14/2024-17956/announcing-issuance-of-federal-information-processing-standards-fips-fips-203-module-lattice-based

[6]. Open Quantum Safe Project, https://openquantumsafe.org/

[7]. Post-Quantum Cryptography Alliance (PQCA), https://pqca.org/

[8]. PQCrypto Project, https://pqcrypto.org/

[9]. D. Bernstein, et al., The SPHINCS+ signature framework, in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 2129-2146.

[10]. L. Ducas, et al., CRYSTALS – Dilithium: Digital Signatures from Module Lattices, Report 2017/633, *IACR Cryptology ePrint Archive*, 2017.

[11]. NIST Computer Security Resource Center (CSRC), Post-Quantum Cryptography: Security (Evaluation Criteria), https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization/evaluation-criteria/security-(evaluation-criteria)

[12]. J. Sowa, B. Hoang, A. Yeluru, S. Qie, et al., Post-quantum cryptography (PQC) network instrument: measuring PQC adoption rates and identifying migration pathways, in *Proceedings of the IEEE International Conference on Quantum Computing and Engineering (QCE'24)*, 2024, pp. 1835-1846.

[13]. D. Joseph, R. Misoczki, M. Manzano, J. P. Schanck, et al., Transitioning organizations to post-quantum cryptography, *Nature*, Vol. 605, Issue 7905, 2022, pp. 237-243.

[14]. G. S. Mamatha, N. Dimri, R. Sinha, Post-quantum cryptography: securing digital communication in the quantum era, *arXiv preprint*, 2024, arXiv:2403.11741.

[15]. V. Ojha, S. Chauhan, S. Yarahmadian, D. Carvalho, Adoption of post-quantum cryptography in communication technologies, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 21-26.

Sensors & Transducers, Vol. 268, Issue 1, April 2025, pp. 66-74

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Transfer Learning-driven Comparative Analysis of GA, PSO, and NSGA-II Over FIS for Enhanced Energy Efficiency in Assisted Living Settings

Anita XHEMALI [1], Elma ZANAJ [2], Gledis BASHA[2] and Lorena BALLIU [2]

[1] Polytechnic University of Tirana, Faculty of Electrical Engineering,
Boulevard 'Dëshmorëte Kombit', Square 'Mother Teresa', 4, Albania
[2] Polytechnic University of Tirana, Faculty of Information Technology,
Boulevard 'Dëshmorëte Kombit', Square 'Mother Teresa', 4, Albania
E-mail: anita.xhemali@fti.edu.al, ezanaj@fti.edu.al,
gledis.basha@fti.edu.al, lorena.balliu@fti.edu.al

Received: 12 Nov. 2024 /Revised: 15 March 2025 /Accepted: 16 April 2025 /Published:30 April 2025

Abstract: This research explores the integration of transfer learning within optimization algorithms Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Non-dominated Sorting Genetic Algorithm II (NSGA-II) over Fuzzy Systems (FIS) for enhancing energy efficiency in assisted living environments. We optimize FIS models by testing various transfer learning combinations: GA to PSO, GA to NSGA-II, PSO to NSGA-II, and GA to NGSA-II. Results show that PSO to NSGA-II delivers the best performance. GA to NSGA-II also showed notable improvement, benefiting from NSGA-II's efficient Pareto front exploration following GA's broad search capabilities. GA to PSO demonstrated slight improvement over GA alone, but PSO after GA performed worse due to premature convergence and reduced genetic diversity. In contrast, GA to NSGA-II retained better solution diversity, improving multi-objective optimization outcomes. These findings highlight the potential of transfer learning to enhance energy optimization in complex assisted living systems and provide deeper insights into its role in improving energy efficiency through strategic algorithmic pairing.

Keywords: Transfer learning fuzzy inference systems, Energy optimization, Genetic algorithms, PSO, Multi-objective optimization, Pareto front.

1. Introduction

The optimization of energy efficiency in assisted living environments is a critical area of research, as it can directly contribute to the sustainability, "green environment", and cost-effectiveness of these settings. To address this, advanced optimization techniques such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Non-dominated Sorting Genetic Algorithm II (NSGA-II) and Fuzzy Inference Systems (FIS) are being increasingly explored to improve energy usage while balancing comfort [1]. Despite improvements, there were limitations due to search space convergence and genetic diversity issues. Several studies have demonstrated the power of MOEAs (Multi-Objective Evolutionary Algorithms) in optimizing complex systems, such as building energy efficiency [2]. For example, SMOPSO/D (Simple Multi-Objective Particle Swarm Optimization with Decomposition), a multi-agent evolutionary algorithm, has been used to optimize energy-efficient building designs by balancing objectives such as energy use, cost, and comfort [3]. Enhanced algorithms like NSGA-II have also been integrated with multi-agent systems to improve solution diversity and convergence rates in building energy design optimization [4]. By underlying multi-agent strategies, these algorithms can dynamically adjust to changing environmental and user conditions, making them ideal for assisted living environments where comfort and

energy efficiency are often at odds. Transfer learning, which highlights past knowledge to accelerate learning in new tasks, has shown promise in improving the efficiency of evolutionary algorithms. In dynamic multi-objective optimization problems (DMOPs), reusing past experiences can enhance both robustness and performance of algorithms like NSGA-II, MOPSO (Multi-Objective PSO), and RM-MEDA (Regularity Model-Based Multi-Objective Estimation of Distribution Algorithm) [5]. For instance, PSO-based controllers have outperformed traditional methods in load frequency control by improving system performance and reducing convergence time [6]. Hybrid approaches have also demonstrated improved optimization performance. For example, the IHOGCP algorithm, which combines GA and PSO, has been used to optimize task offloading in the Internet of Reconfigurable Things (IoRT), enhancing transmission rates and reducing energy consumption under deadlines [7]. Such hybrid models have the potential to improve FIS based energy prediction in assisted living environments by enabling real-time adaptive energy management. Multi-agent systems can dynamically respond to occupant behavior and environmental changes, making them suitable for complex systems like assisted living environments, [8]. Transfer learning has also been applied to dynamic community detection in complex networks, with algorithms like TMOGA (Transfer Learning-Based Multi-Objective Genetic Algorithm) outperforming traditional methods in terms of efficiency and convergence [9]. Despite its success in classification tasks, few approaches have targeted regression problems in fuzzy systems. Traditional fuzzy regression transfer learning often performs well only in constrained domains with limited target data, limiting its practical application [10]. A key challenge in fuzzy transfer learning is aligning data distributions between the source and target domains. Unlike conventional kernel-based nonlinear mappings, fuzzy-based approaches use transformations in the antecedent part of fuzzy rules, making the process more interpretable [11]. Another challenge in fuzzy transfer learning is how to effectively merge and use knowledge from multiple source domains [12]. At the best of our knowledge there are algorithms that explore two approaches for combining fuzzy rules from different domains in regression tasks: one for homogeneous domains and another for heterogeneous domains. In homogeneous cases, knowledge is integrated directly through fuzzy rules, while at heterogeneous cases, knowledge is merged in the form of both data and fuzzy rules. While this research focuses on dynamic FIS surface optimization, its insights are applicable to IoT systems in assisted living environments, where dynamic changes in occupant behavior and environmental conditions require adaptive energy management systems, as demonstrated in our previous work [1]. By incorporating transfer learning from GA (Genetic Algorithm) to PSO (Particle Swarm Optimization), GA to NSGA-II, and PSO to Non-dominated Sorting Genetic Algorithm II (NSGA-II), it is possible to harness previous optimization experiences to enhance the performance of FIS controllers, resulting in more efficient energy management. Reusing past optimization cycles allows the FIS-based models to adapt more effectively to changes in energy usage patterns and occupant behavior. This approach aims to provide improved long-term energy efficiency in assisted living environments, ensuring better comfort, lower costs, and more sustainable operation.

Applying transfer learning to optimize fuzzy control in assisted living settings, specifically balancing energy efficiency and temperature stability, represents a new and practical approach. The structured pipeline of combining multiple algorithms allows for better handling of complex, multi-objective goals in real-world scenarios. The results highlight the potential of transfer learning to significantly improve energy optimization and occupant comfort in assisted living systems, where maintaining a balance between energy use and environmental stability is essential.

The paper is organized as follows: Section 2 addresses the flowchart and creation of triple flow of optimisation covering all three transfer learning combinations (GA → PSO, GA → NSGA-II, PSO → NSGA-II). Section 3 describe simulation results of 12 months of real IoT data, 12 months of generative data for energy consumption, 36 months generative database and also the discussion over data tables, respectively. Sections 4 shows our simulator results for analyse of variance and the optimized surfaces for datasets including visualisastion of Pareto Fronts. Section 5 contains the Conclusions and Future work.

2. Flowchart of Parallel Transfer Learning

The proposed flowchart targets two key objectives: energy conservation and maintaining ambient temperature stable, while increasing fan speeds can mitigate temperature deviations, it also increases energy consumption. Optimization Pathways include three parallel flows for different combinations, Fig. 1. The first optimization pathway explores transfer learning from GA to PSO, targeting both energy conservation and ambient temperature stability. The process begins with GA, which performs a global search across the solution space by analyzing inputs through a Fuzzy Inference System (FIS). GA identifies promising solutions for energy consumption and temperature control, leveraging fuzzy rules to balance energy use with environmental comfort. Once GA establishes a baseline solution, PSO takes over to refine it. PSO adjusts particle positions based on the global best and local best solutions, improving convergence speed and solution accuracy. This transfer learning process generates a new FIS surface that differs from those produced by GA and PSO individually. Finally, the pathway concludes with a variance analysis to evaluate whether the optimized

solution meets the desired performance criteria for energy efficiency and temperature stability.

The second optimization pathway explores transfer learning from GA to (NSGA-II, focusing on balancing energy conservation and ambient temperature stability. GA initiates the process by performing a broad search across the solution space, analyzing inputs through a Fuzzy Inference System (FIS) to identify potential solutions for energy efficiency and environmental comfort. GA's exploration establishes a baseline solution, which NSGA-II then refines through multi-objective optimization.

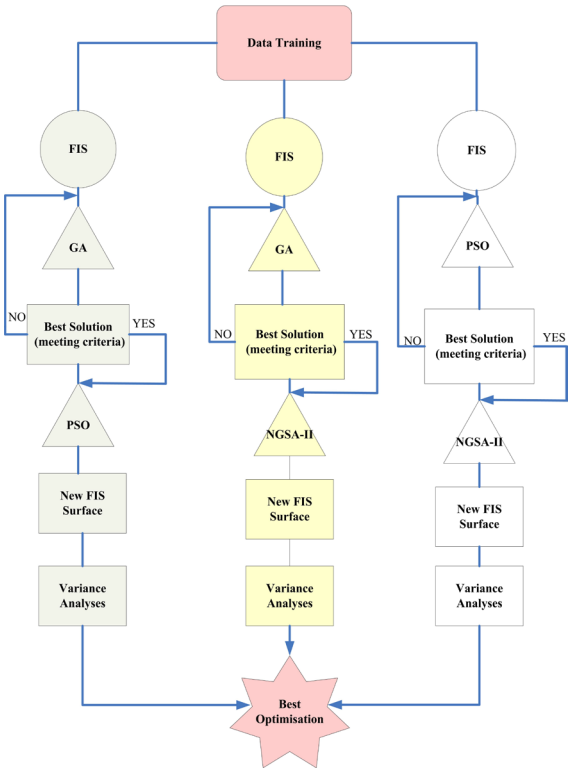


Fig. 1. Flowchart of triple pathways of transfer learning.

NSGA-II improves solution diversity and accuracy using non-dominated sorting and crowding distance mechanisms to maintain a well-distributed Pareto front. The pathway concludes with a variance analysis to assess whether the optimized solution meets the performance criteria for both energy efficiency and temperature control. The third optimization pathway explores transfer learning from PSO to NSGA-II, focusing on optimizing energy conservation and ambient temperature stability. PSO begins the process by rapidly converging to a promising solution through particle adjustments based on the global and local best positions. This enables PSO to identify a near-optimal solution for energy consumption and temperature regulation quickly. NSGA-II then refines this solution by enhancing solution diversity and accuracy using non-dominated sorting and crowding distance mechanisms to generate a well-distributed Pareto front. The pathway concludes with a variance analysis to confirm that the optimized solution meets the target criteria for both energy efficiency and comfort.

To enhance robustness, we integrate default sensor data into the Fuzzy Inference System, creating a training dataset based on real energy consumption data collected over 12 months. This setup ensures a fair comparison across optimization approaches by using a common dataset, consistent population size, and uniform generation count for all optimization paths. This standardized configuration allows for an objective evaluation of each optimization method's performance. The FIS of every step combinations is created using a Mamdani-type fuzzy system, which is one of the most common types of fuzzy inference systems, Fig. 2. In our research, we used three distinct datasets in our optimization process.

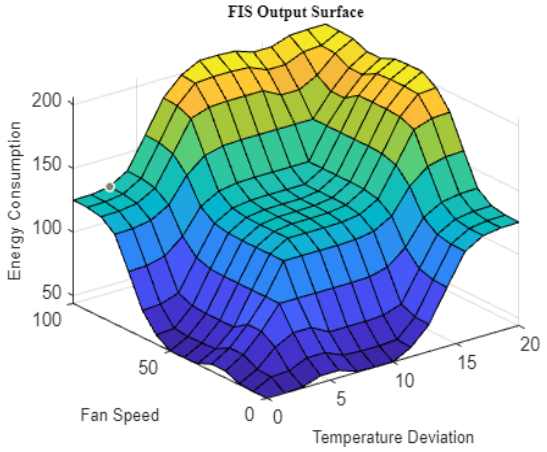


Fig. 2. Before Optimisation FIS surface.

The *first dataset*, which is the real data of IoT consistent across all optimization pathways and transfer learning combinations, serves as a common benchmark for evaluating the overall performance. The *second dataset*, which is generative data for simulations however, is specific to each individual pathway or flowchart, allowing for a more tailored analysis of the optimization process within each transfer learning combination. The *third dataset* is unique for each optimisation pathway but more extended to three years (36 months) of generative data. This approach ensures that each path is evaluated under conditions that reflect its unique characteristics, providing a comprehensive evaluation of each optimization strategy.

3. Simulations Results for Datasets

In this section, we present the simulations results based on two distinct data sources: a 12-months IoT data set for energy consumption selected randomly (Research Data-Dataset of usage pattern and energy analysis of an Internet of Things), [13], a 12-months and 36-months generative data last simulations used

for testing the optimization algorithms behavior and trends.

3.1. Simulations of Real IoT Database

The first 12-month IoT data represents real-world energy consumption (kwatt from fans) data collected from sensors deployed in assisted living environments, providing a realistic foundation for evaluating energy efficiency. This dataset is used consistently across all optimizations approaches to ensure a fair and unbiased comparison, allowing us to measure the performance of each method under identical conditions by using the same IoT dataset for all transfer learning paths, we ensure that the comparison focuses solely on the efficiency and effectiveness of each algorithm, without the influence of differing data sources. We begin with first transfer learning as Table 1 from GA to PSO.

Table 1. Variance of Watt Consumed from GA to PSO.

IoT DATA (Month)	Energy consum	GA	PSO	GA to PSO
M.01	268	150	169	166
M.02	243	150	201	199
M.03	105	129	100	99
M.04	211	147	146	151
M.05	177	150	131	126
M.06	108	150	205	205
M.07	137	129	64	97
M.08	209	134	91	99
M.09	269	133	118	126
M.10	211	145	106	103
M.11	113	132	104	99
M.12	301	150	166	166
TOT	**2352**	**1699**	**1600**	**1637**

The first column represents the actual energy consumption for every month recorded from real-world data. The total real energy consumption is 2352, serving as a baseline for comparison against the optimization outcomes. The total energy consumption using GA is 1699, which reflects a 27.7 % reduction compared to the real consumption. This demonstrates that GA effectively explores the search space and identifies potential solutions for energy reduction, but its tendency to converge too soon to optimal limits affects its overall efficiency. PSO achieves a total energy consumption of 1600, reflecting a 32 % reduction from the real energy consumption. This suggests that PSO's ability to refine solutions through particle adjustments and swarm-based exploration leads to more efficient energy use than GA alone. The GA to PSO combination achieves a total energy consumption of 1637, reflecting a 30.4 % reduction from the real data. While this is slightly higher than PSO alone, the transfer learning process helps improve solution quality and velocity by enriching the results more effectively.

As the *second path* of transfer learning, we have chosen GA to NGSA. Since GA and NSGA-II share a similar nature as evolutionary algorithms, GA to NSGA-II transfer learning powers the global search capability of GA and the multi-objective optimization strength of NSGA-II. While GA explores the search space to identify possible solutions, NSGA-II refines these solutions by balancing multiple conflicting objectives, such as energy consumption as in Table 2. The total energy consumption using NSGA-II is 879, reflecting a significant 62.6 % reduction from the real energy consumption. NSGA-II outperforms GA significantly due to its ability to handle multi-objective trade-offs more effectively, as we mentioned in our previous work [1]. The GA to NSGA-II combination achieves a total energy consumption of 798, which reflects a 66.1 % reduction from the real data and outperforms both GA and NSGA-II individually.

Table 2. Variance of Watt Consumed from GA to NGSA.

IoT DATA (Month)	Energy consum	GA	NGSA	GA to NGSA-II
M.01	268	150	49	68
M.02	243	150	65	68
M.03	105	129	68	49
M.04	211	147	106	49
M.05	177	150	49	58
M.06	108	150	129	50
M.07	137	129	120	74
M.08	209	134	49	102
M.09	269	133	49	63
M.10	211	145	52	50
M.11	113	132	80	88
M.12	301	150	66	80
TOT	**2352**	**1699**	**879**	**798**

Furthermore, this combination is very fast in execution because NSGA-II benefits from the pre-explored solution space provided by GA, allowing it to converge quickly to an optimal solution.

As the *third path* of transfer learning we have chosen the optimisation from PSO to NGSA-II. PSO accelerates the initial convergence due to its fast exploitation of one desired objective function of the search space, meanwhile, NSGA-II improves solution diversity to the final solutions across multiple objectives that in our case are energy consumption and temperature deviation. The combined approach helps in avoiding premature convergence while improving the overall quality of the solution set as in Table 3.

The transfer learning combination of PSO to NSGA-II provides the highest reduction at 75.17 %. and this represents a 43.21 % increase in energy savings compared to PSO alone and a 12.54 % increase over NSGA-II alone.

Table 3. Variance of Watt Consumed from PSO to NGSA.

IoT DATA (Month)	Energy consum	PSO	NGSA	GA to NGSA-II
M.01	268	169	49	49
M.02	243	201	65	49
M.03	105	100	68	49
M.04	211	146	106	49
M.05	177	131	49	49
M.06	108	205	129	49
M.07	137	64	120	49
M.08	209	91	49	49
M.09	269	118	49	49
M.10	211	106	52	49
M.11	113	104	80	49
M.12	301	166	66	49
TOT	**2352**	**1600**	**879**	**584**

3.2. Simulations of Generative 12 Months Database

The second 12 months of generative data is different for every case of transfer learning. It does employ different generative data for each transfer learning combination, so we can evaluate how well each optimization method adapts to and performs under various data conditions, simulating different energy consumption patterns and challenges over time. In order to be fair in comparison we stimulate the three paths as above with the same consistency and same simulator. We begin with first transferring learning from GA to PSO as in Table 4.

Table 4. Variance of Watt Consumed from GA to PSO.

Gen. DATA (Month)	Energy consum	GA	PSO	GA to PSO
M.01	144	116	113	114
M.02	101	122	82	79
M.03	204	125	163	163
M.04	66	107	55	54
M.05	118	124	91	95
M.06	184	125	149	152
M.07	84	110	66	66
M.08	199	125	161	160
M.09	114	118	93	94
M.10	183	125	145	149
M.11	83	107	66	66
M.12	192	125	153	148
TOT	**1673**	**1429**	**1337**	**1340**

SO reduces energy consumption by 6.43 % over GA, while GA-to-PSO achieves a similar 6.23 % reduction. PSO is slightly more energy-efficient, but GA-to-PSO converges faster due to GA's strong initial exploration and the behavior is exactly as with real data. The next transfer learning path is same as we chose real data from Ga to NGSA-II, Table 5. NGSA reduces energy consumption by 17.63 % compared to GA, while GA-to-NGSA improves it further with an 18.30 % reduction. GA-to-NGSA performs slightly better because GA generates a well-diversified initial population, allowing NGSA to refine the solution more efficiently and converge faster. Final transfer learning for this type of dataset we explored the optimization from PSO to NGSA as in Table 6.

Table 5. Variance of Watt Consumed from GA to NGSA.

Gen. DATA (Month)	Energy consum	GA	NGSA	GA to NGSA
M.01	18	84	90	104
M.02	254	125	107	94
M.03	42	92	90	90
M.04	187	124	90	90
M.05	255	125	99	98
M.06	180	125	90	91
M.07	158	125	90	90
M.08	38	84	99	90
M.09	273	125	94	91
M.10	226	125	90	102
M.11	201	125	107	92
M.12	167	119	90	93
TOT	**2000**	**1377**	**1134**	**1125**

Table 6. Variance of Watt Consumed from GA to NGSA.

Gen. DATA (Month)	Energy consum	PSO	NGSA	PSO to NGSA
M.01	196	173	104	91
M.02	196	157	90	90
M.03	218	175	90	91
M.04	61	51	90	90
M.05	196	159	91	90
M.06	246	194	91	91
M.07	114	89	92	90
M.08	192	138	94	90
M.09	74	57	90	90
M.10	249	200	97	90
M.11	253	195	90	90
M.12	192	125	153	148
TOT	**2187**	**1713**	**1173**	**993**

PSO to NGSA shows the highest reduction in energy consumption 54.5 % compared to PSO. NGSA is also more efficient than PSO, with a 46.5 % reduction in energy consumption. PSO is slower compared to NGSA and PSO-to-NGSA. The transfer from PSO to NGSA accelerates convergence by refining the initial exploration provided by PSO, resulting in faster optimization and better energy efficiency.

3.3. Simulations of Generative 36 Months Database

Final approach we chose to explore is the 36 month generative data tailored to each specific combination

of optimization algorithms. This database consists of synthetic data generated to mimic the dynamics of energy consumption over a longer period, providing more varied and extensive scenarios for each optimization path. This generative dataset is used to assess the flexibility and scalability of optimization methods, simulating longer-term trends and potential future behavior of energy consumption. It also provides more robust and thorough analysis of how the algorithms behave under different scenarios, ensuring the validity and reliability of the results. As showing the overall table with data it is out of the interest, we will only display results of three pathways of transfer learning exactly as approaches above in order to explore the behavior of optimisations as Table 7.

Table 7. Energy Consumption Comparison for three transfer learning pathways.

Optimizers (36 month)	Combin. 1	Combin. 2	Combin. 3
No FIS Optimized	5363	5896	5475
GA	4713	-	4468
PSO	4278	4691	-
NGSA	-	2340	2393
GA-PSO	4301	-	-
PSO-NGSA	-	1773	-
GA-NGSA	-	-	2387

As observed across all datasets analyzed, PSO to NGSA in Combination 2 delivers the highest energy reduction of 50.11 %, making it the most efficient learning transfer overall. This is followed closely by Combination 3 (GA, NGSA, GA-NGSA), which demonstrates slightly lower total energy consumption than NGSA alone. However, Combination 1 (GA, PSO, GA-PSO), while exhibiting a slightly less efficient performance compared to GA alone, stands out for its significantly faster execution time, a consistent trend across all datasets tested.

4. Analysis of Variance (ANOVA) and Optimization of Surfaces

In this section, we begin by examining the FIS surfaces as part of the first pathway in transferring learning from GA to PSO. The primary goal of this analysis is to assess how the transfer of knowledge from GA to PSO impacts the optimization surface, particularly in terms of energy efficiency and overall performance.

We first outline the key factors influencing the FIS surfaces, including the membership functions, rule sets, and their impact on the optimization process. Next, we analyze the resulting surfaces for both GA and PSO. The transfer learning approach, which adapts the FIS model initially optimized by GA to the PSO framework, is then evaluated in terms of its ability to refine the optimization process and achieve better performance. This analysis will help identify whether the transfer leads to improvements in energy consumption and overall system efficiency and will provide insights into how the learning from GA can be leveraged to enhance the PSO-based optimization process.

The first FIS surface optimized by GA alone over original FIS is shown in Fig. 3. As a global search method, GA is typically expected to explore widely, producing a surface with moderate roughness but good exploration of the search space. Meanwhile PSO can be smoother due to its particle interactions and exploration balance, Fig. 4. The PSO surface might show less roughness, especially in cases of well-tuned inertia and cognitive/social parameters. In the context of transfer learning, the GA to PSO surface can show a mix of GA's exploratory behavior and PSO's refinement. PSO narrows down the solution space explored by GA, resulting in higher solution density around optimal points and improved consistency, Fig. 5. GA identifies solution regions, and PSO enhances this by performing a more focused search within those regions, leading to optimal solutions closer to each other.

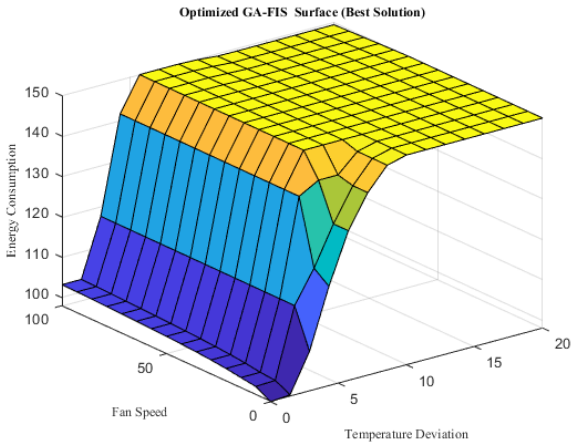


Fig. 3. GA Optimised FIS Surface.

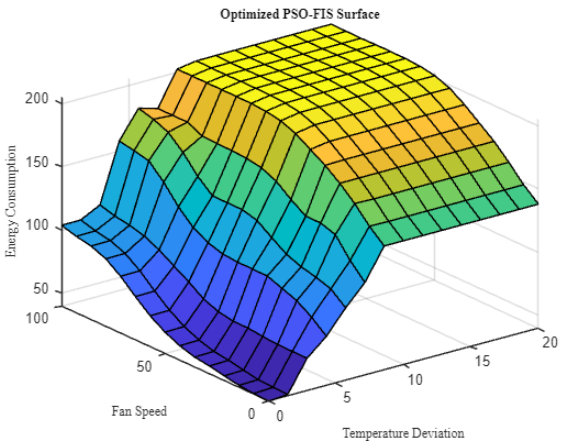


Fig. 4. PSO Optimised FIS Surface.

As a second path of transfer learning, we have chosen to explore the GA to NGSA. FIS surface by NGSA only as Fig. 6 shows smoothness and how gradually the objective function changes in the solution space.

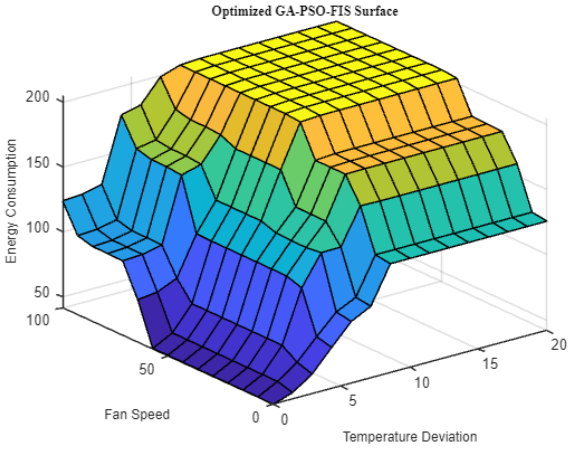


Fig. 5. GA to PSO Optimised FIS Surface.

A smooth surface is stable and easy to optimize, with unchanging improvements. As we can see due to its multi-objective nature, NGSA can have rough surfaces, especially in the beginning, as it searches for a diverse set of solutions. But in our scenarios, we transfer the best solution of GA to NGSA, and we have a smooth surface well from beginning as in Fig. 7. The surface is less rough because GA helps guide the search toward better starting points, reducing large fluctuations in the objective function.

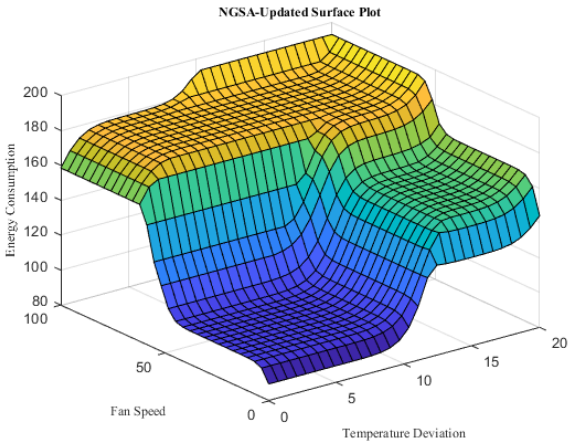


Fig. 6. NGSA Optimised FIS Surface.

One of the key advantages of GA to NGSA transfer learning is that NGSA helps to maintain diversity in the Pareto front, which is crucial for ensuring a better range of optimal solutions. By preserving diversity in this way, NGSA ensures that the final Pareto front is wider and more evenly distributed. This is important because it allows decision-makers to see a broader range of potential solutions, representing different trade-offs between objectives, and making it easier to choose a solution that best meets their needs as see in Fig. 8. In simpler terms, NGSA makes sure the solutions are spread out well across the front, rather than bunched up in just one area, which results in a better overall optimization. In the context of PSO to NGSA transfer learning combined with FIS, the optimization surface's roughness and smoothness evolve in a notable way. PSO alone exhibits roughness causin' the optimization surface to fluctuate, particularly in the early stages of the search, as particles explore the solution space, Fig. 4.

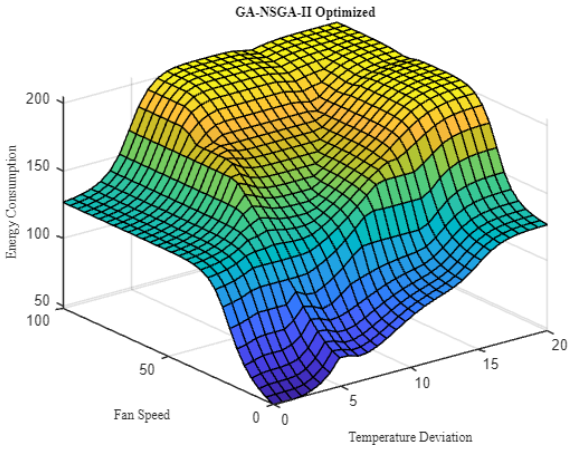


Fig. 7. GA to NGSA Optimised FIS Surface.

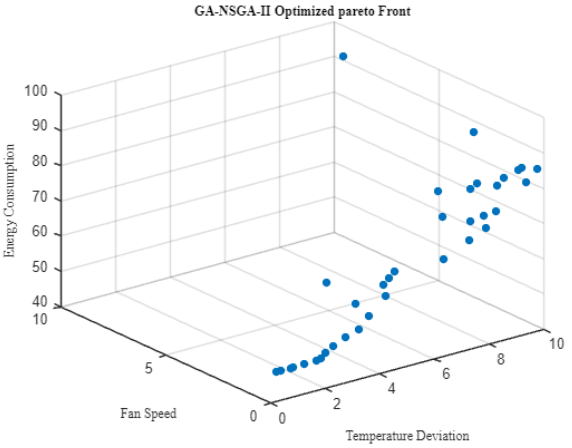


Fig. 8. Pareto front solutions for GA to NGSA.

However, when we apply transfer learning from PSO to NGSA, the process shifts. PSO first explores the solution space, identifying promising solutions, and then these solutions are passed to NGSA for further refinement. The optimization surface becomes smoother as NGSA fine-tunes the solutions, reducing the fluctuations observed in PSO's exploratory phase. This process results in a more stable surface shown in Fig. 9.

The Pareto front obtained from the PSO to NSGA-II transfer learning approach shows a broader

and more evenly distributed spread, as illustrated in Fig. 10. This is due to NSGA-II's ability to maintain solution diversity and empower a well-balanced distribution along the front. By utilizing NSGA-II's non-dominated sorting mechanism and loading distance preservation, the approach effectively relives premature convergence and offers a more comprehensive exploration of the solution space. As a result, the Pareto front not only captures the trade-offs between competing objectives but also demonstrates a more stable convergence toward optimal solutions, leading to improved overall performance in balancing accuracy and efficiency.

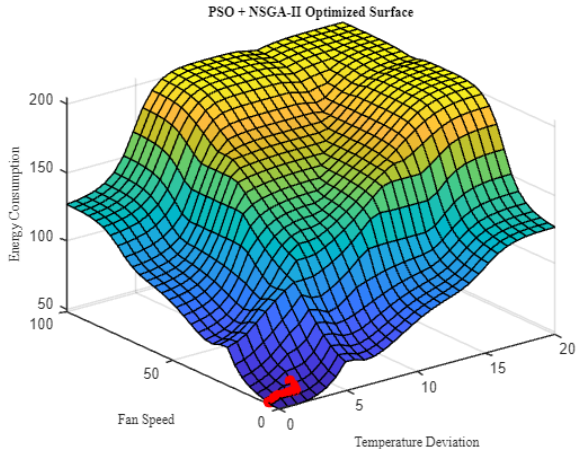


Fig. 9. PSO to NGSA Optimised FIS Surface.

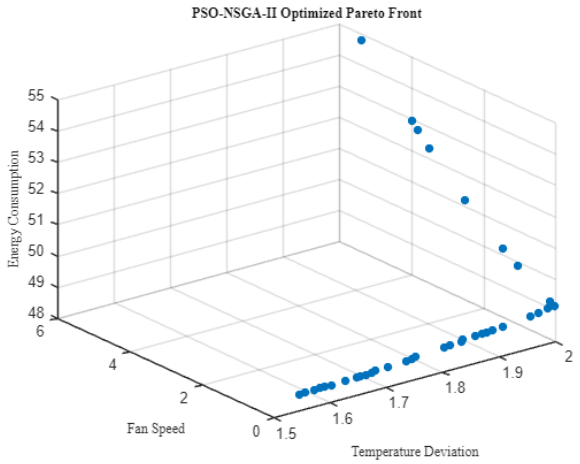


Fig. 10. Pareto front solutions for PSO to NGSA.

5. Conclusions

This research aimed to optimize energy management and occupant comfort in assisted living environments by using dynamic Fuzzy Inference System (FIS) controllers. At our best knowledge, most existing studies focus only on optimizing FIS controllers' parameters or improving the fuzzy rules and not using transfer learning to improve adaptability. The research focused on adapting the controllers to dynamic energy consumption patterns and occupant behavior through transfer learning techniques. Through the incorporation of transfer learning techniques such as GA to PSO, GA to NSGA-II, and PSO to NSGA-II we have shown that leveraging prior optimization experiences enhances the performance of FIS systems. The ability to reuse past optimization cycles significantly improves the adaptability of these optimizers, enabling them to respond more effectively to dynamic changes in energy usage patterns and occupant behavior.

Our results showed that using transfer learning from GA (Genetic Algorithm) to PSO (Particle Swarm Optimization) resulted in slightly lower performance compared to using PSO alone. However, the transfer learning approach was faster than PSO by itself. Even though the performance was a bit lower, the loss was minimal compared to the time saved. In terms of energy conservation, the small decrease in performance was not significant when considering the faster execution time, making this approach better for real-time applications where speed is important.

Meanwhile using NGSA after GA resulted in significantly better performance than using GA or NGSA alone. The execution time was faster, even outperforming GA while using the same number of generations and population size for a fair comparison.

The best performance in this study was achieved by transferring learning from PSO to NGSA. The ability to make real-time decisions using faster execution cycles is highly relevant for assisted living environments where rapid responses are crucial. This approach was highly effective in conserving energy, with energy consumption reduced to approximately half compared to energy conservation alone. The only drawback was the initial execution time for PSO, but once that step was completed, the transition to NGSA proceeded smoothly and efficiently, leading to significant improvements in performance without further delays. While there are challenges related to execution time, the results suggest that further improvements and optimizations in the transfer learning process could make these systems more adaptable and efficient in real-time applications.

Future work will focus on improving execution speed and exploring advanced optimization strategies to enhance the scalability and performance of energy management systems in assisted living environments. This includes developing a hierarchical FIS structure to enable multi-level optimization at both the individual room and building levels, ensuring more adaptive control. Additionally, NSGA-II maybe will be extended to handle many-objective optimization, incorporating inputs such as fan speed to balance complex trade-offs between energy savings, occupant comfort, and operational cost.

References

[1]. A. Xhemali, E. Zanaj, G. Basha, L. Balliu, Enhancing energy efficiency prediction in assisted living through

GA-FIS, PSO-FIS, and NGSA-II-FIS: a comparative evaluation, in *Proceedings of the 5th Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'25)*, 2025, pp. 68-73.

[2]. W. Guo, Y. Dong, Enhancing energy-efficient building design: a multi-agent-assisted MOEA/D approach for multi-objective optimization, *Energy Informatics*, Vol. 7, Issue 1, 2024, 22.

[3]. M. Yang, Y. Liu, X. Shu, Q. Zhang, et al., A multi-objective particle swarm optimization based on maintaining dynamic population size, *Research Square*, 2023 (preprint).

[4]. Z. Zhang, Multi-objective optimization method for building energy-efficient design based on multi-agent-assisted NSGA-II, *Energy Informatics*, Vol. 7, Issue 1, 2024, 10.

[5]. M. Jiang, Z. Huang, L. Qiu, W. Huang, et al., Transfer learning based dynamic multiobjective optimization algorithms, *IEEE Transactions on Evolutionary Computation*, Vol. 22, Issue 4, 2018, pp. 501-514.

[6]. N. Modi, M. Khare, K. Chaturvedi, Performance analysis of load frequency control in single area power system using GA and PSO based PID controller, *International Journal of Electrical, Electronics and Computer Engineering*, Vol. 2, Issue 1, 2013, pp. 108-114.

[7]. N. El Menbawy, H. A. Ali, M. S. Saraya, A. M. T. Ali-Eldin, et al., Energy-efficient computation offloading using hybrid GA with PSO in internet of robotic things environment, *Journal of Supercomputing*, Vol. 79, Issue 17, 2023, pp. 20076-20115.

[8]. C. Bai, Z. Yang, Optimization method of building energy efficiency design based on decomposition multi objective and agent assisted model, *Sustainable Buildings*, Vol. 7, 2024, 4.

[9]. J. Zou, F. Lin, S. Gao, G. Deng, et al., Transfer learning based multi-objective genetic algorithm for dynamic community detection, *arXiv preprint*, 2021, arXiv:2109.15136.

[10]. P. Xu, Z. Deng, J. Wang, Q. Zhang, et al., Transfer representation learning with TSK fuzzy system, *IEEE Transactions on Fuzzy Systems*, Vol. 29, Issue 3, 2021, pp. 649-663.

[11]. M. Xie, Optimization of fuzzy regression transfer learning using genetic algorithm for cross-domain mapping, *WSEAS Transactions on Computers*, Vol. 22, 2023, pp. 316-323.

[12]. J. Lu, H. Zuo, G. Zhang, Fuzzy multiple-source transfer learning, *IEEE Transactions on Fuzzy Systems*, Vol. 28, Issue 12, 2020, pp. 3418-3431.

[13]. Y. Luo, S. Gao, Research Data-Dataset of usage pattern and energy analysis of an Internet of Things-enabled ceiling fan, V3, *Mendeley Data*, 2023.

Theory:

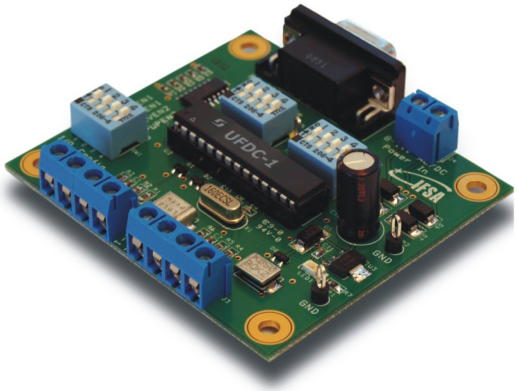
Digital Sensors and Sensor Systems: Practical Design

and

Practice:

Development Board EVAL UFDC-1/UFDC-1M-16

+



Buy book and Evaluation board together. **Save 30.00 EUR**

Development Board EVAL UFDC-1 / UFDC-1M-16

Full-featured development kit for the Universal Frequency-to-Digital Converters UFDC-1 and UFDC-1M-16. 2 channel, 16 measuring modes, high metrological performance, RS232/USB interface, master and slave communication modes. On-board frequency reference (quartz crystal oscillator). Operation from 8 to 14 V AC/DC. Development board software is included.

All existing frequency, period, duty-cycle, time interval, pulse-width modulated, pulse number and phase-shift output sensors and transducers can be directly connected to this 2-channel DAQ system. The user can connect TTL-compatible sensors' outputs to the Development Board, measure any output frequency-time parameters, and test out the sensor systems functions.

Applications:

- Digital sensors and sensor systems
- Smart sensors systems
- Data Acquisition for frequency-time parameters of electric signals
- Frequency counters
- Tachometers and tachometric systems
- Virtual instruments
- Educational process in sensors and measurements
- Remote laboratories and distance education

Order online:
http://www.sensorsportal.com/HTML/BOOKSTORE/Digital_Sensors_and_Board.htm

IFSA

International Frequency Sensor Association Publishing

www.sensorsportal.com

ISSN 1726- 5479