

SENSORS & TRANSDUCERS

ISSN 1726-5479

vol. 264
1 / 24



Sensors, Transducers and Actuators

International Frequency Sensor Association Publishing



Sensors & Transducers

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

Volume 264, Issue 1, February 2024

Editor-in-Chief

Prof., Dr. Sergey Y. YURISH



IFSA Publishing: Barcelona • Toronto

Sensors & Transducers is an open access journal which means that all content (article by article) is freely available without charge to the user or his/her institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author. This is in accordance with the BOAI definition of open access. Authors who publish articles in *Sensors & Transducers* journal retain the copyrights of their articles. The *Sensors & Transducers* journal operates under the Creative Commons License CC-BY.

Notice: No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

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





Editors-in-Chief: Professor, Dr. Sergey Y. Yurish, tel.: +34 696067716, e-mail: editor@sensorsportal.com

Editors for Western Europe

Meijer, Gerard C.M., Delft Univ. of Technology, The Netherlands
Ferrari, Vittorio, Università di Brescia, Italy
Mescheder, Ulrich, Univ. of Applied Sciences, Furtwangen, Germany

Editor for Eastern Europe

Sachenko, Anatoly, Ternopil National Economic University, Ukraine

Editors for North America

Katz, Evgeny, Clarkson University, USA
Datskos, Panos G., Oak Ridge National Laboratory, USA
Fabien, J. Josse, Marquette University, USA

Editor for Africa

Maki K., Habib, American University in Cairo, Egypt

Editors South America

Costa-Felix, Rodrigo, Inmetro, Brazil
Walsole de Reca, Noemi Elisabeth, CINSO-CITEDEF
UNIDEF (MINDEF-CONICET), Argentina

Editors for Asia

Ohyama, Shinji, Tokyo Institute of Technology, Japan
Zhengbing, Hu, Huazhong Univ. of Science and Technol., China
Li, Gongfa, Wuhan Univ. of Science and Technology, China

Editor for Asia-Pacific

Mukhopadhyay, Subhas, Massey University, New Zealand

International Advisory Board

Abdul Rahim, Ruzairi, Universiti Teknologi, Malaysia
Abramchuk, George, Measur. Tech. & Advanced Applications, Canada
Aluri, Geetha S., Globalfoundries, USA
Ascoli, Giorgio, George Mason University, USA
Atalay, Selcuk, Inonu University, Turkey
Atghiaee, Ahmad, University of Tehran, Iran
Augutis, Vyantas, Kaunas University of Technology, Lithuania
Ayesh, Aladdin, De Montfort University, UK
Baliga, Shankar, B., Laser Components DG, Inc., USA
Barlingay, Ravindra, Larsen & Toubro - Technology Services, India
Basu, Sukumar, Jadavpur University, India
Bousbia-Salah, Mounir, University of Annaba, Algeria
Bouvet, Marcel, University of Burgundy, France
Campanella, Luigi, University La Sapienza, Italy
Carvalho, Vitor, Minho University, Portugal
Changhai, Ru, Harbin Engineering University, China
Chen, Wei, Hefei University of Technology, China
Cheng-Ta, Chiang, National Chia-Yi University, Taiwan
Cherstvy, Andrey, University of Potsdam, Germany
Chung, Wen-Yaw, Chung Yuan Christian University, Taiwan
Cortes, Camilo A., Universidad Nacional de Colombia, Colombia
D'Amico, Arnaldo, Università di Tor Vergata, Italy
De Stefano, Luca, Institute for Microelectronics and Microsystem, Italy
Ding, Jianning, Changzhou University, China
Djordjević, Alexander, City University of Hong Kong, Hong Kong
Donato, Nicola, University of Messina, Italy
Dong, Feng, Tianjin University, China
Erkmen, Aydan M., Middle East Technical University, Turkey
Fezari, Mohamed, Badji Mokhtar Annaba University, Algeria
Gaura, Elena, Coventry University, UK
Gole, James, Georgia Institute of Technology, USA
Gong, Hao, National University of Singapore, Singapore
Gonzalez de la Rosa, Juan Jose, University of Cadiz, Spain
Goswami, Amarjyoti, Kaziranga University, India
Guillet, Bruno, University of Caen, France
Hadjiloucas, Sillas, The University of Reading, UK
Hao, Shiyang, Michigan State University, USA
Hui, David, University of New Orleans, USA
Jaffrezic-Renault, Nicole, Claude Bernard University Lyon 1, France
Jamil, Mohammad, Qatar University, Qatar
Kaniusas, Eugenijus, Vienna University of Technology, Austria
Kim, Min Young, Kyungpook National University, Korea
Kumar, Arun, University of Delaware, USA
Lay-Ekuakille, Aime, University of Lecce, Italy
Li, Fengyuan, HARMAN International, USA
Li, Jingsong, Anhui University, China
Li, Si, GE Global Research Center, USA
Lin, Paul, Cleveland State University, USA
Liu, Aihua, Chinese Academy of Sciences, China
Liu, Chenglian, Long Yan University, China
Liu, Fei, City College of New York, USA
Mahadi, Muhammad, University Tun Hussein Onn Malaysia, Malaysia
Mansor, Muhammad Naufal, University Malaysia Perlis, Malaysia

Marquez, Alfredo, Centro de Investigacion en Materiales Avanzados, Mexico
Mishra, Vivekanand, National Institute of Technology, India
Moghavvemi, Mahmoud, University of Malaya, Malaysia
Morello, Rosario, University "Mediterranea" of Reggio Calabria, Italy
Mulla, Imtiaz Sirajuddin, National Chemical Laboratory, Pune, India
Nabok, Aleksey, Sheffield Hallam University, UK
Neshkova, Milka, Bulgarian Academy of Sciences, Bulgaria
Pal, Jitendra, Carnegie Mellon University, USA
Passaro, Vittorio M. N., Politecnico di Bari, Italy
Patil, Devidas Ramrao, R. L. College, Parola, India
Penza, Michele, ENEA, Italy
Pereira, Jose Miguel, Instituto Politecnico de Seteбал, Portugal
Pillarsetti, Anand, Sensata Technologies Inc, USA
Pogacnik, Lea, University of Ljubljana, Slovenia
Pullini, Daniele, Centro Ricerche FIAT, Italy
Qiu, Liang, Avago Technologies, USA
Reig, Candid, University of Valencia, Spain
Restivo, Maria Teresa, University of Porto, Portugal
Rodríguez Martínez, Angel, Universidad Politécnica de Cataluña, Spain
Sadana, Ajit, University of Mississippi, USA
Sadeghian Marnani, Hamed, TU Delft, The Netherlands
Sapozhnikova, Ksenia, D. I. Mendeleev Institute for Metrology, Russia
Singhal, Subodh Kumar, National Physical Laboratory, India
Shah, Kriyang, La Trobe University, Australia
Shi, Wendian, California Institute of Technology, USA
Shmaliy, Yuriy, Guanajuato University, Mexico
Song, Xu, An Yang Normal University, China
Srivastava, Arvind K., Systron Donner Inertial, USA
Stefanescu, Dan Mihai, Romanian Measurement Society, Romania
Sumridetchkajorn, Sarun, Nat. Electr. & Comp. Tech. Center, Thailand
Sun, Zhiqiang, Central South University, China
Sysoev, Victor, Saratov State Technical University, Russia
Thirunavukkarasu, I., Manipal University Karnataka, India
Thomas, Sadiq, Heriot Watt University, Edinburgh, UK
Tian, Lei, Xidian University, China
Tianxing, Chu, Research Center for Surveying & Mapping, Beijing, China
Vanga, Kumar L., ePack, Inc., USA
Vazquez, Carmen, Universidad Carlos III Madrid, Spain
Wang, Jiangping, Xian Shiyu University, China
Wang, Peng, Qualcomm Technologies, USA
Wang, Zongbo, University of Kansas, USA
Xu, Han, Measurement Specialties, Inc., USA
Xu, Weihe, Brookhaven National Lab, USA
Xue, Ning, Agiltron, Inc., USA
Yang, Dongfang, National Research Council, Canada
Yang, Shuang-Hua, Loughborough University, UK
Yaping Dan, Harvard University, USA
Yue, Xiao-Guang, Shanxi University of Chinese Traditional Medicine, China
Xiao-Guang, Yue, Wuhan University of Technology, China
Zakaria, Zulkarnay, University Malaysia Perlis, Malaysia
Zhang, Weiping, Shanghai Jiao Tong University, China
Zhang, Wenming, Shanghai Jiao Tong University, China
Zhang, Yudong, Nanjing Normal University, China

Contents

Volume 264
Issue 1
February 2024

www.sensorsportal.com

ISSN 2306-8515
e-ISSN 1726-5479

Research Articles

Deep Learning and the Modern Radar <i>Webert Montlouis</i>	1
Security of Field Devices in Future Water Management <i>Pekka Koskela and Markku Kylänpää</i>	11
A MEMS Based Large Travel Range Linear Actuator with a Good Minimum Step Size to Be Used in Mirror Shape Correction Applications <i>Ehsan Atash Zaban and Mehmet Kusaf</i>	19
A Novel Large Travel, High Precision MEMS Based Electro-thermal Linear Actuator for Adjusting CCD Camera Lens <i>Ehsan Atash Zaban and Mehmet Kusaf</i>	26
Application of Bragg Light Diffraction for Determination of Acoustic Wave Frequency <i>F. R. Akhmedzhanov</i>	34

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:
<http://www.sensorsportal.com/HTML/DIGEST/Submission.htm>

10th International Conference on Sensors and Electronic Instrumentation Advances

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

25-27 September 2024
Ibiza (Balearic Islands), Spain



The 10th annual SEIA' 2024 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 2024 edition scheduled for 25-27 September in Ibiza, Spain, alongside the 6th IFSA Frequency and Time Conference and the 6th 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' 2024 also emphasizes networking, featuring social events like a welcome cocktail, Gala Dinner, and farewell cocktail.

Sensor Technology

- Accelerometers
- Inclometers
- 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' 2024. 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):
Notification of acceptance:
Registration:
Camera ready and late registration:

1 June 2024
20 June 2024
15 July 2024
31 August 2024



10th International Conference on Sensors and Electronic Instrumentation Advances

10th SEIA'24
IBIZA

<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 one of affiliated open access journal.
- 3) The limited number of full-page papers published in the journals will be selected by the Editorial Board to extend to open access book chapters for the Book Series on 'Advances in Sensors: Reviews', Vol. 9, 'Advances in Biosensors: Reviews', Vol. 4 or 'Advances in Measurements and Instrumentation: Reviews', Vol. 3. These Book Series are indexed in the Book Citation Index (Web of Science) by Clarivate Analytics.

Conference Hotel

The Ibiza Twins Hotel
Av. Pere Matutes Noguera, 79, 07800 Ibiza, Spain

Social Programme

- Welcome Cocktail: 24 September 2024 (20:00-21:30).
- Gala Dinner: 26 September 2024 (20:00-23:30)
- Farewell Cocktail: 27 September 2024 (during the poster session)

Organizing Committee

Chairman

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

Advisory Chairmen

Prof. Vijyakumar Varadarajan (The University of New South Wales, 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)



Deep Learning and the Modern Radar

Webert MONTLOUIS

Johns Hopkins University, Baltimore MD, 21218, USA

Tel.: +1 617-820-4254

E-mail: wmontlouis@jhu.edu

Received: 29 September 2023 /Accepted: 10 January 2024 /Published: 28 February 2024

Abstract: For decades, radar Systems have been essential for detecting and tracking moving targets, such as aircraft, ships, and vehicles. Their tasks increase in difficulty with the development of more complex targets. A radar system operates by transmitting a high-frequency electromagnetic signal that travels through the air and bounces back when it hits an object. By measuring the time it takes for the signal to return, the radar can determine the distance to the object and its speed. Traditionally, radar signals are processed using classical signal processing techniques, such as filtering, matched filtering, and pulse compression. These methods are generally effective but require significant computational resources, particularly when dealing with large datasets or complex signals. As a result, conventional radar systems that rely on direct sampling of Radio Frequency (RF) signals can be too computationally intensive to obtain accurate range estimates within the required time frame. To address these limitations, recent research has focused on applying machine learning techniques for radar signal processing to process the radar return at machine speed. One such approach is the use of neural networks to perform sequence-to-sequence classification on radar signals. The neural network can be trained using a dataset of radar signals, allowing it to learn patterns and features that are indicative of different types of targets. In addition, the introduction of high-speed and highly maneuvering targets necessitates that the response time of the data and signal processor be fast and optimized to meet the close look track of the system requirements. This study's results show that using a trained neural network can significantly improve the accuracy and efficiency of radar range estimation. The neural network approach can accurately detect targets within the radar system's field of regard and estimate their ranges with an overall accuracy of 99 %, making it a promising alternative to traditional radar signal processing techniques. The use of machine learning techniques, particularly neural networks, has the potential to revolutionize the way radar signals are processed and analyzed. The promising results of this study highlight the importance of exploring innovative approaches to radar range estimation and signal processing, particularly in applications where real-time detection and tracking of targets are critical.

Keywords: Radar range estimation, Artificial intelligence, Machine learning, Deep learning, Neural networks, Sequence-to-sequence classification, Temporal Convolutional Network (TCN).

1. Introduction

Recent advancements in artificial intelligence have presented an exciting opportunity to revolutionize various engineering fields by providing new approaches to solving complex problems. With the

emergence of deep learning, there has been a proliferation of innovative network architectures that have significantly enhanced the field's performance and reliability. In fact, deep learning has become a ubiquitous technology integrated into many applications, ranging from vehicle traffic management

[1], medical [2, 3], wireless communications, the internet, retail, finance, and more. Notably, deep learning is particularly useful in applications where computational complexity is a challenge to traditional methods. This technology has already significantly impacted various areas and is expected to continue driving innovation in engineering in the coming years.

Radar systems engineering has a rich history spanning several decades and plays a crucial role in both commercial and military applications. For instance, in commercial applications, airport radar [4-6] facilitates the efficient management of passenger planes' arrivals and departures by providing real-time information on their range and altitude. Similarly, in military applications, radar systems are employed to detect, track, classify, and discriminate targets on the battlefield, with resource managers controlling the various modes of operation.

Radar systems are complex and composed of multiple subsystems [7, 8], including the receiver exciter, data processor, signal processor, and resource manager. This paper specifically focuses on the signal processor subsystem, its vital role, the conventional signal processing approach [7, 9], and recent advancements. The primary aim is to explore innovative techniques that could simplify the signal processor architecture and report target detection quicker than the traditional matched filtering approach.

By exploring novel techniques in signal processing, this paper seeks to identify strategies that can improve radar systems' efficiency, speed, and accuracy, particularly in target detection. Overall, this investigation could lead to the development of more advanced and reliable radar systems with enhanced capabilities that can effectively meet the needs of various applications.

In recent years, the field of radar systems engineering and design has experienced significant transformation thanks to the remarkable advances in CMOS technology [10-12]. These developments have led to the creation of more reliable Radio Frequency (RF) components, thereby enhancing system performance while simplifying receiver/exciter design. In particular, the latest analog-to-digital converters now possess a faster sampling rate [13-15], creating the opportunity to reduce the number of down-conversion stages required, which in turn gives rise to a more efficient receiver design at a reduced cost.

However, despite the significant improvements in receiver design, there are some drawbacks in the data and signal processor, primarily related to data throughput. To address this issue, one potential solution is the introduction of Artificial Intelligence (AI), especially deep learning. Deep learning can leverage the power of advanced algorithms and neural networks to improve data processing speed, accuracy, and efficiency [16-18].

Moreover, deep learning can enable the development of highly sophisticated algorithms capable of detecting patterns in radar signals that

traditional signal processing approaches might overlook [19-21]. By doing so, AI can enhance the performance and accuracy of radar systems in target detection, tracking, and classification. These recent advancements in AI have shown promising results and have the potential to revolutionize radar systems engineering and design.

Deep learning is a type of artificial intelligence and machine learning that uses neural networks to analyze and make predictions about specific types of data [16]. This technology is referred to as "deep" learning because the neural networks employed can have multiple layers, which allows for more sophisticated data processing and classification.

Deep learning has numerous applications, including facial recognition, natural language processing, and autonomous driving systems. In the context of radar systems, deep learning can be used to enhance the accuracy and efficiency of radar signal processing. Specifically, deep learning techniques can be applied to perform sequence-to-sequence classification on radar signals, which can be used to detect, track, and classify various types of targets.

The use of deep learning in radar systems has the potential to revolutionize the field, providing more accurate and efficient detection and classification capabilities. For example, deep learning can be used to identify subtle variations in radar signals that are indicative of specific targets, which traditional signal processing techniques might miss. Additionally, deep learning can be used to analyze large datasets generated by radar systems to extract valuable insights and identify patterns that are difficult for humans to discern. Overall, the application of deep learning in radar systems is an exciting development with significant potential for improving the performance and capabilities of these critical technologies.

Sequence-to-sequence classification is a widely used technique for analyzing temporal data, where each time step in a sequence is assigned a label. In this study, the researchers focus on classifying time steps in a reflected radar signal and use a neural network called the Temporal Convolutional Network (TCN) to perform this classification. TCN is a powerful deep learning architecture that has shown remarkable results in processing sequential data, including speech recognition and natural language processing.

The researchers aim to evaluate the efficacy of TCN in the context of radar signal processing. Specifically, they want to determine the accuracy of the network's predictions and assess whether it is a suitable candidate for this task. The results of this study may have practical implications for the use of TCN in other areas, such as medical signal processing, finance, and environmental monitoring.

This research is significant because radar signal processing is a complex and challenging task, and the development of accurate classification methods could have important applications, such as in national security and weather forecasting. By assessing the accuracy of the network's predictions, the researchers hope to contribute to the development of more

accurate and efficient radar signal processing techniques.

Overall, this study is an important step in exploring the potential of TCN in the context of radar signal processing. By evaluating the accuracy of the network's predictions and assessing its suitability for this task, the researchers aim to provide valuable insights into the effectiveness of this method and guide future research in this area.

The structure of the paper is organized as follows. In Section 2, the radar data model is formulated, which provides a clear understanding of the input data and the information it contains. In Section 3, the Temporal Convolutional Network (TCN) architecture used in the study is presented, which explains the network's design, structure, and key features. In Section 4, the training methodology used to train the TCN is described, including details on the dataset preparation, hyperparameter tuning, and optimization techniques used. Section 5 discusses the conventional radar signal processing. In Section 6, the deep learning signal processing is presented.

Section 7 presents the simulation results of the study, which includes the evaluation of the network's performance and the comparison of results against traditional methods. This section also includes an analysis of the model's strengths and limitations and how it can be improved in future work. Finally, in Section 8, the paper concludes with a summary of the research findings and the implications of the study, including potential applications and future directions for research in this area.

The paper is structured to provide a clear and concise understanding of the research process, including the formulation of the data model, the design of the TCN architecture, the training methodology, and the simulation results. By following this logical flow, the paper provides a comprehensive view of the study and its contributions to the field of radar signal processing.

2. Radar Model

The radar signal data model that will be used in the simulation is based on the constant low-velocity target model. The radar uses a linear frequency modulated (LFM) signal. The transmit signal for a single pulse is modeled by the signal

$$s(t) = u(t)e^{j2\pi f_0 t} = \text{rect}\left(\frac{t}{T}\right)e^{j2\pi(f_0 t + \frac{1}{2}Kt^2)}, \quad (1)$$

where $u(t)$ is the envelope of the LFM waveform, T is the pulsewidth, and K is the linear modulation rate. This transmit signal is then down-converted to a baseband signal given by

$$s_b(t) = \text{rect}\left(\frac{t}{T}\right)e^{j\pi Kt^2} \quad (2)$$

When the signal is incident upon a target, since the constant, low-velocity target model is being used, this

target is taken as static while the wave is reflected. Because of this, the reflected baseband signal is modeled by the equation

$$s_{rb}(t) = \text{rect}\left(\frac{t-t_r}{T}\right)e^{j\pi Kt^2} + N(\sigma), \quad (3)$$

where $t_r = 2R/c$ is the time delay for the signal to propagate to the target and back to the radar receiver, and $N(\sigma)$ is the random Gaussian noise of the environment. Since the detection of the target is of interest here, it is necessary to set up the return signal intensity for the single pulse system.

The intensity of the return will be based on the radar range equation (RRE). For the single pulse case, the RRE is given by

$$SNR = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 k T_0 B F L}, \quad (4)$$

where:

P_t is the Peak transmit power;

G is the Array gain;

λ is the Transmit signal wavelength;

σ is the Target radar cross section;

k is the Boltzman constant;

B is the Radar bandwidth;

F is the Noise figure;

L is the Total system losses.

It is assumed that a 13 dB signal-to-noise ratio is needed on the return for a general detection.

3. Temporal Convolutional Network Architecture

In recent years, deep learning has been a game-changer in many fields, including signal processing. Radar return signals, in particular, pose a significant challenge due to their large amounts of temporal data. This is particularly true when one considers reducing receiver design complexity by sampling the return signal at RF. As such, researchers have explored various deep-learning architectures to improve the processing of these signals, with the TCN architecture being one of the most promising.

The TCN architecture is a type of convolutional neural network (CNN) that incorporates dilated convolutions to capture long-range dependencies in the input signal. Compared to other network architectures, TCN offers several advantages, such as better memory size control and easy adaptability. The network can efficiently process long historical sequences, which is critical for our task of classifying radar signals.

The primary objective of our network is to perform sequence-to-sequence classification on the radar signal, as specified in Section 2. The sequence-to-sequence classification task involves labeling each time step in the input signal as "target" or "noise." To achieve this, we will employ a dilated causal convolutional layer.

The dilated causal convolutional layer is a type of convolutional layer that operates in a causal manner, meaning it only operates on past and present time steps Fig. 1. This characteristic makes it ideal for our task of classifying radar signals, as we only need to label the target and noise time steps based on past and present information. The layer's dilation parameter can be adjusted to expand its receptive field and capture long-range dependencies in the signal, allowing us to extract meaningful temporal features.

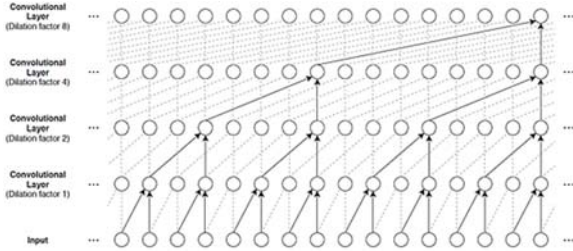


Fig. 1. Dilated Convolutional Layer (Photo Provided by MathWorks).

Overall, the combination of the TCN architecture and dilated causal convolutional layer is expected to significantly enhance our network's ability to classify radar signals accurately. By capturing important temporal features in the signal and efficiently processing long historical sequences, we hope to improve our understanding of radar return signals and support a variety of applications, including target detection, tracking, and recognition.

The defining feature of the convolutional layer is receptive field size. Receptive field size is defined as

$$R = (f - 1)(2^k - 1) + 1 \quad (5)$$

The temporal convolutional network (TCN) is a deep learning model that is particularly well-suited for processing time-domain signals, such as those commonly encountered in radar data analysis. The key idea behind TCNs is to use a series of stacked convolutional layers to extract features from a signal over time while retaining information on long-range dependencies between different temporal patterns. This is achieved by using dilated convolutions, which enable the receptive field of each convolutional layer to grow exponentially with each additional layer.

The two main hyperparameters that can be adjusted to optimize the performance of a TCN are the filter size (f) and the number of stacked convolutional layers (k). By increasing f , one can increase the size of the receptive field, which in turn allows the network to capture more temporal context and potentially improve classification accuracy. However, if the receptive field becomes too large, the network may start to encounter unnecessary data and experience performance degradation. On the other hand, increasing k allows for more complex feature

extraction but can also lead to overfitting and increased computational cost.

In addition to the convolutional layers, TCNs typically include several other layers that are used to improve training efficiency and prevent overfitting. Normalization layers are used to normalize the weights of the parameters in the network, which can help to prevent vanishing or exploding gradients during backpropagation. Spatial dropout layers are used to randomly zero out certain parameters in each layer during training, which helps to reduce the network's reliance on any particular set of features and mitigate the risk of overfitting. The dropout factor hyperparameter associated with each spatial dropout layer determines the proportion of parameters that are zeroed out. Finally, ReLU (rectified linear unit) layers are commonly used as activation functions, which introduce non-linearity into the network and help to capture complex patterns in the data.

The TCN architecture typically consists of a series of "sub-blocks," each of which contains multiple convolutional layers and other associated layers [22, 23]. In the architecture described, there are four main sub-blocks, each of which contains two convolutional layers, as well as normalization, dropout, and ReLU layers. The input radar signal is first passed through the initial sub-block, which contains a 1x1 convolution and skip connection to match the channel dimensions of the subsequent convolutional layers. The signal is then passed through the next three sub-blocks before reaching a fully connected layer, a SoftMax activation function layer, and finally, the classification layer Fig. 2. The number of sub-blocks used is another hyperparameter that must be tuned based on the specific data model being used.

TCNs are a powerful and flexible deep learning architecture that can be used for a wide range of time-domain signal processing tasks [24-26]. By adjusting the various hyperparameters and network architecture components, one can optimize the network's performance for a given dataset and classification task.

The novel radar signal processing technique is characterized by a two-step approach that integrates deep learning and conventional signal processing methodologies. In the initial step, a trained deep learning network rapidly determines the presence of a target or facilitates its detection, thereby yielding a swift response time. The range of the target is deduced in this step. Subsequently, in the second step, conventional signal processing techniques are employed in the vicinity of the target location to extract supplementary features, utilizing only a limited portion of the data. As an illustration, Section 7 demonstrates the utilization of a Range Doppler Map (RDM) to extract the Doppler associated with each target range.

Given the substantial volume of data generated by the employment of a fast ADC and high sampling rate, it is imperative for the signal processor to efficiently process the data stream in real-time to achieve optimal system performance.

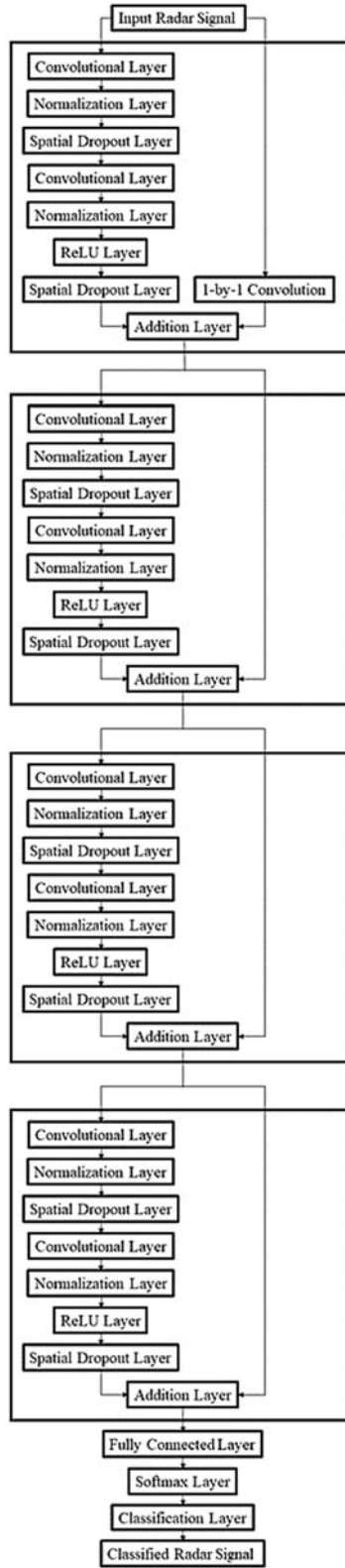


Fig. 2. TCN Architecture.

4. Training Methodology

The process of training a neural network involves feeding it with large amounts of data, and the quality and quantity of the data used can have a significant impact on the performance of the network. For this

particular project, we used radar signals as the training data to train the neural network to detect targets within a range of 100 to 1000 km.

To ensure optimal training results, we varied the ranges of the targets defined in the simulation. By doing so, we were able to generate a training dataset that would provide the neural network with the necessary input to learn how to detect any target within the range band specified. This approach allowed us to cover a wide range of scenarios that the neural network might encounter in real-world applications.

To further optimize the training process, we experimented with two different training methods, the uniform method and the random method. In the uniform method, we generated 10 radar signals with targets located at the lower limit of 100 km. We then incremented the range using a step size equal to the range resolution of the radar system. This approach ensured that the network was accurately trained at every incremental range, allowing it to detect targets with high precision. In contrast, the random method employed a larger and more diverse dataset of radar signals. This approach allowed the neural network to learn to detect targets across a wider range of scenarios and conditions. By doing so, the network was able to generalize better and detect targets within the range band specified more accurately.

Through the use of carefully selected training data and different training methods, we were able to successfully train the neural network to detect targets within a range of 100 to 1000 km with a high degree of accuracy. This achievement has important implications for various applications, such as military and security, where the ability to detect targets from a distance is crucial. This range was then incremented using a step size equal to the range resolution of the radar system, given by

$$\Delta R = \frac{c}{2B}, \quad (6)$$

where c denotes the propagation speed of light, and B denotes the modulation bandwidth of the radar. Again, 10 more radar signals were generated at this range to use as training data. This process is repeated until the upper range limit of 1000 km is reached, training the network along the way.

When it comes to training a neural network, choosing the right training method can significantly impact the model's accuracy, efficiency, and adaptability to different environments. In the case of radar signal detection, two main methods have been experimented with - the uniform method and the random method.

The uniform method involves generating radar signals with targets at different ranges, incrementing the range by a step size equal to the range resolution of the radar system. While this method provides a way to cover a wide range of values, it can lead to overfitting to lower range values and requires an impractical amount of time to train. In contrast, the random method involves generating a large number of

random signals within the defined range band, which reduces the risk of overfitting and makes the training process much faster.

One significant advantage of the random method is that it allows for greater flexibility in the training process. The random generation of signals means that the neural network can be trained to detect targets at various ranges, which makes it more adaptable to different real-world scenarios. This flexibility is crucial in situations where the radar environment is constantly changing, and the network must be able to adapt quickly to new conditions. Additionally, this method ensures that the network is less susceptible to memorizing specific range values and instead learns to identify targets in a more generalized way.

Another benefit of the random method is that it can help the neural network to become more robust to noise. By generating random signals with varying noise levels, the network can learn to recognize patterns in the radar signal despite the presence of noise. This adaptability is essential in real-world scenarios where the radar environment is often noisy and can significantly impact the accuracy of the detection system.

The random method provides many advantages over the uniform method in training neural networks for radar signal detection. Its flexibility, speed, and ability to adapt to changes in the environment make it a suitable approach for real-world applications. By utilizing this method, the neural network can be trained to detect targets over a wide range of ranges while being robust to noise levels, ensuring that the system performs accurately in various conditions.

5. Conventional Radar Signal Processing

Radar signal processing is a complex and sophisticated field that involves a wide range of techniques and methods to extract useful information from radar signals [27, 28]. Its main objective is to detect and interpret radar signals to identify objects or targets in the radar's detection range and to extract information about their location, speed, and size.

The process of radar signal processing begins with the transmission of electromagnetic waves by the radar transmitter. The waves travel through the air and bounce off objects in the radar's detection range, which then return to the radar receiver as echoes. These echoes are then processed by the signal processor to extract information about the detected objects.

One of the key challenges in radar signal processing is filtering the radar signals to remove unwanted noise and interference. Various filtering techniques can be used, such as low-pass, high-pass, and band-pass filters, to remove frequencies outside of the radar's bandwidth or filter out specific types of noise, such as electromagnetic interference.

Another important technique used in radar signal processing is target detection and tracking. This involves identifying and tracking objects in the radar's

detection range by analyzing the radar signals reflected off of them. Target detection algorithms can use various signal processing techniques, such as matched filtering, to distinguish between signals from targets and noise.

Doppler processing is another important technique used in radar signal processing to determine the velocity of a target by analyzing the frequency shift of the radar signals reflected off of it. Pulse compression techniques, such as frequency-modulated continuous wave (FMCW) radar, can increase the range resolution of the radar system by compressing the radar pulse width. Range gating techniques can also be used to isolate radar signals from specific ranges, allowing for more accurate target detection and tracking.

Radar signal processing is a critical component of modern radar systems, and the techniques used play a crucial role in the accuracy, reliability, and efficiency of radar-based detection and tracking systems. As technology advances, the field of radar signal processing continues to evolve, with new techniques and methods being developed to improve the performance and capabilities of radar systems.

6. Deep Learning Signal Processing

Deep learning signal processing combines the power of deep learning with traditional signal processing techniques to process and analyze signals. Alternatively, a Deep learning network can be trained to replace traditional signal processing [29-31]. In the context of sensor signal processing, deep learning techniques can be used to extract features from signals, such as radar signals, that are difficult to detect using traditional signal processing techniques. For example, a deep neural network can be trained to learn the characteristics of radar signals that are associated with specific types of targets or interference and then use this knowledge to detect and classify signals in real-time.

One advantage of deep learning in signal processing is that it can learn complex patterns and relationships in the data, even when the signal processing algorithms do not explicitly define them [32, 33]. This can lead to improved accuracy and reliability of signal detection and classification. Another advantage is that deep learning can be used to process signals in real-time, which is critical in applications such as radar systems for autonomous vehicles, where fast and accurate signal processing is essential for safe operation. More specifically, this is a game changer when it comes to data management and processing in large phased array radars.

Deep learning signal processing is still a relatively new field, and there are many challenges and opportunities for research and development. Some of the key challenges include developing neural network architectures that are optimized for signal processing tasks, dealing with issues such as data scarcity and transfer learning, and understanding the underlying principles and trade-offs involved in using deep

learning for signal processing. Despite these challenges, deep learning signal processing holds great promise for a wide range of applications, from improving the accuracy and reliability of radar systems to enhancing the capabilities of speech and image recognition systems.

7. Performance Analysis

Radar is a crucial technology used in a wide range of applications, including air traffic control, military surveillance, weather forecasting, and automotive safety systems. The performance of a radar system can be evaluated based on several metrics, each of which plays a vital role in determining its effectiveness and accuracy [7, 27].

The range is one of the essential metrics that affect early detection and the maximum distance at which a radar can detect and track a target for given system parameters and radar cross-section. The range depends on various factors, including the transmit power of the radar, the antenna gain, and the sensitivity of the receiver. Higher transmit power and antenna gain result in an increased range for the radar, enabling it to detect targets that are located farther away. Resolution is another crucial metric that measures the radar's ability to distinguish between two targets that are located close to each other. The resolution is determined by the radar's bandwidth, pulse duration, and the number of pulses transmitted. A higher bandwidth and a shorter pulse duration allow the radar to distinguish between targets that are located closer to each other.

Sensitivity is also an essential metric that refers to the minimum detectable signal that a radar can detect. A radar with higher sensitivity can detect weaker signals and is better at detecting small targets. This is particularly important in applications such as weather forecasting, where the radar needs to detect small particles, such as raindrops or snowflakes, and the military where the goal is to avoid detection. Accuracy measures how well a radar can estimate the position, velocity, and other characteristics of a target. Accuracy depends on various factors, including the radar's calibration, signal processing algorithms, and environmental conditions. A radar with high accuracy can provide more precise information about the target, enabling better decision-making in applications such as air traffic control or military surveillance.

Target detection and tracking capability refer to the radar's ability to detect and track targets in real-time. The performance of a radar in this area depends on the radar's signal processing algorithms, tracking filters, and other advanced techniques used for target tracking. A radar with high target detection and tracking capability can accurately track multiple targets simultaneously, even in challenging environments such as clutter and jamming. Largely, the performance of a radar system depends on its design, operating frequency, and the specific application for which it is used. Modern radars use

advanced signal processing techniques, such as adaptive beamforming and waveform diversity, to improve their performance in various applications. These techniques allow the radar to adapt to changing environmental conditions and provide more accurate information about the targets being tracked.

A. Single Target Detection.

Single target detection is an essential process in many radar applications, and the selection of the appropriate radar system and signal processing techniques depends on the specific requirements of the application. To detect a single target, the radar emits a signal and listens for the echoes that bounce back from the target. The radar system analyzes these echoes to determine the distance, direction, and speed of the target.

There are several techniques used for single target detection, including pulse radar, continuous wave radar, and frequency-modulated continuous wave radar. Pulse radar is the most common type of radar used for single target detection. It transmits a short burst of energy, called a pulse, and then listens for the echo. By measuring the time delay between the transmitted pulse and the received echo, the radar can determine the distance to the target [27]. Continuous wave radar, on the other hand, transmits a continuous wave of energy and measures the frequency shift of the reflected wave. This frequency shift, known as the Doppler shift, provides information about the speed of the target. Frequency-modulated continuous wave (FMCW) radar is another technique used for single target detection. FMCW radar transmits a signal with a continuously changing frequency. The radar system analyzes the frequency shift of the reflected signal to determine the distance to the target.

In addition to these techniques, there are several signal processing algorithms used for single target detection, such as matched filtering and thresholding [8]. Matched filtering is a technique that filters out noise and enhances the signal to detect weak echoes from a single target. Thresholding is a technique that sets a threshold level below which signals are considered noise and above which signals are considered a target.

B. Multiple Target Detection.

Multiple target detection refers to the ability of a radar system to detect and track multiple targets simultaneously within its coverage area. This is a crucial capability for many radar applications, including air traffic control, military surveillance, weather monitoring, and more. To achieve multiple target detection and extract additional target features, radar systems typically use a combination of signal processing techniques such as pulse compression, frequency diversity, angle of arrival estimation, and range doppler map Fig. 3. These techniques help to enhance the signal-to-noise ratio of radar returns and to separate the signals from different targets in the radar's field of view. One common approach to multiple target detection is to use a phased array radar system. Phased array radar employs multiple antenna elements that can be individually controlled to steer

the radar beam in different directions. This allows the radar to scan a large area rapidly and to track multiple targets simultaneously [9].

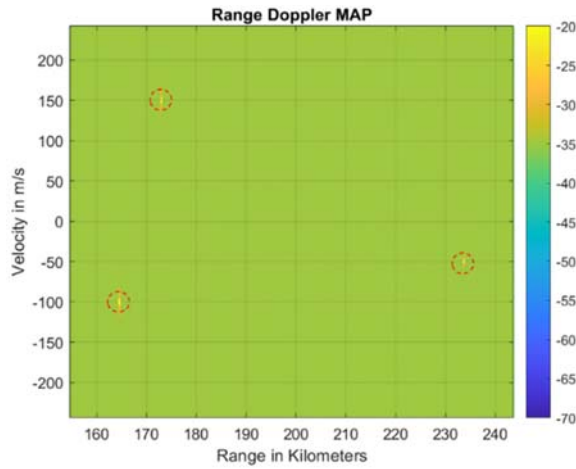


Fig. 3. Range Doppler Map showing three targets as the second step in the signal processor following target detection from the trained deep learning network.

Another approach to multiple target detection is to use advanced signal processing algorithms such as adaptive beamforming and waveform diversity. Adaptive beamforming can help to reduce interference from other radar sources or jamming signals, while waveform diversity can help to improve the detection and tracking of low-observable targets.

Overall, radar multiple target detection is a complex and challenging problem requiring advanced signal processing techniques, radar hardware design, and system integration. However, it is a critical capability for many radar applications and essential for ensuring safety and security in many domains.

C. Probability of Detection.

Radar probability of detection (P_d) is a measure of the likelihood that a radar system will detect a target that is present in its field of view. P_d is an important metric in radar performance evaluation and is typically expressed as a percentage or a decimal between 0 and 1.

The probability of detection depends on a variety of factors, including the radar's transmit power, antenna gain, receiver sensitivity, and signal processing algorithms. In general, higher transmit power and antenna gain will improve the radar's ability to detect targets at longer ranges and in adverse weather conditions. Similarly, more sensitive receivers and advanced signal processing algorithms can help to detect weaker radar returns and improve target discrimination.

In addition to hardware and software factors, the probability of detection also depends on the characteristics of the target itself, such as its size, shape, and reflectivity. Larger targets with higher radar cross-sections (RCS) will generally be easier to detect than smaller or low-observable targets.

To estimate the probability of detection for a particular radar system and target scenario, radar engineers typically use computer simulation or empirical testing. Simulation models can help to predict the expected performance of a radar system under different operating conditions Fig. 4, while testing can provide more accurate measurements of actual radar performance.

Radar probability of detection is a key performance parameter that is critical for many radar applications, including air traffic control, military surveillance, weather monitoring, and more. Improving the probability of detection is an ongoing challenge for radar engineers and requires a combination of hardware and software advancements, as well as a deep understanding of target characteristics and radar system behavior.

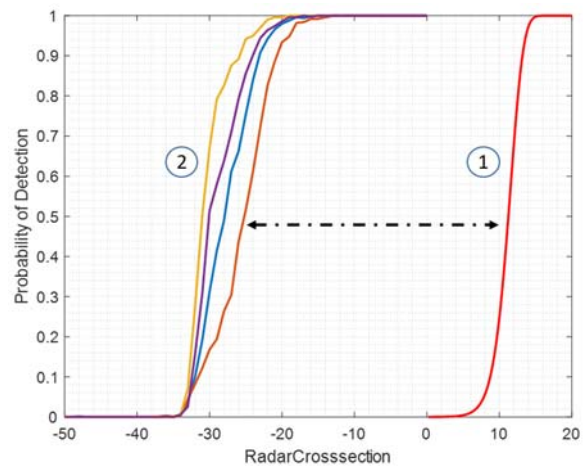


Fig. 4. Probability of detection curves for four probability of false alarms [1e-4 1e-6 1e-7 1e-8] and (1) the theoretical probability of detection curve in conventional radars. The trained deep-learning network shows considerable improvement in target detection for the same system parameters.

8. Conclusions

The paper introduces a state-of-the-art method for detecting radar targets, which utilizes deep learning and employs a sequence-to-sequence evaluation with a temporal convolutional network. The approach is novel and offers several advantages over traditional radar target detection methods. One key advantage is that recent advances in faster analog-to-digital converters have made it possible to simplify radar receiver designs, increase efficiency, and reduce costs.

However, these advances come with some challenges, such as high computational complexity for the signal processor when sampling the radar return at RF. The proposed detection scheme overcomes these challenges and demonstrates that real-time target detection on raw radar data can be achieved without the need for preprocessing. The experimental results show that this approach achieves high accuracy and

can be easily integrated into existing radar signal processors.

This method is particularly significant as it enables accurate detection of targets in complex and dynamic environments where traditional methods may fail to perform. Additionally, this approach offers flexibility in the design of radar systems, allowing for better adaptation to diverse applications. Overall, this work represents a major contribution to the field of radar target detection and offers a promising direction for future research in this area.

References

- [1]. V. Katariya, M. Baharani, N. Morris, O. Shoghli, H. Tabkhi, DeepTrack: lightweight deep learning for vehicle trajectory prediction in highways, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, Issue 10, 2022, pp. 18927-18936.
- [2]. S. Yang, *et al.*, Learning COVID-19 pneumonia lesion segmentation from imperfect annotations via divergence-aware selective training, *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, Issue 8, 2022, pp. 3673-3684.
- [3]. L. A. Thomaz, S. M. M. Faria, L. M. N. Távora, L. Marcenaro, Automatic medical image diagnosis: highlights from the 2021 IEEE 5-minute video clip contest [SP competitions], *IEEE Signal Processing Magazine*, Vol. 39, Issue 3, 2022, pp. 130-134.
- [4]. T. Martelli, C. Bongioanni, F. Colone, P. Lombardo, L. Testa, A. Meta, Security enhancement in small private airports through active and passive radar sensors, in *Proceedings of the 17th International Radar Symposium (IRS'16)*, 2016, pp. 1-5.
- [5]. P. Lanzkron, E. Brookner, Solid state X-band airport surface surveillance radar, in *Proceedings of the IEEE Radar Conference*, 2007, pp. 670-676.
- [6]. H. Ghadaki, R. Dizaji, Target track classification for airport surveillance radar (ASR), in *Proceedings of the IEEE Conference on Radar*, 2006.
- [7]. D. K. Barton, *Radar Equations for Modern Radar*, Artech House, 2012.
- [8]. B. R. Mahafza, *Radar Systems Analysis and Design Using MATLAB*, 3rd Ed., CRC Press, 2013.
- [9]. M. A. Richards, *Fundamentals of radar signal processing*, 2nd Ed., McGraw-Hill Education LLC, Chicago, Ill., 2014.
- [10]. T. B. Ngo, Q. H. Do, S. W. Yoon, A wideband circulator leakage canceler for retro-directive RF system, *IEEE Microwave and Wireless Components Letters*, Vol. 32, Issue 10, 2022, pp. 1211-1214.
- [11]. S. Park, S. Hong, Millimeter-wave wideband differential four-way Wilkinson power divider with 90° rotational symmetric layout, *IEEE Microwave and Wireless Components Letters*, Vol. 32, Issue 10, 2022, pp. 1163-1166.
- [12]. F. Yu, *et al.*, A low phase noise VCO with placement reordered cores, high-Q bond wire array inductors, *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 69, Issue 9, 2022, pp. 3754-3758.
- [13]. V. Hiremath, S. Ren, An ultra high speed encoder for 5GSPS Flash ADC, in *Proceedings of the IEEE Instrumentation & Measurement Technology Conference Proceedings*, 2010, pp. 136-141.
- [14]. T. Li, Y. Zhang, Y. Ni, An ultra-fast ADC linearity test and calibration method, in *Proceedings of the IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID'19)*, 25-27 Oct. 2019, pp. 282-285.
- [15]. Y. Zhang, T. Li, Z. Huang, Y. Ni, D. Fu, High-gain and fast-setting OTA for high speed and high resolution ADC, in *Proceedings of the IEEE International Conference on Electron Devices and Solid State Circuits (EDSSC'18)*, 6-8 June 2018, pp. 1-2.
- [16]. H. C. Kaskavalci, S. Gören, A deep learning based distributed smart surveillance architecture using edge and cloud computing, in *Proceedings of the International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML'19)*, 26-28 Aug. 2019, pp. 1-6.
- [17]. H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, M. D. Vos, SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 27, Issue 3, 2019, pp. 400-410.
- [18]. Y. Tang, J. Xu, K. Matsumoto, C. Ono, Sequence-to-sequence model with attention for time series classification, in *Proceedings of the IEEE 16th International Conference on Data Mining Workshops (ICDMW'16)*, 12-15 Dec. 2016, pp. 503-510.
- [19]. Y. Deng, L. Wang, H. Jia, X. Tong, F. Li, A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance, *IEEE Transactions on Industrial Informatics*, Vol. 15, Issue 8, 2019, pp. 4481-4493.
- [20]. S. Arniriparian, M. Freitag, N. Cummins, M. Gerczuk, S. Pugachevskiy, B. Schuller, A Fusion of Deep Convolutional Generative Adversarial Networks and Sequence to Sequence Autoencoders for Acoustic Scene Classification, in *Proceedings of the 26th European Signal Processing Conference (EUSIPCO'18)*, 3-7 Sept. 2018, pp. 977-981.
- [21]. T. Ni, X. Ding, Y. Wang, J. Shen, L. Jiang, G. Zhang, Spectrum sensing via temporal convolutional network, *China Communications*, Vol. 18, Issue 9, 2021, pp. 37-47.
- [22]. M. Liu, H. Qin, R. Cao, S. Deng, Short-term load forecasting based on improved TCN and DenseNet, *IEEE Access*, Vol. 10, 2022, pp. 115945-115957.
- [23]. X. Yuan, S. Qi, Y. Wang, K. Wang, C. Yang, L. Ye, Quality variable prediction for nonlinear dynamic industrial processes based on temporal convolutional networks, *IEEE Sensors Journal*, Vol. 21, Issue 18, 2021, pp. 20493-20503.
- [24]. F. Luo, S. Poslad, E. Bodanese, Temporal convolutional networks for multiperson activity recognition using a 2-D LIDAR, *IEEE Internet of Things Journal*, Vol. 7, Issue 8, 2020, pp. 7432-7442.
- [25]. H. Wang, D. Hu, M. Zhang, N. Li, Y. Yang, Multiphase flowrate measurement with multimodal sensors and temporal convolutional network, *IEEE Sensors Journal*, Vol. 23, Issue 5, 2023, pp. 4508-4517.
- [26]. Y. Cheng, Y. Xu, H. Zhong, Y. Liu, Leveraging semisupervised hierarchical stacking temporal convolutional network for anomaly detection in IoT communication, *IEEE Internet of Things Journal*, Vol. 8, Issue 1, 2021, pp. 144-155.
- [27]. W. L. Melvin, Principles of modern radar, in *Radar Applications* (W. L. Melvin, J. A. Scheer, Eds.), Vol. 3, SciTech Publishing Inc., 2013.

- [28]. F. E. Nathanson, Radar Design Principles: Signal Processing and the Environment (J. P. Reilly, M. Cohen, Eds.), 2nd Ed., McGraw-Hill, 1991.
- [29]. Y. Liu, Y. Li, Y. Zhu, Y. Niu, P. Jia, A brief review on deep learning in application of communication signal processing, in *Proceedings of the IEEE 5th International Conference on Signal and Image Processing (ICSIP'20)*, 23-25 Oct. 2020, pp. 51-54.
- [30]. P. Y. Hsu, C. K. Cheng, Arrhythmia Classification using deep learning and machine learning with features extracted from waveform-based signal processing, in *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC'20)*, 20-24 July 2020, pp. 292-295.
- [31]. W. Li, K. Wang, L. You, Z. Huang, A new deep learning framework for HF signal detection in wideband spectrogram, *IEEE Signal Processing Letters*, Vol. 29, 2022, pp. 1342-1346.
- [32]. A. Jagannath, J. Jagannath, T. Melodia, Redefining wireless communication for 6G: signal processing meets deep learning with deep unfolding, *IEEE Transactions on Artificial Intelligence*, Vol. 2, Issue 6, 2021, pp. 528-536.
- [33]. A. M. Elbir, DeepMUSIC: multiple signal classification via deep learning, *IEEE Sensors Letters*, Vol. 4, Issue 4, 2020, pp. 1-4.



Published by International Frequency Sensor Association (IFSA) Publishing, S. L., 2024
(<http://www.sensorsportal.com>)





10 Top Reasons

to Publish Open Access Books with IFSA Publishing

- Indexed in Book Citation Index (Web of Science)
- Copyrights belong to Authors (CC-BY)
- The maximum number of pages is not limited
- Very reasonable publication fees
- High visibility
- All book types accepted
- Available in different formats: electronic and print
- Freely available online
- High quality standards
- Authors benefit from IFSA Membership

https://www.sensorsportal.com/HTML/IFSA_Publishing.htm

Security of Field Devices in Future Water Management

^{1,*} Pekka KOSKELA and ² Markku KYLÄNPÄÄ

¹ VTT Technical Research Centre of Finland Ltd., Kaitoväylä 1, Oulu, 90570, Finland

² VTT Technical Research Centre of Finland Ltd., Kemistintie 3, Espoo, 00240, Finland

¹ Tel.: +358 20 722 111, fax: +358 20 722 7001

* E-mail: Pekka.Koskela@vtt.fi

Received: 13 December 2023 / Accepted: 5 January 2024 / Published: 28 February 2024

Abstract: Water management as a part of critical infrastructure is undergoing transformation alongside the advancement of digitalization. Future water management systems will incorporate both edge and cloud services. Increased connectivity of systems and the use of remote management together with growing heterogeneity and complexity of systems will bring new demands and challenges for security systems. In order to address these future security challenges, we study the zero trust approach and its possible realization with a physical unclonable function facility. Especially in our focus are resource-constrained devices like sensors in the field and their safety.

Keywords: Sensors, Energy efficiency, Zero trust, Physical unclonable function, Hashgraph, Water management.

1. Introduction

In the future, the needs of remote connections will increase, where users use and connect to the water management and control systems remotely. Increased connectivity also enables new services, like remote software updates for Internet of Things (IoT) devices in the field. Monitoring, control, and management operations can also be done over the Internet. These needs must be fulfilled both in normal and in abnormal situations. For example, during a pandemic situation control room and office work was restricted. This also caused changes in work practices in water management and accelerated technology adoption [1].

These requirements and technology development will lead water management systems towards cloud environments, where both edge computing and cloud services will be available. However, this development will also increase attack surface, emphasizing the need to develop holistic security architecture.

There are few security architecture proposals for water management systems. Ntuli and Abu-Mahfouz describe an architecture that is focused on securing

access to IoT devices [2] utilizing common IoT protocols, e.g., Constrained Application protocol (CoAP) and MQ Telemetry Transport (MQTT) and publish-subscribe communications model. The importance of secure boot and secure firmware update is also mentioned. Xia et al. [3] describe a blockchain-based system that is focusing on decentralizing transactions between the actors and is also proposing the use of peer-to-peer (P2P) networks for communications. Lee et al. [4] describe the future water management platform. Their system relies on bi-directional communication, intelligent network, and centralized management of smart water grid.

In this paper, we present our work towards security architecture for future water management systems. The main contributions of our work are:

- Utilization of Physical Unclonable Function (PUF) as basis for immutable device identity;
- Impact of Zero Trust security architecture and its realization;
- Discussion of system integrity verification and secure software update topics.

The rest of this paper is structured as follows. First, in Section 2 we give a brief overview of water management. Next, in Section 3 we introduce a key security primitive called Physical Unclonable Function (PUF). In Section 4 we describe elements of the Zero Trust (ZT) security architecture. Then, in Section 5 we discuss potential threats and limitations of our approach. Finally, in Section 6 we conclude the paper.

2. Water Management

Water management consists of both fresh water supply and distribution and wastewater treatment and disposal. These operations are part of the critical infrastructure of society that should work reliably in every situation. Water management IoT devices are sensors, pumps, and valves. These can be controlled and monitored on site or remotely from a control room. Water management system architecture depends on the size of the management system and the level of digitalization. It is expected that future water management systems utilize multiple new technologies including cloud-based services [5]. This trend has also been recognized by cloud providers [6].

In Fig. 1 we present our generic view of the potential future water management architecture, where remote connections and cloud services, both internet and edge clouds, are used. Some of the field devices can be directly connected to edge services to achieve fast response time or improved security in the sense of reduced attacking surface, where information is used locally near the source. The next operation level devices are connected to the control room with field buses, for instance industrial ethernet (Ethernet/IP, Profinet), Modbus or Profibus. The control room is connected to the Internet so that it is possible to utilize cloud services. However, cloud dependency should be analyzed from a reliability and security point of view. Critical functionality should also work without cloud connectivity. Moreover, the control room can also support remote connections. In cases, when the movement of citizens is restricted, like during the COVID-19 epidemic, it may be necessary to have capability to do remote management outside of the control room. Also digitalization of water management systems and increased heterogeneity of suppliers and devices will increase requirements to have remote connections and to keep the devices up-to-date and secure.

Cloud environments with remote connections will open new attack surfaces and increase potential vulnerabilities of the system.

In that kind of environment, it is not enough to trust firewalls as barriers. Trust between all actors including users, services and applications of the system must be kept up continuously. In order to tackle this challenge, we study exploitation of a zero trust concept together with a physical unclonable function (PUF) concept for future water management.

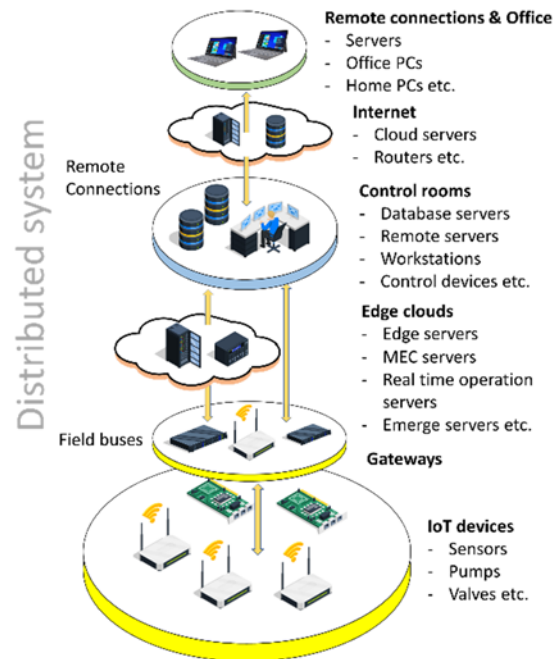


Fig. 1. Future water management architecture.

3. Physical Unclonable Function

PUF (physical unclonable function) is a security technology that creates unique, secure, and unclonable identifiers or keys from the physical characteristics of a device or component. PUFs have the following features: they are robust (they do not change over time), unique (no two PUFs are the same), easy to evaluate (they can be implemented practically), hard to replicate (they cannot be duplicated), and very difficult or impossible to predict (their responses cannot be guessed). The physical characteristics that PUFs use can be based on various types of differences, such as electric, optical, chemical, acoustic, and thermal noise. For example, electric PUFs can originate from differences in bistable state, time, voltage-current, and capacitance [7]. One type of electric PUFs is silicon based PUFs, which use the natural variations and imperfections that occur during the fabrication of integrated circuits or other physical devices. Some examples of silicon-based PUFs are SRAM-PUFs, ring oscillator PUFs, and arbiter PUFs. For instance, SRAM-PUF assigns a value to each SRAM cell based on its unique bistable state when it is powered up. The value can be zero, one, or unstable and unreachable, as shown in Fig. 2.

During manufacturing, the system can create digital fingerprints based on its unique physical properties i.e. PUFs. These fingerprints are the result of applying different challenges (such as bistable state, see above, voltage-current, time etc.) to the system and recording the corresponding responses. When the system faces the same challenge again, it produces the same response, which verifies its identity. PUFs can be classified into weak and strong PUFs, depending on how many challenge-response pairs they can generate.

Weak PUFs have only one or a few pairs, while strong PUFs have many pairs. The main advantage of PUFs is that they are theoretically unclonable because the physical variations that cause them cannot be easily replicated. This makes them useful for various security-critical applications. However, PUFs also have challenges, such as noise, reliability, and the need for proper error correction and management to ensure their practical utility in secure systems. The main purpose of a PUF is to establish a local root of trust in a device or a system, which can then be used for a wide range of applications, such as secure key storage, device authentication, remote attestation, secure bootstrapping, anti-counterfeiting, and more.

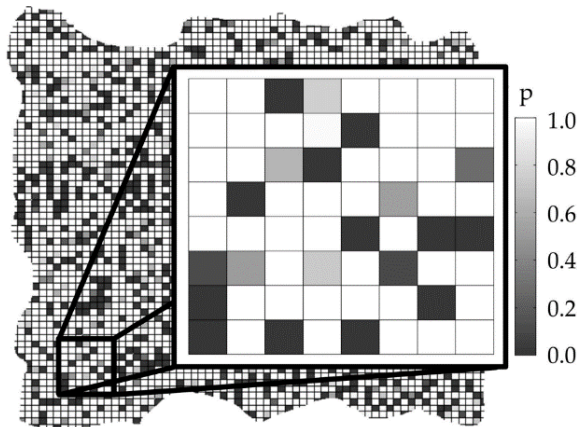


Fig. 2. A 64-bit fingerprint, shown within a larger fingerprint for context. The lightness of the shading of each cell indicates p , the probability of powering-up to 1 [8].

4. Zero Trust

Rose et al. [9] proposed a set of basic principles for Zero Trust Architecture (ZTA). ZTA treats all resources (such as data sources, computing services, and communication channels) as valuable assets that need protection. Access to these resources is granted on a per-session basis, with dynamic policies that consider various factors such as client identity, application/service identity, and behavioral attributes. Authentication and authorization of resources are strictly enforced, and all communication is secured regardless of network location. The enterprise monitors and measures the integrity and security posture of all owned and associated assets, collecting information about the current state of assets, network infrastructure, and communication channels to continuously improve its security posture. Fig. 3 shows a generic architecture model of the zero trust concept and one of its realizations.

Zero trust is based on the idea of sharing resources on a per-session basis, where privileges are minimally shared to perform services and minimize the impact of attacks. To achieve this, the system must have means of isolation and micro-segmentation, such as binding roles to the actors, granular access, provisioning

control, key hierarchy and prohibiting lateral movement. In the next sections, we will discuss possible realizations of zero trust, especially in field devices, where resources are scarce.

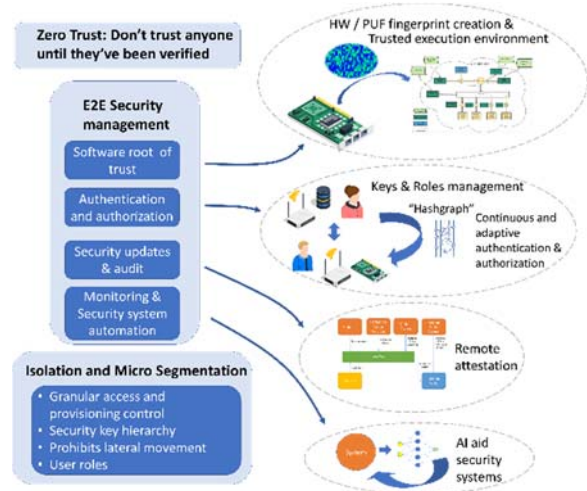


Fig. 3. Zero trust model.

4.1. Root of Trust

A Root of Trust (RoT) is a fundamental concept in computer security that refers to a reliable and secure origin of a computing system. The RoT must be inherently trusted, and it cannot be validated. The RoT can be used to establish a chain of trust that ensures the integrity, authenticity, and confidentiality of system components and applications. Based on the RoT and device certificates the actors can verify the security posture of each other using remote attestation. The actors can also authenticate and authorize to share resources and services. In evolving infrastructure, to keep the system secure, there is a need for a secure update mechanism and system audits monitoring the system's health. A simple RoT can be an immutable code block (e.g., boot ROM implementing secure boot) and a device-specific secret. The code block should guard access to a device-specific secret that is then used to derive the root key of the system also known as a Hardware Unique Key (HUK).

The RoT can also be used in isolated and secure execution areas such as secure enclave, secure element, trusted execution environment (TEE), trusted platform module (TPM) or hardware secure module (HSM). Resource constrained IoT devices should be relatively cheap. Due to the cheapness requirement, additional hardware components like TPM or custom hardware-based security modules are not preferred. TEE-based approaches like ARM TrustZone are integrated to a CPU and are nowadays available also in micro-controller class devices [10]. There are also specific trusted computing concepts for micro-controllers, e.g., the DICE concept [11]. A set of minimal functionality that is required to support remote attestation is discussed in [12].

A device-specific key must be immutable and different in each device instance. It must be kept secret and be directly accessible only from the RoT code. Chip manufacturers have used specific one-time-programmable (OTP) memory areas called electronic fuses to implement a device-specific secret. Early versions were factory-programmable but now most systems are field-programmable. These techniques can raise the bar of attacks but physical attacks utilizing electron microscope are still possible and additional steps are needed in chip manufacturing [13].

A root of trust device-specific secret can also be created with the aid of a PUF concept, where the device secret is built using existing inherent unique physical properties such as small manufacturing variations in silicon chips, which are used to generate unique, unclonable identifiers that can be used to derive cryptographic keys. From a resource point of view, PUF alternatives, which exploit existing hardware, like SRAM (Static Random Access Memory) based, are interesting, because most of the field devices have some form of RAM. In RAM based PUFs each time the SRAM block powers on, the memory cells come up as either 1 or 0. The start-up values create a random and repeatable pattern, which is unique to each chip [14].

In our study, we used the root of trust realization setup called *asvin* [15], the E1 development board from OKDO. This development board has an LPC55S69xx microcontroller from NXP. The microcontroller contains onboard PUF using SRAM to form a unique repeatable pattern. This pattern is further added to an activation code, which is then turned into a digital fingerprint i.e. root of trust device secret, which provides foundation for security subsystems.

4.2. Authentication and Authorization

The ZT principles state that there is no inherent trust among the actors of a system (users, applications, services, and hardware). This means that every actor must have a provable identity for authentication and authorization between each other. To achieve a lightweight authentication protocol, the communication between field devices and authentication services should be minimal and the cryptographic algorithms should be energy efficient such as, based on symmetric encryption. For this purpose, we chose a lightweight authenticated encryption scheme (ASCON) to implement. The ASCON is NIST recently proposed to standardization based on open competition, where addition of security, the selection criterion was performance and flexibility in terms of speed, size and energy use [16, 17]. ASCON can perform authenticated symmetric encryption with associated data (AEAD) and hashing [18]. As the name implies, AEAD ensures the privacy, integrity, and authenticity of ciphertext data and the integrity of unencrypted associated data. AEAD

implements a method where data is first encrypted and then an authentication tag, i.e. message authentication code (MAC), is calculated. Fig. 4 shows the procedure where: 1) The key encrypts messages, and a random number is added to ensure privacy. 2) Calculates an authentication tag (MAC) that ensures that both encrypted and unencrypted parts of a message are not tampered with. Because MAC is done after encryption it makes it possible to exclude all tampered messages before encryption.

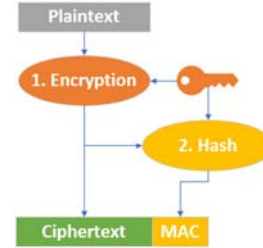


Fig. 4. Authenticated encryption with associated data (AEAD) in encrypt-then-MAC method.

4.3. Key and Role Management

As a part of authentication, we also study a distributed ledger-based approach, where field device identity information is typically stored in a blockchain. The selected ledger implementation depends on the use case. The ledger can be either private or public and either permissionless or permissioned, see Fig. 5.

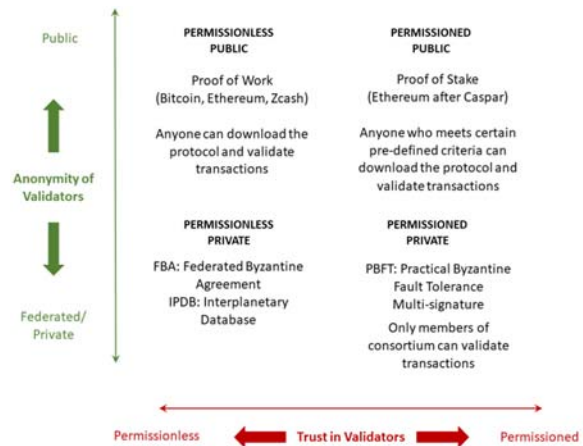


Fig. 5. Type of blockchain [19].

Concerning the nature of critical infrastructure, where access to a network is restricted and only permissioned actors are allowed, the best approach is a private and permissioned ledger. There are numerous consensus mechanisms like proof of work, proof of stake etc., which are popular in public networks but need to be excluded, because those mechanisms are energy and resource hungry [20]. However, there are still many consensus mechanisms based on Byzantine fault tolerance or other voting mechanisms that

perform better in terms of speed, energy consumption, and resource usage. One such example is Redundant Byzantine Fault Tolerance (RBFT) [21], where a preselected primary node creates a new block and other nodes follow only if the primary node is not malicious.

The performance of a consensus protocol has a significant impact on various aspects of the system, including fast transaction speed, quick recovery time from failures, scalability, low computation overhead, and energy consumption [22, 23]. Because of efficiency, we decided to study hashgraph-based approaches [24] as they provide fast consensus mechanism [25] operating even 120000 transactions per second (TPS) [26] with minimal communication overhead [24], and low energy consumption 3 mWh/transaction [27] due to minimal radio operations [28]. The hashgraph consensus mechanism is based on virtual voting, where only data of data (i.e., hashes) is shared between the nodes (i.e. gossip of gossiping). The data is shared using a flooding mechanism, where every node randomly sends its own information to neighbor nodes. Thus, during the consensus, there is a need to distribute two hashes (The size of the hashes is between 160 to 384 bits, depending on the security level) and timestamps (64 bits). This is a small amount compared to the actual data and the flooding mechanism provides fast consensus. However, if only hashes are distributed and the original data is only available in the source node, there must be other means to replicate data, e.g., using mirror nodes. Otherwise, the original data is lost if the source node is lost.

We propose to use node identities and database addresses, such as URL or IP addresses, as transaction data on the hashgraph, as shown in Fig. 6. By including the hash of the data in the database, we can verify its integrity. The data stored in the database can be specific to each node and sensor, and can be provided by different actors in the hashgraph community, such as fresh water companies, sensor manufacturers, sensor sellers, software companies, etc. For example, sensor manufacturers can provide calibration and SRAM-PUF data. The data can be either public (e.g., calibration data) or private (e.g., SRAM-PUF data), and can be accessed with appropriate methods.

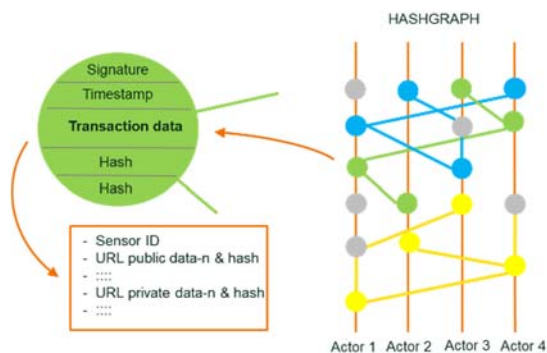


Fig. 6. Hashgraph and transaction data.

4.4. System Integrity Verification

When the system is up and running, the integrity of the field devices must be verified. Trustworthy way requires that the field devices have hardware-based root of trust that is protecting a device secret. In our case, the device secret is based on SRAM-PUF. The root-of-trust is based on ARMv8-M Trusted Firmware-M (TF-M) security architecture that is an implementation of ARM Platform Security Architecture (PSA) [29]. The security architecture provides isolation between secure and non-secure environments and allows secure boot implementation. Based on these building blocks, remote parties can request integrity verification of the system using a remote attestation protocol.

The Internet Engineering Task Force (IETF) RATS working group has specified an architecture for remote attestation that is specifying actors and data flows between the actors [30]. The architecture data flow is presented in Fig. 7.

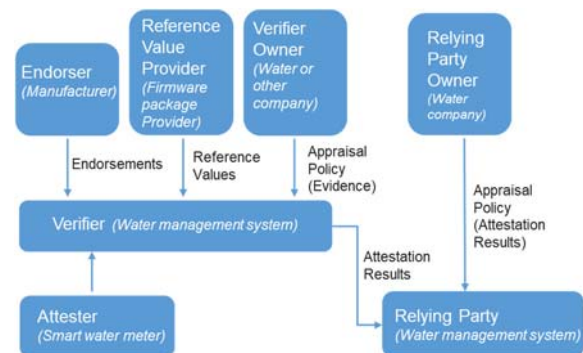


Fig. 7. IETF Remote Attestation ProcedureS (RATS) Architecture conceptual data flow.

In the context of water management, the attester actor can be a smart water meter that needs to provide a proof of its integrity to the water management system that acts as the relying party actor of the architecture. The endorser actor can be a manufacturer of the smart water meter that could, e.g., issue a device certificate for the smart water meter. The attester should provide signed integrity measurements of the system components, e.g., cryptographic hashes. These cryptographic hashes can be compared to known good reference values provided by the reference value provider. For example, a firmware package provider could provide these reference values. The verifier actor is verifying the attestation evidence and then enforcing the policy (also known as appraisal) set by the verifier owner, e.g., omitting verification of some components, and then passing the attestation result to the relying party (now the water management system). The water management system can then act according to the appraisal policy, e.g., logging attestation errors or blocking access from components whose verification has failed.

We propose that distributed ledger technology can be exploited for remote attestation systems in context

of water management (see Fig. 8). The figure presents W3C (World Wide Web Consortium) model of digital verification architecture [31], which is fit to remote attestation verification model, where the water meter acts as Holder, verifier owner and manufacturer of meter as Issuers, water management system as Verifier, reference value provider located in Verifiable Data Registry implemented with distributed ledger technology. The functionality of reference value providers can be automated by using smart contracts, where verification is made.

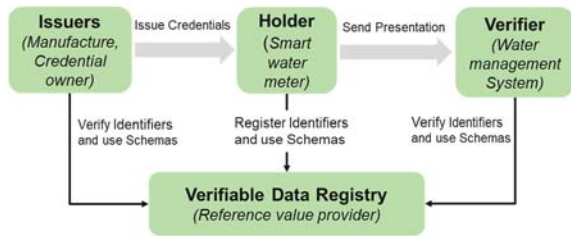


Fig. 8. Remote attestation model (RATAS) fit to W3C's Verifiable Credentials Data Model.

5.3. Secure Software Updates

Software updates are either functionality or security updates. There is a need to either deploy new functionality or fix problems in old functionality. Security updates are fixing detected vulnerabilities. Update may be either field update or over-the-air (OTA) update. Especially OTA updates should be optimized by size. Update packages must be signed, and packages should be verified before installation. Packages should have a version number and rollback prevention should prevent installation of old vulnerable versions to already updated devices. Trusted Firmware-M includes Firmware Update (FWU) component that allows testing the installed update and either rolling back or accepting the update. Software update requires additional storage as the system must be able to store the original and the updated system image.

5. Discussion

5.1. Security and Zero Trust Concept

In Future water management system with an untrusted cloud environment, the attack surface will be much larger than in conventional water management systems having only some remote connections. The development and digitalization will lead to cloud-based systems, where clear barrier areas with firewalls cannot be defined. In that kind of environment, a zero trust approach will enrich methods to create sufficient protection and improve management of the system, especially under attacks

and even attacks coming from inside of the system. One key point to build up security based on zero trust is that every actor, like users, devices, services, and programs, need to authenticate before getting rights to operate and use the resources. To do authentication actors will need a robust root of trust. Especially, field devices with a few energy resources will need energy efficient and reliable means to have root of trust like SRAM-PUF fingerprint.

SRAM-PUFs are a hardware security primitive that generate unique responses based on the physical variations in SRAM cells. These responses are not stored as conventional software-based cryptographic keys, making them resistant to software attacks that exploit vulnerabilities or weaknesses in software programs. However, SRAM-PUFs are not immune to side channel attacks, which are attacks that exploit the physical characteristics or behavior of the hardware. For example, cloning attacks [32, 33] can copy the SRAM-PUF responses by measuring the power consumption or electromagnetic radiation of the device. Data remanence attacks [34] can recover the SRAM-PUF responses by reading the residual charge left in the SRAM cells after power-off.

Data remanence is the phenomenon of data persisting in some physical form even after it has been erased. This can depend on the temperature, as data can remain for about 1.5 hours at 75 °C, 3 days at 50 °C, almost two months at 20 °C, and about 3 years at 0 °C [35, 36]. A data remanence attack is a type of fault injection attack that tries to recover data by introducing faults into the target system. They have shown that the data remanence effect can also be used to inject faults by changing the temperature of the SRAM [37].

The goal of cloning attacks is to clone the original SRAM and replace it with a fake SRAM that acts like the real one even if the attacker controls it. One way of cloning attack is to exploit a complementary metal-oxide-semiconductor aging mechanism, called Bias Temperature Instability (BTI), to alter the original start-up values of an SRAM-PUF and create physical clone [32]. Another method is to use a focused ion beam circuit to measure SRAM-PUFs and make a physical clone of SRAM. To measure the SRAM, they captured the near infrared (NIR) photonic emissions of SRAM and used them to make a clone.

Both types of side-channel attacks require deep knowledge of the specific hardware and often require laboratory equipment and settings, making them difficult to execute in the field. In our implementation, if the device is turned off and on again, the new fingerprint is also generated and the old one is discarded, which requires a new authentication. This will effectively prevent side-channel attacks that reboot the system, such as cloning attacks, and make it hard to inject faults that cause system rebooting and require new authentication. However, if the attacker can easily power off and on the field devices and the reauthentication requires administrative support, this will open up a new vulnerability in the system.

5.2. SRAM Energy Efficiency

SRAM stands for static because it does not need constant power bursts like DRAM (dynamic random-access memory) to keep the data stored. It can retain data as long as a small steady current is supplied, which reduces static energy dissipation to less than one to tens of nanowatts (nW) [37, 38]. The main source of static energy dissipation is leakage current, which causes continuous power consumption in the idle state. However, SRAM consumes more energy in the active state, when it is accessed for reading or writing, ranging hundreds of microwatts (μ W) [39]. The amount of energy consumed during a read or write operation depends on factors such as the size of the memory cell, the operating voltage and temperature, and the access time. If SRAM-based caches are used for fast data access, they need to be kept in standby mode, where they may be partially active, consuming a small amount of power to maintain data integrity and respond quickly to requests.

5.3. Hashgraph

To avoid a single point of failure, the system should be decentralized. One way to achieve secure distribution is to use distributed ledger technology, which includes blockchain and hashgraph technologies. From a security perspective, the distributed ledger technology provides a verified data registry, which has data integrity assurance and support for different consensus mechanisms as inherent properties. This ensures data consistency even if some actors are malicious. Several public and commercial ledgers (e.g., Bitcoin, Ethereum, Hedera) have cryptocurrencies, which are used to reward participants of the ledger block creation. However, from a private distributed ledger perspective, cryptocurrency is an extra functionality, which is not necessarily required. Nevertheless, if the platform service is bought from an external provider or if there is a need to create some incentive or exchange system among community members or online shops, a cryptocurrency can facilitate those functionalities.

6. Conclusions

The research paper proposes a security architecture for future water management systems based on zero trust and physical unclonable function (PUF) concepts. Water management is a critical infrastructure that needs to operate reliably and securely in various situations. Future water management systems will use cloud and edge services, remote connections, and heterogeneous devices, which will increase the complexity and vulnerability of the systems. PUF is a technology that generates unique and unclonable identifiers or keys based on the physical properties of a device or component. PUF can

provide a local root of trust for resource-constrained devices, such as sensors, in the field. As the resource efficient and suitable sensor devices was used SRAM-PUF. Zero trust is a security paradigm that assumes no inherent trust between actors in a system, such as users, devices, services, and applications. Zero trust requires continuous authentication and authorization of resources, secure communication, and dynamic policies. The paper describes the elements of zero trust architecture and how they can be applied to water management systems. The paper discusses how to verify the integrity of field devices using remote attestation protocols and how to perform secure software updates using trusted firmware components. The paper also considers the challenges and limitations of these approaches.

Acknowledgements

We would like to thank Jukka Julku for his main contributions to the remote attestation implementation and Aada Illikainen for her work with the Hashgraph.

References

- [1]. S. Renukappa, A. Kamunda, S. Suresh, Impact of COVID-19 on water sector projects and practices, *Utilities Policy*, Vol. 70, 2021, 101194.
- [2]. N. Ntuli, A. Abu-Mahfouz, A simple security architecture for smart water management system, *Procedia Computer Science*, Vol. 83, 2016, pp. 1164-1169.
- [3]. W. Xia, X. Chen, C. Song, A framework of blockchain technology in intelligent water management, *Frontiers in Environmental Science*, Vol. 10, 2022, pp. 1-12.
- [4]. S. W. Lee, S. Sarp, D. J. Jeon, J. H. Kim, Smart water grid: the future water management platform, *Desalination and Water Treatment*, Vol. 55, 2015, pp. 339-346.
- [5]. C. V. Martinez, Turning the tide: The digital future of water management, <https://www.autodesk.com/design-make/articles/water-management>
- [6]. E. Bindler, P. Keaney, K. Martz, Securing Water Utilities with AWS, Amazon Cloud, <https://aws.amazon.com/blogs/industries/new-cybersecurity-white-paper-securing-water-utilities-using-aws/>
- [7]. T. McGrath, I. E. Bagci, Z. M. Wang, U. Roedig, R. J. Young, A PUF taxonomy, *Applied Physics Reviews*, Vol. 6, Issue , February 2019, 011303.
- [8]. D. E. Holcomb, W. P. Burleson, K. Fu, Power-up SRAM state as an identifying fingerprint and source of true random numbers, *IEEE Transactions on Computers*, Vol. 58, 2008, pp. 1198-1210.
- [9]. S. Rose, O. Borchert, S. Mitchell, S. Connelly, Zero Trust Architecture, Special Publication (NIST SP), *National Institute of Standards and Technology*, Gaithersburg, MD, 2020.
- [10]. ARM, TrustZone for Cortex-M, ARM, <https://www.arm.com/technologies/trustzone-for-cortex-m>
- [11]. DICE Layering Architecture, *Trusted Computing Group*, Jul. 2020.

- [12]. A. Francillon, Q. Nguyen, K. B. Rasmussen, G. Tsudik, A minimalist approach to Remote Attestation, in *Proceedings of the Design, Automation & Test Europe Conference & Exhibition (DATE'14)*, 2014, pp. 1-6.
- [13]. A. Shamsoshoara, A. Korenda, F. Afghah, S. Zeadally, A survey on physical unclonable function (PUF)-based security solutions for Internet of Things, *Computer Networks*, Vol. 183, 2020, 107593.
- [14]. D. E. Holcomb, W. P. Burleson, K. Fu, Initial SRAM state as a fingerprint and source of true random numbers for RFID tags, in *Proceedings of the Conference on RFID Security*, Vol. 7, 2007, p. 01.
- [15]. Asvin, Physically Unclonable Function (PUF)-Setup, <https://asvin.io/physically-unclonable-function-setup/>
- [16]. NIST Selects 'Lightweight Cryptography' Algorithms to Protect Small Devices, *National Institute of Standards and Technology*, 2023.
- [17]. I. Elsayed, S. Aftabjahani, D. Gardner, E. MacLean, J. R. Wallrabenstein, E. Y. Tawfik, Hardware and energy efficiency evaluation of NIST lightweight cryptography standardization finalists, in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'22)*, 2022, pp. 133-137.
- [18]. M. S. Turan, K. McKay, D. Chang, J. Kang, N. Waller, J. M. Kelsey, L. E. Bassham, D. Hong, Status Report on the Final Round of the NIST Lightweight Cryptography Standardization Process, NIST Interagency/Internal Report (NISTIR), *National Institute of Standards and Technology*, Gaithersburg, MD, 2023.
- [19]. B. Varghese, M. Villari, O. Rana, P. James, T. Shah, M. Fazio, R. Ranjan, Realizing edge marketplaces: challenges and opportunities, *IEEE Cloud Computing*, Vol. 5, 2018, pp. 9-20.
- [20]. J. Sedlmeir, H. U. Buhl, G. Fridgen, R. Keller, The energy consumption of blockchain technology: Beyond myth, *Business & Information Systems Engineering*, Vol. 62, 2020, pp. 599-608.
- [21]. P.-L. Aublin, S. B. Mokhtar, V. Quéma, RBFT: Redundant byzantine fault tolerance, in *Proceedings of the IEEE 33rd International Conference on Distributed Computing Systems*, 2013, pp. 297-306.
- [22]. S. M. H. Bamakan, A. Motavali, A. B. Bondarti, A survey of blockchain consensus algorithms performance evaluation criteria, *Expert Systems with Applications*, Vol. 154, 2020, 113385.
- [23]. J. Zhang, Y. Rong, J. Cao, C. Rong, J. Bian, W. Wu, DBFT: A Byzantine fault tolerance protocol with graceful performance degradation, *IEEE Transactions on Dependable and Secure Computing*, Vol. 19, 2021, pp. 3387-3400.
- [24]. L. Baird, The SWIRLDS hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance, Tech Reports SWIRLDS-TR-2016-01, *SWIRLDS*, Vol. 34, 2016, pp. 9-11.
- [25]. L. Baird, A. Luykx, The hashgraph protocol: efficient asynchronous BFT for high-throughput distributed ledgers, in *Proceedings of the International Conference on Omni-layer Intelligent Systems (COINS'20)*, 2020, pp. 1-7.
- [26]. N. Gao, R. Huo, S. Wang, T. Huang, Y. Liu, Sharding-hashgraph: a high-performance blockchain-based framework for industrial internet of things with hashgraph mechanism, *IEEE Internet of Things Journal*, Vol. 9, 2022, pp. 17070-17079.
- [27]. J. I. Ibañez, F. Rua, The energy consumption of Proof-of-Stake systems: A replication and expansion, *arXiv Preprint*, 2023, arXiv:2302.00627.
- [28]. P. Koskela, Energy-efficient solutions for wireless sensor networks, PhD Thesis, *Acta Universitatis Ouluensis*, 2018.
- [29]. T. Martin, Designing Secure IoT Devices with the Arm Platform Security Architecture and Cortex-M33, *Elsevier Science*, Oxford, 2022.
- [30]. H. Birkholz, D. Thaler, M. Richardson, N. Smith, W. Pan, Remote ATtestation procedureS (RATS) Architecture, IETF RFC 9334, <https://www.rfc-editor.org/info/rfc9334>
- [31]. M. Sporny, D. Longley, D. Chadwick, Verifiable Credentials Data Model v1.1, W3C Recommendation, 03 March 2022, <https://www.w3.org/TR/2022/REC-vc-data-model-20220303>
- [32]. S. Duan, G. Sai, BTI aging-based physical cloning attack on SRAM PUF and the countermeasure, *Analog Integrated Circuits and Signal Processing*, Vol. 117, Issue 1-3, Dec 2023, pp. 45-55.
- [33]. C. Helfmeier, C. Boit, D. Nedospasov, J.-P. Seifert, Cloning physically unclonable functions, in *Proceedings of the IEEE International Symposium on Hardware-Oriented Security and Trust (HOST'13)*, 2013, pp. 1-6.
- [34]. C. Yehoshuva, R. Raja Adhithan, N. Nalla Anandakumar, A survey of security attacks on silicon based weak PUF architectures, in *Proceedings of the 8th International Symposium on Security in Computing and Communications (SSCC'20)*, Chennai, India, 2020, pp. 107-122.
- [35]. P. Gutmann, Data remanence in semiconductor devices, in *Proceedings of the 10th USENIX Security Symposium (USENIX Security'01)*, 2001, pp. 39-54.
- [36]. N. A. Anagnostopoulos, T. Arul, M. Rosenstihl, A. Schaller, S. Gabmeyer, S. Katzenbeisser, Attacking SRAM PUFs using very-low-temperature data remanence, *Microprocessors and Microsystems*, Vol. 71, 2019, 102864.
- [37]. G. Prasad, A. Anand, Statistical analysis of low-power SRAM cell structure, *Analog Integrated Circuits and Signal Processing*, Vol. 82, 2015, pp. 349-358.
- [38]. D. Anitha, M. M. Ahmad, Ultra-low leakage static random access memory design, *International Journal of Reconfigurable and Embedded Systems*, Vol. 12, 2023, pp. 60-69.
- [39]. P. Sandeep, P. A. H. Vardhini, V. Prakasam, SRAM utilization and power consumption analysis for low power applications, in *Proceedings of the International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT'20)*, 2020, pp. 227-231.



A MEMS Based Large Travel Range Linear Actuator with a Good Minimum Step Size to Be Used in Mirror Shape Correction Applications

Ehsan Atash ZABAN and Dr Mehmet KUSAF

Department of Electrical and Electronic Engineering,
Cyprus International University, 99258, Nicosia, Northern Cyprus, Mersin 10, Turkey
Tel: 00989375437106
E-mails: 22009479@student.ciu.edu.tr, mkusaf@ciu.edu.tr

Received: 16 September 2023 / Accepted: 28 December 2024 / Published: 28 February 2024

Abstract: In this paper, a high-precision thermal linear micro actuator (*LMA*) with a large travel range is analyzed, simulated and presented. Physical functions can be performed by MEMS thanks to micro-actuators. The operation of various types of micro-actuators in many applications has been successful. There are two perspectives to categorize micro-actuators; one focused on driving forces and the other on mechanisms. Electric, magnetic, and flow fields can be utilized to generate force in the space between stationary and moving parts. The intrinsic actuation capabilities of materials like piezoelectric, magnetostrictive, and photo stretched materials also can be used in them. Changes in shape or volume caused by thermal expansion and phase transformations; such as the shape-memory effect and bubble formation are another important ways to reach a deformation in these types of actuators. In our proposed *LMA*, electro-thermal technique has been selected. The main components of the electro-thermal *LMA* proposed are substrate, slider, stopper, and forks. The slider is a long rectangle beam placed on a rail so that can move bidirectional, either up-ward or down-ward. A complex interesting technique, by using the stoppers and forks, results bidirectional movement of the slider. The stopper consists of arrayed electro-thermal chevrons (*EC*) and springy clutch. By applying a DC voltage of 1.5 V to the Red fork's *EC*s arrays, the minimum step size approximately 20 μm is achieved. Repeating steps sequentially generate a travel range of about 2 mm. For all simulations, the ANSYS software has been used. The *LMA* has the ability to displace heavy loads linearly and accurately. Minimum step size and large travel range of the *LMA* proposed make it useful to be used for Mirror Shape Correction application.

Keywords: MEMS, Mirror shape correction, Linear micro actuator (*LMA*), Electro-thermal chevrons (*EC*), The slider, The stopper, The forks, The springy clutch.

1. Introduction

Numerous researchers have studied *LMAs* and micro-deflections extensively for the purpose of arranging individual elements connected to the slider's head in micro electromechanical systems (MEMS). By repeating small deflections [1, 2], it is

possible to move a slider linearly on a rail using certain techniques. Electrostatic, magneto-electric, piezoelectric, and electro-thermal techniques are the most effective ways to generate these deflections. It's unfortunate that micro scale applications tend to generate electrostatic and magneto-static forces that are too small. The devices can be operated in a resonant mode or applied a high voltage to achieve a

large travel range and overcome this issue. The Piezoelectric and Electro-thermal *LMAs* have the ability to create significant propulsion forces and deformation. Commercial MEMS foundries typically don't support piezoelectric materials during their fabrication processes. Electro-thermal *LMAs* [4] were the reason why we were drawn to them. Low power consumption and a large and linear travel range can be achieved using electro-thermal actuators as a promising solution. The IC voltage is the standard for standardizing the operating voltage of these actuators, unlike electrostatic-driven actuators. It is now easier to design them with no strict dimensional tolerances. Sliders in electro-thermal linear actuators allow for precise movement forward or backwards. Microsystem devices can generate linear movement through a variety of techniques. To move sliders, the initial method involves transferring momentum through a series of impacts beginning with number [1, 5, 16, 17]. The second method involves employing impact heads and notches the edges of the sliders, which significantly improves grip for sliders with a thickness of 6 to 8 mm. In the third method, the gripping occurs due to the frictional contact between the impact heads and the slider [9, 10] and leads to a beneficial resolution. The focus of this paper is on a sophisticated electro-thermal *LMA* that can travel up to 2 mm, and has a minimum step size of 20 mm and a travel range of 2 mm. An *LMA* for mirror shape correction application has been reported by [11, 19], but it does not have favorable power consumption or travel range. [12] has been reported a bulky *LMA*. Compared to [12], the proposed actuator is lighter and can be utilized in various important applications. In the following sections, several issues related to the *LMA* are discussed. The actuator mechanism and stopper are examined in Sections 2

and 3 with care, accuracy, and theoretical reasoning. The design of the slider is presented in Section 4. Section 5 is the conclusion.

2. Mechanism

Blue and green forks, along with the slider, stopper, and substrate, are all main components in this *LMA*. The schematic of the *LMA* can be seen in Fig. 1. Several EC arrays are used to move the slider in two directions and also release the springy clutches from the side wall. Two supports at an angle, known as 'anchors', secure the long beams that make up the ECs. Current passing through the beams from one terminal to another is what causes them to expand due to Joule heating. A deflection occurs in the y-direction of mathematical axis due to thermal expansion; consider Red ECs as an example. Deflecting the EC is affected significantly by both the length and angle of the beams. A smaller angle causes a greater deflection, while a larger angle causes a smaller deflection [7, 13, 20, 25]. The substrate acts as a chassis, and the slider is a solid long rectangular beam that can be moved on the substrate. Totally, two forks are available, one blue and one green. As depicted in Fig. 1, the stoppers are located on the four locations of the fork surface. The stopper's composition consists of the springy clutch and the ECs. To prevent the slider from shaking and moving randomly, the clutches are used. Besides, the stopper's ECs aid in generating the required force to connect the clutches to the substrate side walls. All clutches must be touched to the side walls of the slider during normal mode to prevent it from moving or shaking.

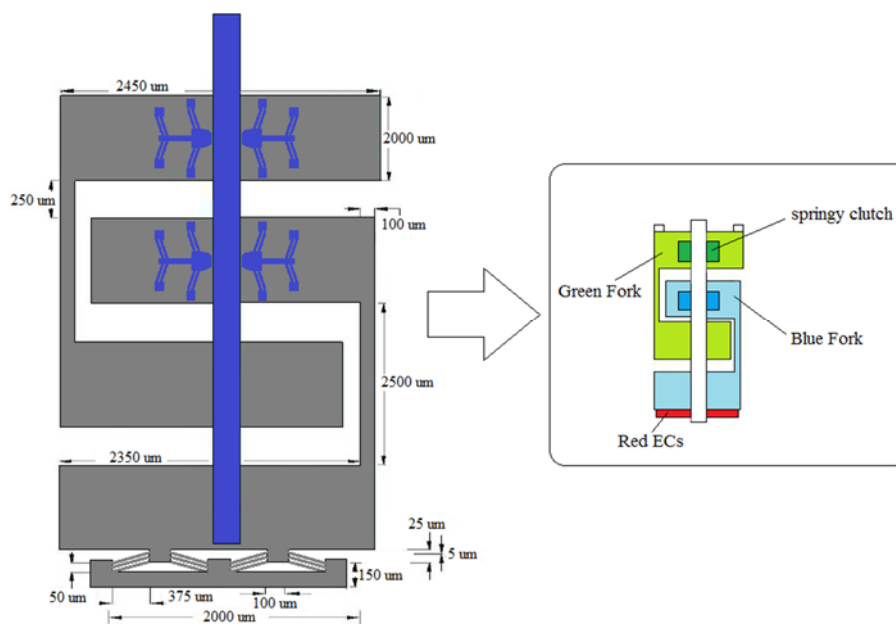


Fig. 1. The schematic of the *LMA* proposed.

Fig. 2 (a) shows that the slider is currently in rest mode, indicating that the entire fork's Green and Blue ECs are ON or activated. ON is the term for passing electrical current through electro-thermal cells and expanding polysilicon cells utilizing Joule's law. To start the operation, the Green fork's ECs are ON and the Blue fork's ECs are OFF, as shown in Fig. 2 (b). Upward or downward movement is achieved by applying electrical signals to all EC electrodes in a regular and sequential manner. The Blue fork moves forward and the slider dribbles in an interesting way when the voltages of the Green-Red and Blue forks are turned on and off, respectively. It is important to note that all the clutches are the same as in Fig. 2(c) in the previous step. Fig. 2(d) shows that the voltages applied to the ECs of the Green and Blue forks will be ON. As the temperature rises and cools the ECs, the Red fork's EC voltage drops to zero, causing the slider to contract smoothly. Refer to Fig. 2(e). This contraction leads to the Blue fork deflecting downward and the slider moving down-trend, shown in Fig. 2(e). A significant travel range could be achieved by repeating the sequences. To achieve upward movement, all steps can be reversed. In Fig. 3, you can see the schematic for the springy clutch. To obtain high precision and a large travel range, all parameters must be calculated carefully and separately [19, 9, 23, 24]. In short, generating a prototype of a complicated linear actuator necessitates the precise calculation, analysis of all dimensions, geometries, and material properties. Take into account which the generated friction force, appropriately, is enough to ensure the slider latching and stabilizing is successful during operation. To keep the sliders steady, the clutches can be fitted into the notches of LMAs. In these structures, despite the favorable force generated, resolution is limited. The expected resolution for these types of LMAs is approximately 100 nm. The LMA suggested reposition the slider with a force that could reach 8 mN either forward or backward.

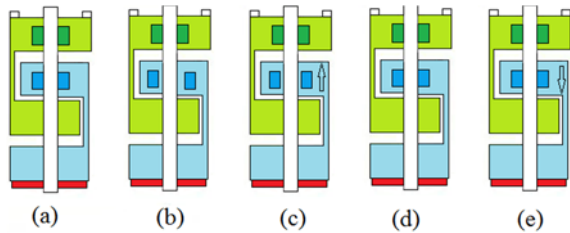


Fig. 2. Operation sequences (a) normal mode, (b) the Green fork's ECs are ON and the Blue fork's ECs are OFF, (c) The Green-Red and Blue fork's ECs voltages applied are ON; OFF, (d) the slider is getting ready to contract, (e) downward deflection of Red fork's ECs.

3. Analysis of the Stopper

By analyzing an inclined beam at the center of the chevrons shuttle, the springy clutch can be designed.

Taking into account node A and assume that the shuttle is solid, as depicted in Fig. 4. When designing a springy clutch, it's crucial to take important parameters into account. Node A is deflected by the force generated F electro thermally in the central shuttle in the y-direction, which is equivalent to [14].

$$\begin{aligned}
 U^F &\equiv U_{y \text{ total}}^A = U_{y1}^A + U_{y2}^A = \\
 &= F \frac{1}{2 \left(\sin \theta_s \frac{EA_s^2}{l_s^3} + \cos \theta_s \frac{12EI_s^2}{l_s^3} \right)} + \\
 &+ F \frac{1}{2 \left(\sin \theta_c \frac{EA_c^2}{l_c^3} + \cos \theta_c \frac{12EI_c^2}{l_c^3} \right)} = \\
 &= \frac{Fl_s}{EA_s} \frac{1}{2 \left(\sin \theta_s^2 + \frac{\cos \theta_s^2}{\gamma_s} \right)} + \\
 &+ \frac{Fl_c}{EA_c} \frac{1}{2 \left(\sin \theta_c^2 + \frac{\cos \theta_c^2}{\gamma_c} \right)}
 \end{aligned} \quad (1)$$

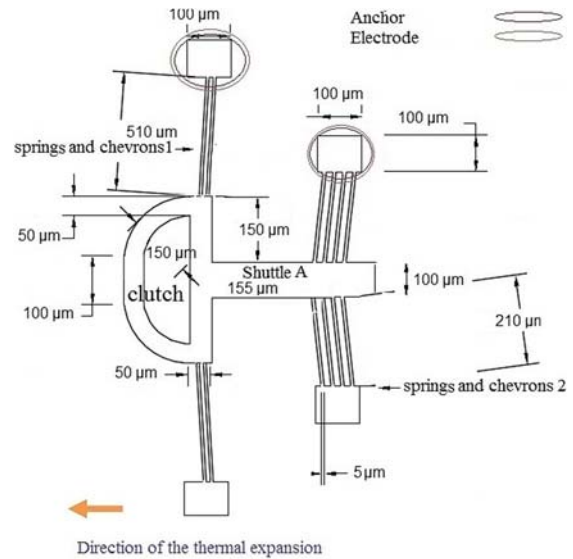


Fig. 3. The schematic of the springy clutch.

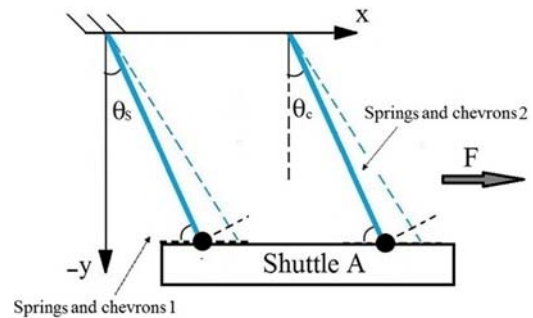


Fig. 4. A beam inclined joined at the central of chevrons' shuttle.

The expression identified by 's' and 'c' is linked to chevrons 1 and 2, as indicated in Fig. 4. E is the Young's modulus of the material, while l_s and l_c are

the beam length. A_c represents the beam cross section and I_s and I_c represents the moment of inertia of the cross section with respect to the out of plane axis z . As it said already, F is the electro-thermal force generated. Fig. 5 demonstrates the deflection of the springy clutch compared to the force applied. The reaction force in the x -direction is:

$$\begin{aligned} R_x^F &\equiv R_x^A = \\ &= \cos \theta_s \sin \theta_s \left(\frac{EA_s}{l_s} - \frac{12EI_s}{l_s^3} \right) U_{y1}^A + \\ &+ \cos \theta_c \sin \theta_c \left(\frac{EA_c}{l_c} - \frac{12EI_c}{l_c^3} \right) U_{y2}^A = \\ &= F \frac{\sin \theta_s \cos \theta_s (\gamma_s - 1)}{2(\gamma_s \sin \theta_s^2 + \cos \theta_s^2)} + \\ &+ F \frac{\sin \theta_c \cos \theta_c (\gamma_c - 1)}{2(\gamma_c \sin \theta_c^2 + \cos \theta_c^2)}, \end{aligned} \quad (2)$$

and

$$\gamma_s = \frac{A_s l_s}{12 I_s}, \gamma_c = \frac{A_c l_c}{12 I_c} \quad (3)$$

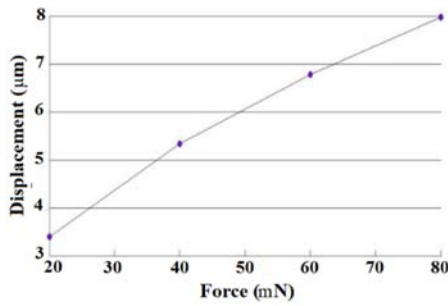


Fig. 5. Deflection of the springy clutch vs. the force applied.

The *SSFF* (static surface friction force) generated by the clutch's head is the same as:

$$F_{friction} = \mu_{static} F_u, \quad (4)$$

where F_u is the elastic potential force of the beam. Equations (1-3) can be taken into account in a primitive way to calculate the *SSFF*. The force generated by the chevrons1 and the chevrons2 causes the springy clutch to move toward the slider side wall. The electro-thermal mechanism and output heat conduction are responsible to actuate the chevrons. Following equations can be used theoretically to explain related operation.

$$Q_{output} = -K w_s h_s \left(\frac{\partial T_s}{\partial x} \right)_{x+dx} \quad (5)$$

where K , w_s , h_s , T and x represents the thermal conductivity of poly-silicon, thickness, the height of the beams, temperature, and the fluctuating length of the beams, respectively. Equation (1) is substituted by equation (2), which means heat passing through thin beams. By neglecting the heat loss of the convection and radiation to the ambient, heat conduction is as following [15]:

$$Q_{input} = -K w_s h_s \left(\frac{\partial T}{\partial x} \right)_x \quad (6)$$

By passing electrical current through the beams of the springy clutch:

$$Q_{j_s} = j_s^2 \rho w_s h_s dx, \quad (7)$$

$$\rho = \rho_0 [1 + \zeta (T_s - T_0)], \quad (8)$$

$$J_s = \frac{V}{2 \rho l_s}, \quad (9)$$

$$V = 2 l_s E_F, \quad (10)$$

where J_s , ρ , T_0 , ζ and V represents current density, resistivity, substrate temperature, the temperature's coefficient, *DC* voltage, respectively. T_0 should be considered 293K. Figs. 6 and 7 display the springy clutch's deflection in relation to the *DC* voltage applied and the release time, respectively. Fig. 8(a-d) shows Model meshed, the clutch deformation, electrical voltage and current contours, respectively.

$$Q_{input} + Q_{j_s} = Q_{output} \quad (11)$$

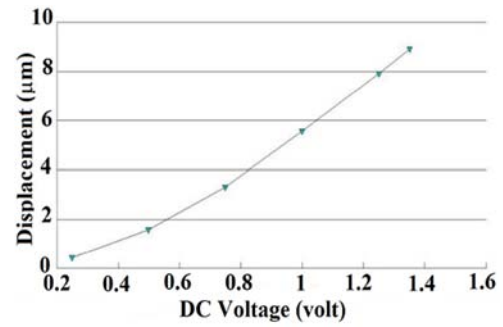


Fig. 6. Deflection of the clutch vs. the *DC* voltage applied.

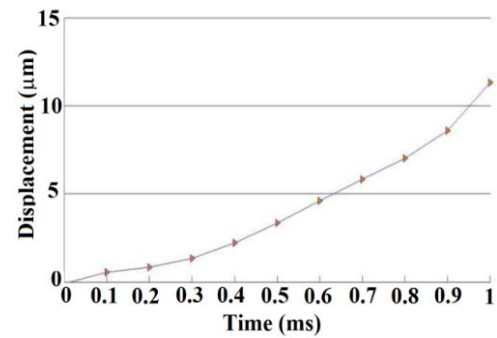


Fig. 7. The deflection of the clutch vs. the time.

The following equation can be formulated by substituting Equations (5) and (7) into Equation (11) and by limiting dx to 0.

$$K_p w_s h_s \left(\frac{\partial^2 T_s}{\partial x^2} \right) + j_s^2 \rho w_s h_s = 0 \quad (12)$$

The conductive cross-sectional area of the springy clutch is identical, so:

$$\frac{\partial T_s^2}{\partial x^2} = \frac{-j_s^2 \rho}{K}, \quad (13)$$

$$T(0) = T(l_s) = T_0, \quad (14)$$

$$T_s(x) = \frac{\rho j_s^2}{2K} (2l_s x - x^2) + T_0, \quad (15)$$

$$\Delta T_s = \int_0^{2l_s} \frac{T(x)_s - T_0}{2l_s} dx = \frac{V^2}{3K\rho} = \frac{P_r l_s}{3K A_s} \quad (16)$$

$$\Delta T_{s_{max}} = T_{s_{max}} - T_0 = \frac{V^2}{2K\rho} = 1.5 \Delta T_s, \quad (17)$$

$$P_r = I_r V \quad (18)$$

where P_r , I_r are electrical power and electrical current, respectively. Fig. 8(e) shows temperature's distribution of the springy clutch. The average temperature of the springy clutch increases to ΔT_s , and ΔT_c , so deflection at node A in the y-direction is equal to [14]. By considering ΔT_s and ΔT_c , there is a compressive force $F_{\Delta T}$ in the hot arms [21]:

$$\begin{aligned} U^{\Delta T} &\equiv U_{y_{total}}^A = \\ &= \alpha \Delta T_s l_s \frac{\sin \theta_s}{\left(\sin \theta_s^2 + \cos \theta_s \frac{12l_s}{A_s l_s^2} \right)} + \\ &+ \alpha \Delta T_c l_c \frac{\sin \theta_c}{\left(\sin \theta_c^2 + \cos \theta_c \frac{12l_c}{A_c l_c^2} \right)} = \\ &= \alpha \Delta T_s l_s \frac{\sin \theta_s}{\left(\sin \theta_s^2 + \frac{\cos \theta_s^2}{\gamma_s} \right)} + \\ &+ \alpha \Delta T_c l_c \frac{\sin \theta_c}{\left(\sin \theta_c^2 + \frac{\cos \theta_c^2}{\gamma_c} \right)}, \end{aligned} \quad (19)$$

where α is the coefficient of the thermal expansion. The contour plots of clutch deflection are shown in Fig. 8(a) when applying a DC voltage of 1.35 volt. The meshed model of the springy clutch by using couple field elements in *ANSYS* software is shown in 8(b). The reaction force generated in the x-direction is equal to:

$$\begin{aligned} R_x^{\Delta T} &\equiv R_{x_{total}}^A = \\ &= -\alpha \Delta T_s E A_s \frac{\cos \theta_s}{\left(\sin \theta_s^2 + \frac{12l_s^2}{A_s^2} + \cos \theta_s^2 \right)} - \\ &- \alpha \Delta T_c E A_c \frac{\cos \theta_c}{\left(\sin \theta_c^2 + \frac{12l_c^2}{A_c^2} + \cos \theta_c^2 \right)} = \\ &= -\alpha \Delta T_s E A_s \frac{\cos \theta_s}{\left(\sin \theta_s^2 + \frac{\cos \theta_s^2}{\gamma_s} \right)} - \\ &- \alpha \Delta T_c E A_c \frac{\cos \theta_c}{\left(\sin \theta_c^2 + \frac{\cos \theta_c^2}{\gamma_c} \right)} \end{aligned} \quad (20)$$

When considering the clutch release time, the total deflection is:

$$U^{\Delta T+F} = (U^{\Delta T} + U_s^F) + (U_c^{\Delta T} + U_c^F) \quad (21)$$

Total stiffness is:

$$U^{\Delta T+F} = \frac{(2\alpha \Delta T_s E A_s \sin \theta_s + F) + (2\alpha \Delta T_s E A_s \sin \theta_s + F)}{K_{tb}}, \quad (22)$$

$$\begin{aligned} K_{tb} &\equiv \frac{F}{U_s^F} + \frac{F}{U_c^F} = \frac{2 \left(\sin \theta_s^2 + \frac{\cos \theta_s^2}{\gamma_s} \right) E A_s}{l_s} + \\ &+ \frac{2 \left(\sin \theta_c^2 + \frac{\cos \theta_c^2}{\gamma_c} \right) E A_c}{l_c} \end{aligned} \quad (23)$$

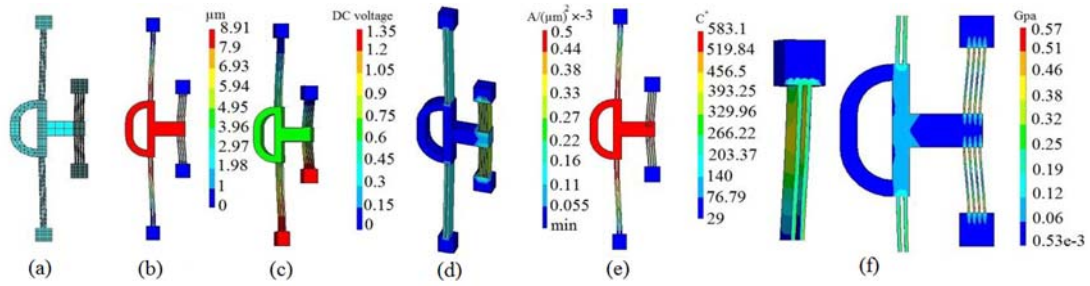


Fig. 8. Model meshed (a), the clutch deformation (b), the DC voltage (c), electrical current (d), temperature distribution (e), Stress counter of the clutch (f).

The expressions above can be approximated to the following:

$$K_{TA} = 2 \times m K_{tb} + 2 \left(\sin \theta_c^2 + \frac{\cos \theta_c^2}{\gamma_c} \right) E A_c l_c^{-1}, \quad (24)$$

$$F_{\Delta T} \cong \frac{\alpha \Delta T_s}{\sin \theta_s^2} \frac{12 E I_s}{L_{1s}^2} + \frac{\alpha \Delta T_c}{\sin \theta_c^2} \frac{12 E I_c}{L_{1c}^2} \quad (25)$$

$$L_{1s} = 0.5 \alpha \frac{L_s}{\cos \theta_s}, L_{1c} = 0.5 \alpha \frac{L_c}{\cos \theta_c}, m = 2 \quad (26)$$

By compressing the springy clutch, approximately 5 m, a force of 40 mN can be generated. The two clutches have a total SSFF of 80 mN, by assuming the SSFC is roughly 0.2 [22]. Equation (4) allows the

slider to generate propulsion upward or downward force about 16 mN which guarantee the slider stability during operation. In other words, if the slider's propulsion forces exceed the total clutch's friction force, the grip would be faulted. A force of 80 mN can be achieved by applying a DC voltage of 1.25 volt. Fig. 8(f) illustrates the stress contours of the clutch that are less than 0.57 GPa due to the applied voltage which causes a deflection of approximately $7.9 \mu\text{m}$.

4. Analysis of the Blue fork's ECs

Equations of Section 3 are used to design the ECs of the Red fork. The couple field elements of ANSYS software are used to mesh the model and the electro-thermal analysis to be carried out. Table 1 shows the material properties of poly-silicon. The slider has a stroke of $2 \mu\text{m}$ when using a voltage of 1.5 V, as shown in Fig. 9.

Table 1. The Polysilicon properties used in the simulation.

Para	Value	Unit
Dens	2.33	$\text{pg. } \mu\text{m}^{-2}$
Ex	169	$\text{pg. } \mu\text{m}^{-1}\text{ns}^{-2}$
Prxy	0.28	—
Th ex	2.6e^{-6}	$^{\circ}\text{C}^{-1}$
Res	e^9	$\mu\text{m. } \Omega$
Thermal conductivity	14.9e^{-5}	$\text{pg. } \mu\text{m. ns}^{-3}\text{C}^{-1}$

The slider's deflection in relation to the force applied and time is shown in Figs. 10, 11. The temperature of the entire machine considers 293 K. Applying the appropriate voltage can result a minimum step size of $20 \mu\text{m}$. The sequences can be repeated to achieve a large travel range, as demonstrated in Fig. 2. The contours of the deflection, stress, and temperature distribution in the Blue fork are shown in Fig. 12 (a-c), respectively. It

should be noted that the stopper can be separated from the Blue fork's body in terms of electrical and thermal issues. The neighborhood of the stopper and the Blue fork electrical circuits can lead to an error. The SiO_2 could be utilized to separate boundaries.

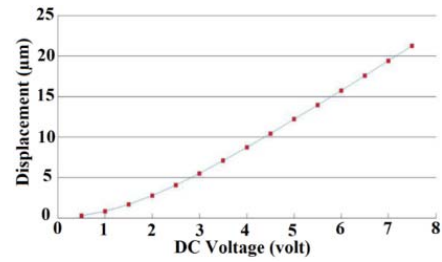


Fig. 9. Deflection of the *LMA*'s red fork vs. the DC voltage applied.

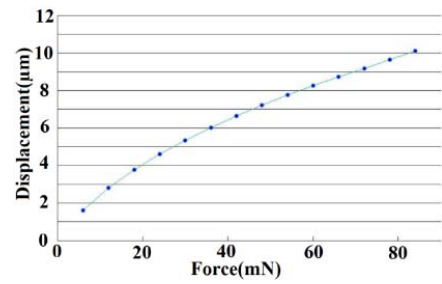


Fig. 10. The force required to deform the *LMA*'s red fork.

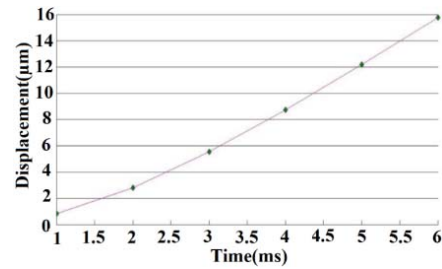


Fig. 11. Deflection of the red fork's vs. the time.

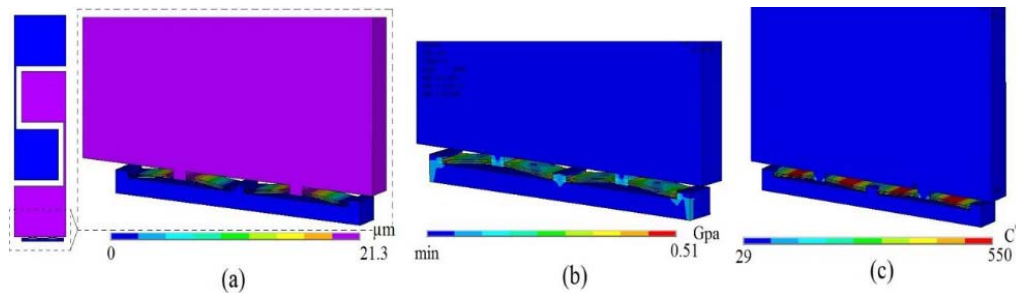


Fig. 12. The contour of the blue and red fork's deflection (a), the contours of the stress (b), the temperature distribution (c).

5. Conclusion

A favorable propulsion upward or down-ward force can be generated by the LMA, which can cause

it to displace heavy loads on the micro and millimeter scales. By repeating sequences, the LMA can reach large travel ranges and has a minimum step size of $20 \mu\text{m}$, which is useful for mirror shape correction

applications. The upward or downward propulsion force is approximately 8 mN when using a DC voltage equal to 1.5 volt.

References

- [1]. M. J. Daneman, N. C. Tien, O. Solgaard, A. P. Pisano, K. Y. Lau, R. S. Muller, Linear microvibromotor for positioning optical components, *J. Microelectromech. Syst.*, Vol. 5, 1996, pp. 159-165.
- [2]. N. R. Tas, T. Sonnenberg, R. Molenaar, M. Elwenspoek, Design, fabrication and testing of laterally driven electrostatic motors employing walking motion and mechanical leverage, *J. Micromech. Microeng.*, Vol. 13, 2003, pp. N6-N15.
- [3]. L. Ristic (Ed.), Sensor Technology and Devices, *Artech House*, Norwood, MA, 1994.
- [4]. C. S. Pan, W. Hsu, An electro-thermally and laterally driven polysilicon microactuator, *J. Micromech. Microeng.*, Vol. 7, 1997, pp. 7-13.
- [5]. M. Pai, N. C. Tien, Low voltage electrothermal vibromotor for silicon optical bench applications *Sensors Actuators A*, Vol. 83, 2000, pp. 237-243.
- [6]. J. R. Reid, V. M. Bright, J. T. Butler, Automated assembly of flip-up micromirrors, *Sensors Actuators A*, Vol. 66, 1998, pp. 292-298.
- [7]. J. S. Park, L. L. Chu, A. D. Oliver, Y. B. Gianchandani, Bent-beam electrothermal actuators: II. Linear and rotary microengines, *J. Microelectromech. Syst.*, Vol. 10, 2001, pp. 255-262.
- [8]. H. N. Kwon, S. H. Jeong, S. K. Lee, J. H. Lee, Design and characterization of a micromachined inchworm motor with thermoelastic linkage actuators, *Sensors Actuators A*, Vol. 103, 2003, pp. 143-149.
- [9]. M. Baltzer, T. Kraus, T. Obermeier, A linear stepping actuator in surface micromachining technology for low voltages and large deflections, in *Proceedings of the IEEE Conference Solid-State Sensors and Actuators*, Chicago, IL, 1997, pp. 781-784.
- [10]. Naga Manikanta Kommanaboina, Teferi Sitotaw Yallew, Alvise Bagolini & Maria F. Pantano, A C-shaped hinge for displacement magnification in MEMS rotational structures, *nature Journal*, Published: 04 January 2024.
- [11]. M. P. de Boer, D. L. Luck, W. R. Ashurst, R. Maboudian, A. D. Corwin, J. A. Walraven, J. M. Redmond, A High-performance surfacemicromachined inchworm actuator, *J. Microelectromech. Syst.*, Vol. 13, 2004, pp. 63-74.
- [12]. T. Risaku, Y. Eui-Hyeok, A normally latched large-stroke, inchworm microactuator, *Micromech. Microeng.*, Vol. 17, 2007, pp. 1715-1720.
- [13]. J. K. Luo, A. J. Flewitt, S. M. Spearing, N. A. Fleck, W. I. Milne, Modelling of microspring thermal actuator, in *Proceedings of the NSTI-Nanotech Conference*, Vol. 1, 2004, pp. 355-358.
- [14]. Z. Yong, C. Alberto, D. E. Horacio, A thermal actuator for nanoscale in situ microscopy testing: design and characterization, *J. Micromech. Microeng.*, Vol. 16, 2006, pp. 242-253.
- [15]. B. S. Ang, D. Zuraini, S. Othman, M. Muhamad Azman, Design and analysis of thermal microactuator European, *Journal of Scientific Research*, Vol. 35, Issue 2, 2009, pp. 281-292.
- [16]. A. Potekhina, C. Wang, Review of electrothermal actuators and applications school of engineering and physical sciences, *Actuators*, Vol. 8, Issue 4, 2019, 69.
- [17]. S. E. Osman, M. Zarog, Optimized V-shaped beam micro-electrothermal actuator using particle swarm optimization (PSO) technique, *Micro and Nanosystems*, Vol. 11, Issue 1, 2019, pp. 62-67.
- [18]. M. Chiao, L. Lin, Self-buckling of micromachined beams under resistive heating, *J. Microelectromech. Syst.*, Vol. 9, 2000, pp. 146-151.
- [19]. O. Ulkir, Design and fabrication of an electrothermal MEMS micro-actuator with 3D printing technology, *Materials Research Express*, Vol. 7, Issue 7, 2020, 075015.
- [20]. P. D. Pour, Dr. M. Ghommam, Dr. A. Abdelkefi, Modeling and design enhancement of electrothermal actuators for microgripping applications, *Appl. Sci.*, Vol. 13, Issue 18, 2023, 10140.
- [21]. D. Hill, W. Szyszkowski, E. Bordatchev, On modeling and computer simulation of an electrothermally driven cascaded nickel microactuator, *Sensors and Actuators A*, Vol. 126, 2006, pp. 253-263.
- [22]. I. H. Hwang, J. H. Lee, Novel measurement system of the friction coefficients for the DRIE sidewalls, in *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems*, 2006, pp. 210-213.
- [23]. S. Ramya, S. Praveen Kumar, T. Aravind, U. Rasan, G. Dinesh Ram, D. Lingaraja, Design of chevron electrothermal actuator for high force defense applications, in *Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC'23)*, 2023, pp. 1-3.
- [24]. S. Memon, S. Khan Baloch, S. Rehman Soomro, I. Ali Tunio, O. Ur Rehman, S. Ahmed, Design and finite element modeling of electro-thermal actuator for biological applications, *Journal of Computing & Biomedical Informatics*, Vol. 5, Issue 1, 2023.
- [25]. H. Xie, H. Yang, A. R. Y. Ding, L. Xiao, T. Pan, Y. Yan, W. Jiao, A robust lateral shift free (LSF) electrothermal micromirror with flexible multi-morph beams, *Microsystems & Nanoengineering*, Vol. 9, 2023, pp. 1-12.



A Novel Large Travel, High Precision MEMS Based Electro-thermal Linear Actuator for Adjusting CCD Camera Lens

Ehsan Atash ZABAN and Dr Mehmet KUSAF

Cyprus International University, Nicosia, Cyprus

Tel: 00989375437106

E-mail: 22009479@student.ciu.edu.tr, mkusaf@ciu.edu.tr

Received: 16 December 2024 /Accepted: 28 December 2024 /Published: 28 February 2024

Abstract: In this paper, a linear micro actuator (*LMA*) with a stroke of $2.1\ \mu\text{m}$ is designed and presented. Substrate, springy slider, and stopper are main parts of the *LMA* proposed. A large travel range can be achieved, by repeating the generated strokes. All simulations are done with *ANSYS* finite element. The springy slider, particularly and interestingly, can be deformed bidirectional either up-ward or down-ward. In fact, the slider creates such a linear deformation in one direction by electro-thermal stimulation. Without a stabilizer there is no value in even making such linear deformation. To stabilize and to prepare the slider for a large travel range, the four stoppers includes the springy clutches and electrostatic comb drives are used. The comb drives works with the electrostatic actuation and generates the propulsion forces required of the springy clutches. The clutches touches the side wall of the substrate to hold and stable the slider's stroke. Combining the stroke of the slider and the stoppers operation, results a large travel range interestingly. The *LMAs* with large travel range are so important for different kinds of applications in micrometer scales. This *LMA* can push and pull the loady objects connected to the slider's head accurately and can be used in adjusting CCD camera lens.

Keywords: MEMS, CCD camera lens, The *LMA*, The slider, The stopper, Stroke, The Clutch, The electro-thermal actuators.

1. Introduction

The *LMAs* play an important role in micro-electro-mechanical-systems (*MEMS*). They can be the main components of the artificial arms and the muscles so that can strengthen the dynamic parameters of the robots because of their light-weight, micrometric scale and their own output force generated. These actuators have many applications in the industrials and the scientific equipment's; for example, the MRI resonance image scanners and the electron beam lithography systems. There are different kinds of actuators working on complex and sensitive micro systems such as the electro-thermal, the electrostatic,

the piezoelectric, and the magnetic and so on [3, 4, 7]. All the micro actuators mentioned have their relative advantages and disadvantages, particularly. As an example, the magnetic *LMAs* can have a significant negative effect on the entire micro scale systems, so the non-magnetic ones can be proposed and substituted [15, 2]. The resolution, the stroke and the speed are the parameters to be taken into account during designing. The force generated by the electro-static micro-actuators is too small. Therefore, to increase the deformation, the voltage applied should be increased. The piezoelectric micro actuator are powerful and can displace heavy loads but due to the problems of using the piezoelectric materials in the

process of layering, casting in micrometric cases and achieving the displacement of more than one millimeter, the specialists of such systems are focused on the electro-thermal systems, instead. To create a linear motion actually in a certain direction, there are some methods. Firstly, the slider of micro-motor can be moved through transferring motions between a series of impacts. Second methods are the impact heads toothed engaged with the slider, which causes the slider to have a high output force [11, 10, 16, 7]. In the third method, the slider is held through the friction generated between the clutch of a stopper and either the contact surfaces or the substrate side walls [1, 14]. In this work, a novel *LMA* with a force generated up to 2 mN and a large travel range, with high precision, is going to be proposed. One of the interesting goals in the designing such micro actuators is to achieve high precision stroke. Another point in designing of these actuators is to achieve the large travel ranges so that *LMAs* can be applied in some important applications like adjusting CCD Camera Lenses [21, 12]. In the Section 2.1 of the paper, the mechanism of the linear actuator proposed is explained. Afterward, the slider is analyzed and equations due to designing are presented (Section 2.2); paper explains the stopper (Section 2.3). In the final, the conclusion is presented.

2. Design

2.1. Mechanism

The micro-actuator proposed, as it can be shown in Fig. 1, has three main parts; the substrate, the slider and the stoppers. The substrate acts as a chassis and the springy slider consists of the beams (hot arms), the electrical electrodes, and the constrainer. The stopper assemblies on the four corners of the springy slider's upper surface. In other words, there are the four stoppers placed on the surface of the slider. The stoppers include the springy clutch, and electrostatic comb drives. The clutches are used to prevent the slider from moving or randomly shaking and the

electrostatic comb drives are used to generate the force required for connecting the clutches to the substrate side walls. As it said, in the normal state, all the clutches' heads are connected to the substrate side wall to prevent the slider from shaking and vibrating. Thickness of the slider, the comb drives and the clutch are 5 μm , 3.5 μm and 3.5 μm , respectively. Two sets of the electrodes are located in up and down sections of the slider to stimulate the springy electro-thermal actuator. When the slider is in the rest mode, all of the clutches are connected to the substrate side wall. For an up-ward or down-ward deformation, an actuation voltage is required to be applied to the electro-thermal slider electrode which causes an electrical current flowing through the hot arms. To lock the four corners of the slider from any randomly shaking or instability, the four comb drives remain activated.

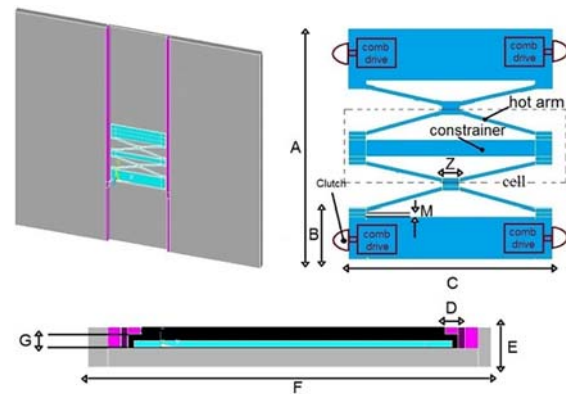


Fig. 1. Structure of the linear micro actuator.

In the normal mode, the entire stoppers are activated and positive voltages are connected to all of the electrostatic comb-drives' electrodes 1, 2, 3, 4 and the electro-thermal slider electrodes' voltage would be zero, please see Fig. 2 (a). To start an up-ward deformation, the DC positive voltage is connected to the electro-thermal slider electrode which causes up-ward expansion; takes into account the comb drives 3, 4 should be cut off, see Fig. 2 (b).

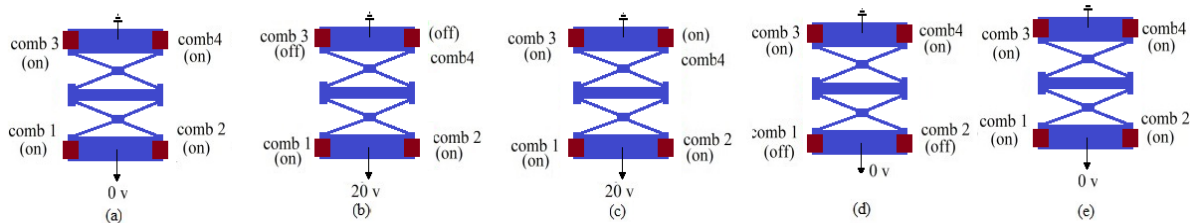


Fig. 2. Motion cycles of the slider.

To save the deformation generated after expanding the slider, the voltages of the comb drives 1, 2, 3, 4 remain ON; ON, actually, here means the positive DC voltage is connected, see Fig. 2(c). To stabilize the whole system and store the built-in up-ward

deformation, the positive DC voltage is applied to all comb drives. The electrode's voltage of the slider and the combs 1, 2 remain cut off so that the slider can be ready to contract through the temperature returning, see Fig. 2 (d). This contraction causes the down part

of the slider have an up-ward movement. To achieve a backward movement, the orders of the applied signals can be reversed. Fig. 3 shows the different voltages applied to the LMA; T_{sw} is the rest time of the slider. In this period of time, the slider stays off., $T_{gw1,2}$ is ON till the slider goes cold. T_{go1} is rest time of the stoppers 1, 2 and T_{go2} is rest time of the stoppers 3, 4 to next cycle. The combs 1, 2, 3, 4 are ON and the slider voltage remains 0 volt. By repeating the process above, the slider can achieve a long linear up-ward movement. For a down-ward movement, all previous steps can be done reversely. To reach a high precision and large travel range, all parameters carefully and separately should take into consideration [19, 9]. To fabricate a prototype of such complex linear actuator, all dimensions, geometries, material properties of the linear actuator narrowly must be calculated and analyzed.

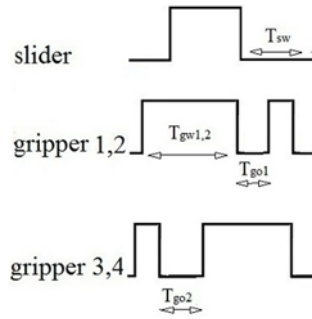


Fig. 3. The signals applied.

2.2. Slider Thermal Equations

Before embarking into designing and simulation of the slider, we take into account analysis of Electro-thermal equations related to the slider. Since length of the hot arms inclined is much larger than height and width, a model simplified shown in Fig. 4, and related deformation calculations can be considered, see Fig. 5. L , h and w are length, height and the width of the model mentioned. By neglecting the heat losses caused by radiation and convection than the ambient, heat conduction would be as following [17, 7, 8]:

$$Q = -KA \frac{\partial T}{\partial x} \quad (1)$$

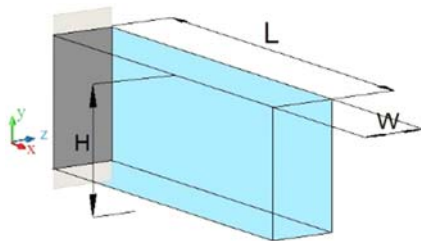


Fig. 4. The Model simplified.

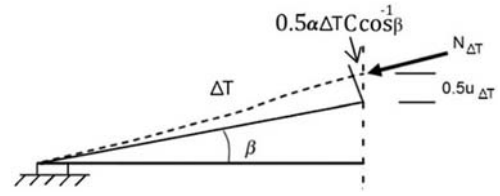


Fig. 5. Mathematical model of the beams.

In the equation above, K , A and x presents the thermal conductivity of polysilicon, temperature, cross-section and variable length of the mode, respectively. By passing the heat through the structure, we have:

$$Q_{\text{input}} = -Kwh \left(\frac{\partial T}{\partial x} \right)_x \quad (2)$$

Output heat conduction is (3):

$$Q_{\text{output}} = -Kwh \left(\frac{\partial T}{\partial x} \right)_{x+dx} \quad (3)$$

The structure can be treated as an electrical circuit by applying an actuation voltage between two electrodes:

$$Q_j = \left(\frac{v}{\rho L} \right)^2 (\rho_0 [1 + \zeta(T - T_0)]) wh dx, \quad (4)$$

where v , ρ_0 , T_0 and ζ are the actuation voltage, the special resistance of the model at temperature of T_0 , the normal temperature of substrate electrodes that consider $293K_0$ and temperature coefficient, respectively.

$$Q_{\text{input}} + Q_j = Q_{\text{output}} \quad (5)$$

By substituting Equations (2) and (4) into Equation (5), and considering $dx \rightarrow 0$, the following equation achieves:

$$K_p \left(\frac{\partial^2 T}{\partial x^2} \right) = -j^2 \rho \quad (6)$$

Boundary conditions are equals to $T(0) = T(L) = T(0)$. Paying attention to the equation (6) and the conditions mentioned, the temperature function can be as following:

$$T(x) = \frac{v^2}{2\rho KL^2} (L-x)x + T_0 \quad (7)$$

By imposing the thermal gradient perpendicular to the model axis, the thermal transfers are:

$$T(x) = (T - T_0)x/L + T_0 \quad (8)$$

$$T(y) = \frac{T_{\text{hot}} - T_{\text{cold}}}{h} y + \frac{T_{\text{hot}} - T_{\text{cold}}}{2}, \quad (9)$$

where T_{cold} and T_{hot} are the cold and hot temperatures, respectively. Deformation in

coordinates $X - Y$ can be shown by the parameter of $v(x)$. Assume waiving the $Y - Z$ planes deformation, the theoretical calculations can be as following:

$$\frac{\partial^2}{\partial x^2} \left(EI \frac{\partial^2 v}{\partial x^2} \right) = F - P \frac{\partial^2 v}{\partial x^2}, \quad (10)$$

where I, P and f is moment of inertia, Young's modulus, load of the model and force, respectively, hence:

$$F = P(x) - \frac{\partial^2 \int \alpha E T y dA}{\partial x^2} \quad (11)$$

Thermal expansion coefficient is α . By solving equation (11) following equation can be found:

$$\int \alpha E T y dA = \frac{\alpha E w h^2}{12} (T_{\text{hot}} - T_{\text{cold}}) \quad (12)$$

Because of waiving the movement in $Y - Z$:

$$\int \alpha E T z dz = M_{Ty} \quad (13)$$

So:

$$\frac{\partial^2 \int \alpha E T y dA}{\partial x^2} = 0 \quad (14)$$

Substituting Equation (13) into (10), shows $F = 0$, hence:

$$\frac{\partial^4 v}{\partial x^4} EI + P \frac{\partial^2 v}{\partial x^2} = 0 \quad (15)$$

Assuming the right end is free or unsupported, $P = 0, \frac{\partial v(0)}{\partial x} = v(0) = 0$. By considering the thermal stresses, the boundary conditions at the free end of a solid beam are:

$$EI \frac{\partial^2 v}{\partial x^2} (L) = - \int \alpha E T y dA, \quad (16)$$

$$EI \frac{\partial^3 v}{\partial x^3} (L) = -P \frac{\partial v}{\partial x} - \frac{\partial \int \alpha E T y dA}{\partial x} \quad (17)$$

According to equation (12):

$$12I \frac{\partial^2 v}{\partial x^2} (x) = -\alpha w h^2 (T_{\text{hot}} - T_{\text{cold}}) \quad (18)$$

Using the boundary conditions and integrating the equation (18), deformation equation is:

$$v(x) = \frac{-\alpha w h^2 (T_{\text{hot}} - T_{\text{cold}}) x^2}{24 \int y^2 dA}, \quad (19)$$

and

$$\int y^2 dA = \frac{w h^2}{12} = I \quad (20)$$

Finally, Equation (19) converts to (21):

$$v(x) = \frac{-\alpha (T_{\text{hot}} - T_{\text{cold}}) x^2}{24h} \quad (21)$$

By substituting equation (7) in equation (21), the deformation value of the slider can be calculated. To calculate deformation of the slider, there is another solution. Variable T is the temperature of the constrainer block, and the temperature variable of the hot arm is $+\Delta T$. Variable T causes deformation in the structure. In other words, the temperature difference cause arm to be elongated. Dimensions of the constrainer especially remain no change and cause the slider to an up-ward deformation. The vertical deformation disp_α , the arms dimensions disp_T and the temperature difference ($\text{disp}_{\Delta T}$) can be approximated by [5, 19]:

$$\text{disp}_\alpha = \text{disp}_T + \text{disp}_{\Delta T} \approx \alpha C (T \sin \beta + \frac{\Delta T}{\sin \beta \cos \beta}) \cong \alpha C (T \beta + \Delta T \beta^{-1}) \quad (22)$$

Assuming similarity of T and ΔT of 2 cells:

$$\text{disp}_\alpha \cong 2 \text{disp}_T = \alpha (T \beta + 2 \Delta T \beta^{-1}) \quad (23)$$

ΔT and $N_{\Delta T}$ approximately can be found:

$$N_{\Delta T} \cong \frac{\alpha \Delta T}{(\sin \beta)^2} + \frac{12 E p (\cos \beta)^2}{0.25 C^2} \quad (24)$$

where p is moment of inertia; this force lonely may cause the arm to be buckled. The axial compressive force in the arm can approximately be found by [5]:

$$N \cong \frac{\alpha \Delta T}{(\sin \beta)^2} + \frac{12 E p (\cos \beta)^2}{0.25 C^2} + \frac{F}{2 \sin \beta} \quad (25)$$

The *LMA* proposed is made of polysilicon. The dimension of the whole *LMA*'s structure clearly has given in Table 1.

Table 1. Dimensions of the slider and the comb drive.

Parameter	Value (μm)	Parameter	Value (μm)
M	35	Z	55
A	1600	G	10
B	350	U	24
C	800	O	14
D	50	R	3
E	100	S	2
F	1000	Y	3

Table 2 shows the material properties of polysilicon. Dimensions are in the micrometer scale and the parameters have been normalized due to *ANSYS* software technical conditions. The slider experiences a stroke of $2.1 \mu\text{m}$, as shown in Fig. 6, by applying a DC voltage 13 volt. A minimum step size

of 1 μm can be achieved, by applying a DC voltage of 7 volt to electrode of the slider. Figs. 7 and 8 illustrates the slider Expansion normally versus to the voltage applied and the time, respectively. Fig. 9, also shows the slider force generated versus the voltage applied. The maximum temperature distribution over the structure is 450 $^{\circ}\text{C}$.

Table 2. The Polysilicon properties used in the simulation.

Para	Value	Unit
Dens	2.33	$\text{pg.}\mu\text{m}^{-2}$
Ex	169	$\text{pg.}\mu\text{m}^{-1}\text{ns}^{-2}$
Prxy	0.28	-
Th ex	2.6e^{-6}	$^{\circ}\text{C}^{-1}$
Res	e^9	$\mu\text{m.}\Omega$
Thermal conductivity	14.9e^{-5}	$\text{pg.}\mu\text{m.}\text{ns}^{-3}\text{C}^{-1}$

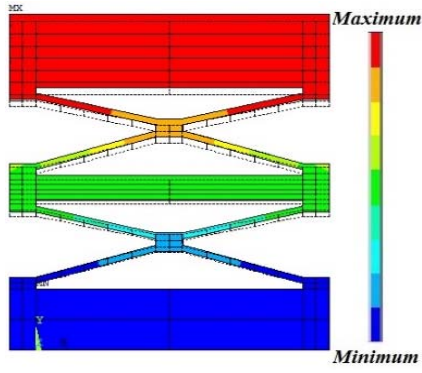


Fig. 6. Contour plots of the slider deformation.

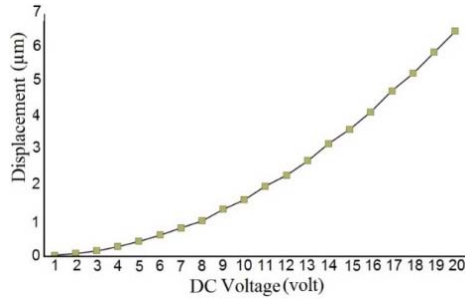


Fig. 7. Expansion of the slider vs. the DC voltage applied.

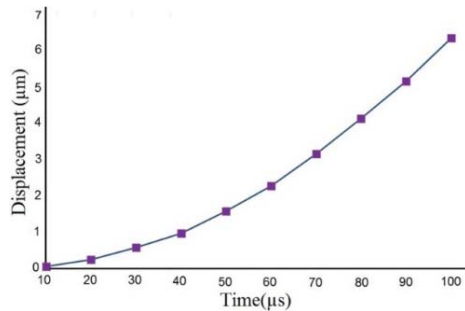


Fig. 8. Expansion of the slider vs. the time.

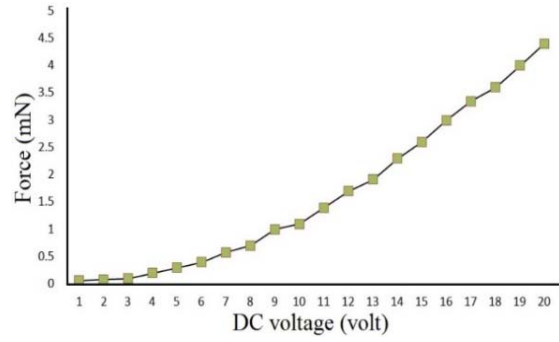


Fig. 9. The force generated of the slider vs. the DC voltage applied.

2.3. Designing of the Stopper

One of the main parts of the actuator proposed, shown in Fig. 10, is the stopper. It mainly consists of three parts; the clutch, the electrostatic comb drives and anchors. The friction generated by clutches, causes the slider stop moving. Some the linear actuators use the clutch and the notch. These types of the actuators generate powerful static force. At beginning of the design, the maximum amount of the force generated by the slider should be considered. Take into consideration which the static friction force generated by the clutches must overcome the slider propulsion force generated. There are two ideas to design the springy clutch. The low springy stiffness of the clutch results a required low electrostatic force; and vice versa. With the high springy stiffness, the clutches require more the actuation voltage to be activated, so there is a trade-off between the springy stiffness presented in the equation (26) and the static friction force; where N is elastic potential of the clutch. Because of using the notches [20], the friction force generated is appropriately large. The comb drives are used to trigger the clutch to prevent the slider from any shaking or movement. The comb drive can ideally be modeled as a set of m conductors. These conductors are embedded in a uniform lossless dielectric medium (see Fig. 11). Conductors have constant electrostatic potential ability. The charges are equally distributed on conductor's surface [14].

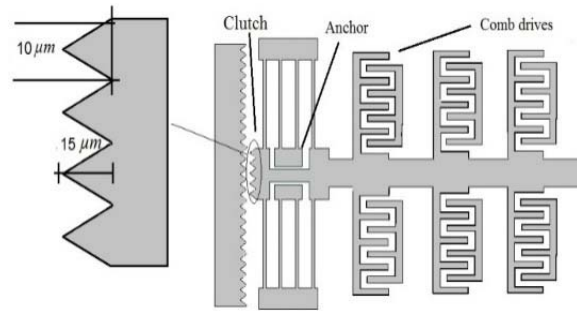


Fig. 10. Schematic of the stopper.

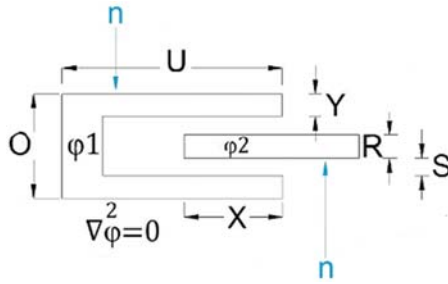


Fig. 11. Schematic view of a comb drive cell.

$$F = \mu_s N, \quad (26)$$

$$q_{i(v)} = \varepsilon \frac{\partial \varphi_i(v)}{\partial n}, \quad (27)$$

where $q_{i(v)}$, ε , φ_i , i and n are the charge density on the surface, the dielectric constant, the conductor electrostatic potential, the internal region and the external region of the conductors, respectively:

$$\nabla^2 \phi = 0 \quad (28)$$

Here i is the number of conductors. By considering the boundary conditions $\varphi = \varphi_i$ and having $i = 1, 2, 3, m$, the surface charge density can be obtained:

$$\phi_i = \sum_{j=1}^m \partial_{sl} \int \frac{\partial \phi}{\partial n}(v') G(v, v') d_{\text{surf}}(v') + \hat{K}, \quad (29)$$

$$Q = \sum_{j=1}^m \partial_{sl} \int \frac{\partial \varphi}{\partial n}(v') d_{surf}(v') \quad (30)$$

where Q , G , v and v' are the total charge, the Green's function, the position vector of source point and position vector of field point, respectively. ∂S_1 is the surface of conductor J , which is zero and K is a constant. Electrostatic force is equal to:

$$f = -\frac{1}{2} \frac{q^2}{\varepsilon} n, \quad (31)$$

$$F = \int f_x ds_\Gamma, \quad (32)$$

where f_x is the component x of the force, Γ is the surface of the moving finger [14]. Therefore, the comb drive dimensions can be established; dimensions are presented in Table 1. Fig. 12(a) and 12(b), respectively, shows the model meshed of the springy clutch and its deformation by applying a force equal to 50 μN . Fig. 13 depicts deformation of the clutch versus the force applied by the comb drives. Fig. 14 shows the force generated by single cell of the comb drive. A force more than 50 μN can be generated by applying a voltage 10 volt to the comb drives which is enough to push the clutches toward the substrate side walls. It should be noted that the electrostatic actuation block must be isolated from the slider's body. This isolation is necessary to prevent errors due to interaction of the electro-thermal and electrostatic

circuits. The material selected to be used as an isolator can be SiO_2 . Figs. 15 and 16 show stress contours of the springy clutch during process; there is a maximum stress of 0.56 GPa and 0.6 GPa, respectively, which is logically acceptable.

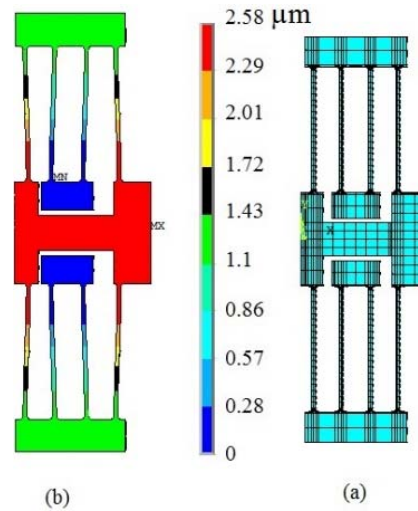


Fig. 12. A model deformed of the springy clutch (a), contour plots of the clutch deformation by applying a force equal to 50 μN and a deformation equal to 2.58 μm .

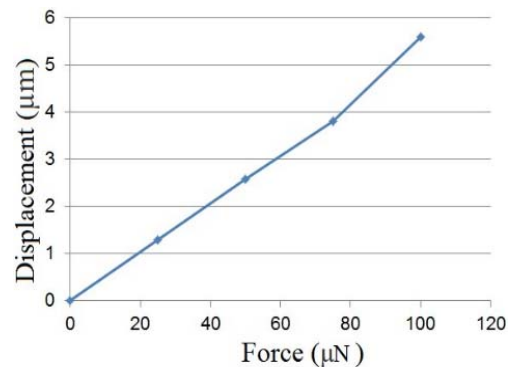


Fig. 13. Deformation of the clutch vs. the force generated of the comb drives.

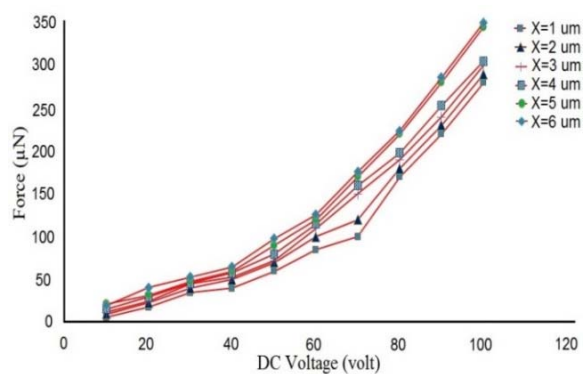


Fig. 14. The force generated of a comb drive vs. the DC voltage applied.

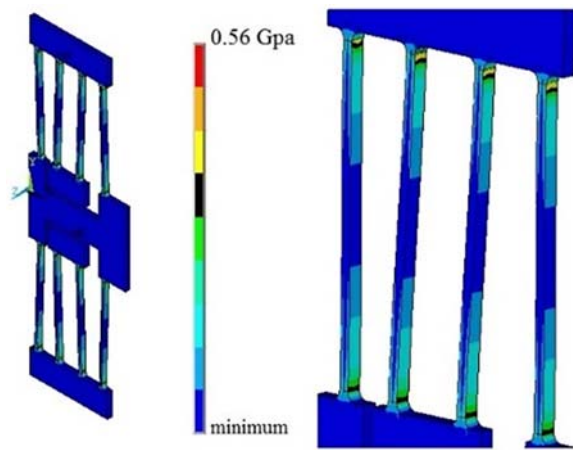


Fig. 15. Stress contour plots of the springy clutch by applying a force equal to 2 mN (+y direction).

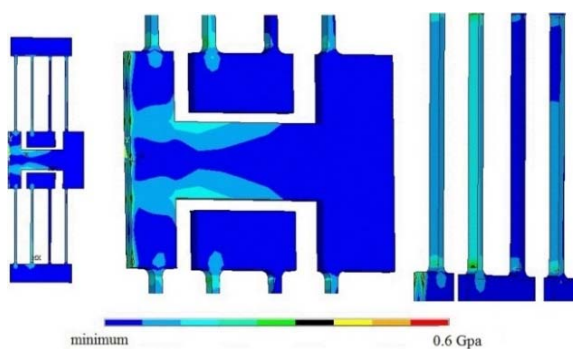


Fig. 16. Stress contour plots of the springy clutch while applying a force equal to 2.2 mN (+y direction).

3. Conclusion

The *LMA* proposed can achieve stroke of $2.1\ \mu\text{m}$ with the acceptable force generated of 2 mN. By repeating the operational sequences numerous times, large travel range can be achieved. The actuator presented has minimum step size about $2.1\ \mu\text{m}$ so that makes this opportunity for CCD Camera's Lens to have a high resolutions and accuracy. To increase the friction force, there are many techniques, but in this work, two ideas were followed; first, increasing the numbers of the comb drives which results increasing whole the system size. Secondly, using impact stator head and the sliders with the edges notched. These two ideas have pros and cons so it should be taken into consideration technologically. The simulation results show that the *LMA* proposed has good speed, high accuracy and large travel range precisely to be used in adjusting CCD Camera's Lens.

References

- [1]. M. Baltzer, T. Kraus, T. Obermeier, A linear stepping actuator in surface micromachining technology for low voltages and large deflections, in *Proceedings of the IEEE Conference Solid-State Sensors and Actuators*, Chicago, IL, 1997, pp. 781-784.
- [2]. K. Chinzei, R. Kikinis, F. Jolesz, Compatibility of mechatronic devices: designs criteria, in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, 1999, pp. 1020-1030.
- [3]. M. J. Daneman, N. C. Tien, O. Solgaard, A. P. Pisano, K. Y. Lau, R. S. Muller, Linear microvibromotor for positioning optical components, *J. Microelectromech. Syst.*, Vol. 5, 1996, pp. 159-165.
- [4]. Q. A. Huang, N. K. S. Lee, Analytical modeling and optimization for a laterally driven polysilicon thermal actuator, *Microsystem Technologies*, Vol. 5, Issue 3, 1999, pp. 133-137.
- [5]. D. Hill, W. Szyszkowski, E. Bordatchev, On modeling and computer simulation of an electro-thermally driven cascaded nickel microactuator, *Sensors and Actuators A*, Vol. 126, 2006, pp. 253-263.
- [6]. Y. Kuang, Q. A. Huang, N. K. S. Lee, Numerical simulation of a polysilicon thermal flexure actuator, *Microsystem Technologies*, Vol. 8, 2002, pp. 17-21.
- [7]. H. N. Kwon, S. H. Jeong, S. K. Lee, J. H. Lee, Design and characterization of a micromachined inchworm motor with thermoelastic linkage actuators, *Sensors Actuators A*, Vol. 103, 2003, pp. 143-149.
- [8]. C. S. Pan, W. Hsu, An electro-thermally and laterally driven polysilicon microactuator, *J. Micromech. Microeng.*, Vol. 7, 1997, pp. 7-13.
- [9]. M. Pai, N. C. Tien, Low voltage electrothermal vibromotor for silicon optical bench applications *Sensors Actuators A*, Vol. 83, 2000, pp. 237-243.
- [10]. J. S. Park, L. L. Chu, A. D. Oliver, Y. B. Gianchandani, Bent-beam electrothermal actuators: II. Linear and rotary microengines, *J. Microelectromech. Syst.*, Vol. 10, 2001, pp. 255-262.
- [11]. J. R. Reid, V. M. Bright, J. T. Butler, Automated assembly of flip-up micromirrors, *Sensors Actuators A*, Vol. 66, 1998, pp. 292-298.
- [12]. T. Risaku, Y. Eui-Hyeok, A normally latched large-stroke, inchworm microactuator, *Micromech. Microeng.*, Vol. 17, 2007, pp. 1715-1720.
- [13]. N. R. Tas, T. Sonnenberg, R. Molenaar, M. Elwenspoek, Design, fabrication and testing of laterally driven electrostatic motors employing walking motion and mechanical leverage, *J. Micromech. Microeng.*, Vol. 13, 2003, pp. N6-N15.
- [14]. Y. Wenjing, M. Subrata, C. M. Noel, Optimal shape design of an electrostatic comb drive in microelectromechanical systems, *Journal of Microelectromechanical Systems*, Vol. 7, Issue 1, 1998, pp. 16-26.
- [15]. H. Yamashita, H. Nozue, E. Nomura, K. Itoh, T. Ema, S. Hirasawa, K. Kojima, T. Tamura, K. Nakajima, Recent progress in electron-beam cell projection technology, *Jpn. J. Appl. Phys. 1: Regul. Pap.*, Vol. 35, Issue 12B, 1996, pp. 6404-6414.
- [16]. R. Yeh, S. Hollar, K. S. J. Pister, Single mask, large force, and large displacement electrostatic linear inchworm motors, *J. Microelectromech. Syst.*, Vol. 11, 2002, pp. 330-336.
- [17]. C. Shao, S. Liu, Y. Hao, K. Ma, Research on electrothermal actuator applied in MEMS safety system, in *Proceedings of the 2nd International Conference on Machine Learning and Computer Application (ICMLCA'21)*, Shenyang, China, 17-19 December 2021, pp. 1-8.
- [18]. R. Saba, S. Iqbal, R. I. Shakoob, M. M. Saleem, S. A. Bazaz, Design and analysis of four-jaws

- microgripper with integrated thermal actuator and force sensor for biomedical applications, *Rev. Sci. Instrum.*, Vol. 92, 2021, 045007.
- [19]. A. S. Algamili, M. H. Md. Khir, J. O. Dennis, A. Y. Ahmed, S. S. Alabsi, S. S. Ba Hashwan, M. M. Junaid, A review of actuation and sensing mechanisms in MEMS-based sensor devices, *Nanoscale Research Letters*, Vol. 16, 2021, 16.
- [20]. S. Evstafyev, V. Samoylikov, S. Timoshenkov, P. Gornostaev, Designing thermal MEMS on a system level, *Proceedings of SPIE*, Vol. 12157, 2022, 121570M.
- [21]. N. M. Kommanaboina, M. F. Pantano, A. Bagolini, Optimization of an amplification mechanism enabling large displacements in MEMS-based nanomaterial testing devices, *Micro and Nano Engineering*, Vol. 15, June 2022, 100131.

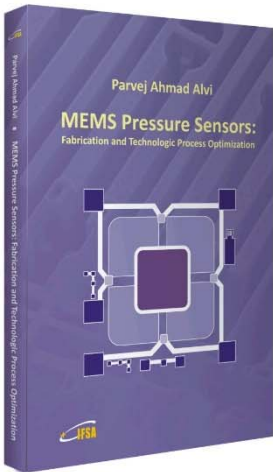


Published by International Frequency Sensor Association (IFSA) Publishing, S. L., 2024
(<http://www.sensorsportal.com>).

Parvej Ahmad Alvi

MEMS Pressure Sensors:
Fabrication and Technologic Process Optimization





Hardcover: ISBN 978-84-616-2207-8
e-Book: ISBN 978-84-616-2438-6

So far, no book has described the step by step fabrication process sequence along with flow chart for fabrication of micro pressure sensors, and therefore, the book has been written taking into account various aspects of fabrication and designing of the pressure sensors as well as fabrication process optimization. A complete experimental detail before and after each step of fabrication of the sensor has also been discussed. This leads to the uniqueness of the book.

Features include:

- A complete detail of designing and fabrication of MEMS based pressure sensor.
- Step by step fabrication and process optimization sequence along with flow chart, which is not discussed in other books.
- Description of novel technique (lateral front side etching technique) in terms of chip size reduction and fabrication cost reduction, and comparative study on both the techniques (i.e. Front Side Normal Etching Technology and Front Side Lateral Etching Technology) for the fabrication of thin membrane.
- Discussion on issues of sealing of conical tiny cavity; because the range of pressure applied (i.e. greater or less than atmospheric pressure) can be decided by methodology of sealing of tiny cavity.
- A complete theoretical detail regarding aspects of designing and fabrication, and experimental results before and after each step of fabrication.

MEMS Pressure Sensors: Fabrication and Process Optimization will greatly benefit undergraduate and postgraduate students of MEMS and NEMS courses. Process engineers and technologists in the microelectronics industry including MEMS-based sensors manufacturers.

Order: http://www.sensorsportal.com/HTML/BOOKSTORE/MEMS_Pressure_Sensors.htm

Application of Bragg Light Diffraction for Determination of Acoustic Wave Frequency

F. R. Akhmedzhanov

Institute of Ion-Plasma and Laser Technologies, TMS Laboratory, 33 Durmon Yuli Str.,
100125 Tashkent, Uzbekistan

Tel.: +998-902121998, fax: +998-712623183

E-mail: akhmedzhanov.f@gmail.com

Received: 5 December 2023 / Accepted: 13 January 2024 / Published: 28 February 2024

Abstract: The possibility of determining the frequency of an acoustically active acoustic wave has been studied. The studies were carried out using the Bragg diffraction method of light on a high-frequency acoustic wave in a lanthanum gallosilicate crystal oriented with high accuracy along the third-order crystallographic axis. Transverse acoustic waves were excited in the range of 0.8-1.6 GHz using lithium niobate piezoelectric transducers. The attenuation coefficient of acoustic waves was previously determined by the pulse ultrasonic method. An expression is obtained for the magnitude of the specific rotation of the plane of polarization of an acoustic wave depending on the effective constant of the acoustic activity tensor and the effective elastic constant. It has been established that specific acoustic activity depends on the frequency of the acoustic wave according to a quadratic law. The results of the study showed that the proposed method makes it possible to determine the frequency of an acoustically active acoustic wave in a dynamic mode with an accuracy of up to 1 %.

Keywords: Acoustic wave, Acoustical activity, Bragg light diffraction, Frequency, Diffracted light intensity, Lanthanum gallosilicate crystal.

1. Introduction

In this work, to determine the frequency of an acoustic wave, it is proposed to use the phenomenon of acoustic activity, which is observed in crystals belonging to one of 21 non-centrosymmetric point symmetry groups. To understand the proposed technology, the phenomenon of acoustic activity is briefly described below. This phenomenon, associated with the spatial dispersion of elastic moduli, was first considered by Andronov [1] and then studied in a number of theoretical and experimental works [2-14], in which some of its regularities and features were revealed, in comparison with optical gyrotropy.

In general, acoustic activity provides useful information about the physical properties of crystals

associated with the periodic structure. In addition, spatial dispersion leads to interesting acoustic phenomena, in particular to the dependence of the phase velocity of an acoustic wave on frequency and other effects [2, 9, 10]. However, in contrast to optical measurements, polarization measurements in crystal acoustics are extremely complex, and therefore this phenomenon has been studied experimentally in a limited number of crystals [8-14].

As in the case of optical activity, acoustic activity manifests itself in the rotation of the plane of polarization of a transverse acoustic wave propagating in a gyrotropic crystal in some special directions [9-12]. In particular, in crystals of point symmetry group 32, including quartz (SiO_2) and lanthanum gallosilicate ($\text{La}_3\text{Ga}_5\text{SiO}_{14}$), this direction is the

crystallophysical axis [001], which is an acoustic as well as an optical axis [8, 9, 13, 14].

Thus, acoustic gyrotropy, like optical gyrotropy, can be considered as one of the types of anisotropy of the medium. Mathematically, it manifests itself in taking into account the first derivative in the expansion of the elastic modulus tensor with respect to the wave vector and is described by a fifth-rank tensor. The equation describing the contribution of spatial dispersion to elastic moduli can be written as [2, 9]:

$$c_{ijkl}(\omega, \vec{q}) = c_{ijkl}(\omega) + i\gamma_{ijklm}(\omega)q_m \quad (1)$$

Here ω , q are the frequency and wave vector of the acoustic wave; c_{ijkl} are the components of the elasticity tensor. γ_{ijklm} are the components of the acoustic activity tensor. It can be shown that the components of the tensor γ_{ijklm} are equal in order of magnitude to the lattice constant [1, 8]. It follows that the influence of spatial dispersion of elastic constants on the propagation of acoustic waves will manifest itself at frequencies approximately equal to or greater than 1 GHz [8, 9]. Taking spatial dispersion into account leads to the following entry for Hooke's law [9]:

$$\sigma_{ij} = (c_{ijkl} + iq_m \gamma_{ijklm} \kappa_m) u_{kl}, \quad (2)$$

where u_{kl} are the components of the strain tensor, κ_m are the components of the wave normal. From here, in the approximation of plane harmonic waves, and using the wave equation for acoustic waves, it can be obtained the following Green-Christoffel equations taking into account spatial dispersion:

$$(\Gamma_{ik} + iqG_{ik} - \rho V^2 \delta_{ik}) u_k = 0, \quad (3)$$

where δ_{ik} is the Kronecker tensor, Γ_{ik} are the components of the Green-Christoffel tensor:

$$\Gamma_{ik} = c_{ijkl} \kappa_j \kappa_l \quad (4)$$

The introduced second-rank tensor G_{ik} is written as [2, 3]:

$$G_{ik} = \gamma_{ijklm} \kappa_j \kappa_l \kappa_m \quad (5)$$

Expression (5) can be conveniently written in terms of the rank 4 pseudotensor G_{mpjl}

$$G_{ik} = \delta_{ikp} G_{mpjl} \kappa_m \kappa_j \kappa_l, \quad (6)$$

where δ_{ikp} is the Levi-Civita tensor.

It should be taken into account that due to the symmetry of the acoustic activity tensor:

$$G_{ik} = -G_{ki} \quad (7)$$

It follows that the diagonal components of the tensor G_{ik} are equal to zero, i.e. $G_{ii} = 0$.

As a result, spatial dispersion leads to the fact that not many transverse waves with the same phase

velocity V_0 propagate along the acoustic axis, but two circularly polarized waves with polarization vectors rotating in opposite directions and with slightly different velocities V_r and V_l .

$$V_{r,l} = V_0 \pm q \frac{G}{\rho}, \quad (8)$$

where ρ is the crystal density, q is the wave number, G is the effective acoustic gyrotropy pseudotensor constant.

Thus, if a plane-polarized transverse wave propagates from the point $z = 0$ along the acoustic axis, then such a wave can be represented as a superposition of two waves, right- and left-polarized. Moreover, at the point $z = 0$ the phases of these waves can be taken equal to zero:

$$\varphi_l = \varphi_r = 0,$$

and after moving the waves to point $z = L$, their phases will change, respectively, by the amount:

$$\varphi_l = \frac{\omega L}{v_l}, \quad (9)$$

and

$$\varphi_r = \frac{\omega L}{v_r} \quad (10)$$

As a result, the resulting vector will rotate by an angle:

$$\varphi = \frac{\varphi_r - \varphi_l}{2} \quad (11)$$

This angle of rotation of the polarization plane can be calculated through the wave velocities of two circular polarizations [2, 9]:

$$\varphi = \frac{\omega L}{2} \left(\frac{1}{v_r} - \frac{1}{v_l} \right) = \frac{\omega L}{2} \cdot \frac{(v_l - v_r)}{v_r \cdot v_l} \quad (12)$$

After some transformations, taking into account relation (8) and that $V_r \approx V_l = V_0$, we obtain an expression for the specific rotation of the plane of polarization δ :

$$\delta = \frac{\varphi}{L} = \frac{\omega^2}{2V_0^2} \cdot \frac{G}{c_{44}} \quad (13)$$

It can be seen that the value of δ is proportional to the square of the frequency of the acoustic wave. Note that in relation (13) for the lanthanum gallosilicate crystal we used, the effective pseudotensor constant G is equal to the component of the acoustic activity tensor γ_{543} .

The influence of spatial dispersion on the angles characterizing the deviation of the energy flow from the acoustic axis in the (100) and (010) planes in crystals of trigonal symmetry has been studied in [10]

on the basis of the phenomenological theory of acoustical activity. The cumbersome relations obtained in [10] are significantly simplified for lanthanum gallosilicate crystals. As a result, the expression for determining the angle of deviation of the energy flow from the acoustic axis has the form:

$$\psi = \pm \arctg\left(\frac{c_{14}}{c_{44}\sqrt{2}}\right) \quad (14)$$

Relationship (14) helps to determine the limitation imposed by the refraction of acoustic waves on the experimental conditions.

2. Samples and Experimental Methods

To experimentally study acoustic activity, the Bragg method of light diffraction by acoustic waves was used. The research sample was prepared from an optically pure single crystal. To exclude the influence of elastic anisotropy on the manifestation of spatial dispersion, the parallelepiped-shaped sample was oriented with an accuracy of no worse than $10'$ along the [001] crystallographic direction, coinciding with the acoustic axis of the crystal.

To correctly set up the experiments, the maximum distance L_{lim} from the beginning of the crystal at which rotation of the plane of polarization can be observed was initially determined. According to [10], this distance is determined by the expression:

$$L_{lim} = 0.5 \cdot S \cdot \text{ctg}|\psi| \quad (15)$$

Here S is the aperture of the piezoelectric transducer, ψ is the largest angle of deviation of the energy flow from the acoustic axis in the (100) and (010) planes. To determine the distance L_{lim} , first, using expression (14), the angle of deviation of the energy flow from the acoustic axis was calculated using the values of the elastic constants c_{14} and c_{44} from [13, 15], which turned out to be equal to 10.5 degrees. Then, using expression (14) and (15) with the aperture of the piezoelectric transducer $S = 4$ mm, it was determined that the effective value of L_{lim} is equal to 1.1 cm.

Thus, measurements of the intensity of diffracted light had to be carried out at distances from the piezoelectric transducer no more than the calculated limit value. As a result, a sample with a length of 1.05 cm and transverse dimensions of 5×5 mm was prepared for the study.

To excite plane-polarized transverse acoustic waves with frequencies from 0.3 to 1.6 GHz, piezoelectric transducers were used in the form of lithium niobate X-cut plates with an aperture of 4×4 mm², with which measurements could be carried out at the indicated frequencies. The piezoelectric transducer was oriented so that the initial polarization of the transverse wave was directed along the X axis.

Previously, the attenuation coefficient of transverse acoustic waves α in the lanthanum

gallosilicate sample was determined using an ultrasonic method in a pulsed mode. The measurements were carried out at frequencies of 0.36, 0.42 and 0.72 GHz. In the ultrasonic system, echo pulses arrived at an amplitude-time selector introduced into the measuring circuit, which opens at a specified time interval, selecting a specific pulse, the amplitude of which is measured by a digital voltmeter. As a result, the attenuation coefficient of acoustic waves is determined from the measured amplitudes of neighboring pulses A_1 and A_2 [16]:

$$\alpha = \frac{\ln\left(\frac{A_1}{A_2}\right)}{2L_0}, \quad (16)$$

where L_0 is the length of the sample. The accuracy of determining the attenuation coefficient due to multiple measurements of pulse amplitudes and subsequent averaging was $\sim 5\%$.

The diagram of the acousto-optic cell used to determine the intensity of diffracted light is shown in Fig. 1. Optical waves with a wavelength of 632.8 nm were generated by a helium-neon laser. Pulses of diffracted light were received, amplified, and converted into electrical pulses using a photomultiplier tube. As a result, it was possible to determine the intensities of diffracted light I_1 and I_2 at different points of the lanthanum gallosilicate sample along the direction of propagation of the acoustic wave, by moving the sample perpendicular to the direction of incidence of the laser beam. The intensities of diffracted light were measured automatically using a computer controlled by specially developed software.

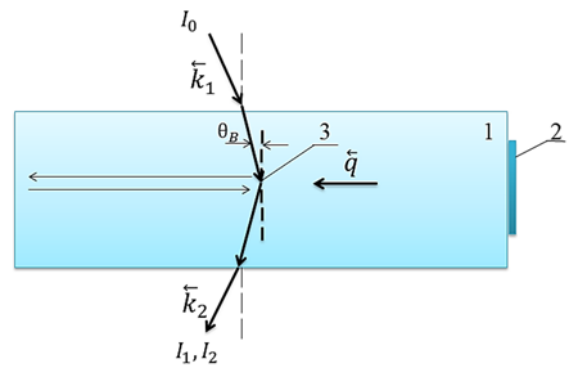


Fig. 1. Diagram of an acousto-optic cell. The geometry of Bragg diffraction is shown at one of the transmission points along the direction of propagation of an acoustically active wave.

In Fig. 1, the following notations are introduced: 1 is $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ sample, 3 is acoustic wave front, 2 is piezoelectric transducer, θ_B is the internal Bragg angle of incidence of the light beam. In the experiment, the magnitude of the specific rotation of the plane of polarization and the attenuation coefficient of transverse acoustic waves propagating along the

acoustic axis were determined from measurements of the dependence of the intensity of diffracted light (I) on the distance (z) from the piezoelectric transducer along the direction of propagation of the acoustic wave. Note that the geometry of Bragg light diffraction, presented in Fig. 1, corresponds to the so-called isotropic diffraction of light by sound [17].

When using Bragg light diffraction to determine the characteristics of acoustic waves, an important question is the intensity of the diffracted light, which is determined by the acoustic-optical quality factor M_2 . This coefficient was introduced by Dixon as a characteristic of the efficiency of Bragg diffraction by acoustic waves [18].

The expression for this coefficient in the case of an anisotropic medium has the form [18]:

$$M_2 = \frac{n_1^3 n_2^3 p_{eff}^2}{\rho V^3}, \quad (17)$$

where n_1 and n_2 are the refractive indices of incident and diffracted light, respectively, ρ is the density, V is the velocity of the acoustic wave, p_{eff} is the effective photoelastic constant. The larger the M_2 value, the greater the intensity of the diffracted light and the higher the sensitivity of the acoustic-optical method compared to other methods of recording acoustic waves.

The effective photoelastic constant in relation (17) is a convolution of the values of the photoelastic constants of the crystal under consideration according to the normalized polarization vectors of the incident and diffracted light μ , β and the direction and polarization of the acoustic wave κ and η [18]:

$$p_{eff} = p_{ijkl} \mu_i \beta_j \eta_k \kappa_l \quad (18)$$

In the considered Bragg diffraction geometry, the wave vector of the transverse acoustic wave is directed along the [001] axis; the polarization of this wave is perpendicular to the wave vector and makes a certain angle ψ with the [100] axis in the plane (001). In turn, the wave vector of light is directed along the [010] axis (up to a small Bragg angle) and the polarization of the incident light is set parallel to the wave vector of the acoustic wave.

It is easy to show that in such a light diffraction geometry, the effective photoelastic constant is determined by the photoelastic tensor component p_{44} and the angle ψ using the relation:

$$p_{eff} = p_{44} \cos \psi, \quad (19)$$

where, in turn, due to the rotation of the plane of polarization of the transverse acoustic wave, the value of the angle ψ is a variable value depending on the specific rotation of the plane of polarization δ , namely $\psi = \delta - z$.

To correctly set up experiments on the use of Bragg diffraction of light to determine the

characteristics of the propagation of acoustic waves, it is also necessary to take into account the following. Crystals of lanthanum gallosilicate are optically uniaxial, therefore, during the diffraction of light on transverse acoustic waves with a rotation of the plane of polarization, features should be observed in these crystals depending on the angles of incidence θ_1 and diffraction θ_2 on the frequency of acoustic waves [19, 20]. This is due to the fact that in an optically anisotropic medium, the refractive indices of incident and diffracted light are generally different.

As a result, the geometry of anisotropic diffraction differs from conventional Bragg geometry in that the angles of incidence and diffraction are not equal [19]. In the geometry we are considering, the plane of light diffraction is the crystallophysical plane (100). In such a geometry, an acoustically active transverse wave propagates strictly along the [001] direction, and its polarization rotates in the (001) plane. The laser beam is incident and diffracted by this acoustic wave at small angles to the [010] direction.

Cross-sections of the surfaces of wave vectors for anisotropic diffraction in optically uniaxial crystals and expressions for calculating the angles of incidence and diffraction of light are considered in detail in [19]. The dependence of the angles of incidence θ_1 and diffraction θ_2 on the frequency of transverse acoustic waves propagating along [001] was determined using the calculation formulas given in [19, 20], but taking into account that in lanthanum gallosilicate crystals $n_e > n_0$ and that the incident light is extraordinary.

The calculated dependences of the angles of incidence θ_1 and diffraction θ_2 on the frequency are presented in the Fig. 2, which shows the pattern of changes in these angles in the studied frequency range.

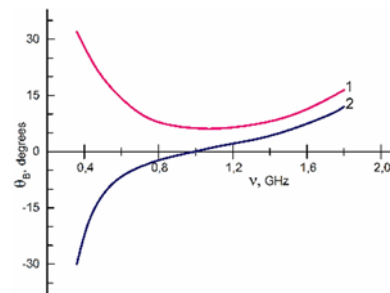


Fig. 2. Dependence of the angles of incidence θ_1 (1) and diffraction θ_2 (2) on the frequency of transverse acoustic waves along the [001] direction at the Bragg diffraction of light in lanthanum gallosilicate crystals.

It can be seen that in the frequency range 0.8-1.6 GHz, the diffraction angle θ_2 changes by approximately 10° with a practically unchanged angle of incidence. Using the values of the refractive indices of lanthanum gallosilicate crystals at the light wavelength $\lambda_0 = 632.8$ nm, equal to $n_e = 1.9107$ and $n_0 = 1.8993$ [15], we obtained the value of the characteristic frequency ν_m^* when the diffraction angle θ_2 is equal to zero:

$$\nu_m^* = \frac{\nu \cdot n_0}{\lambda_0 \cdot n_e} \cdot (n_e^2 - n_0^2)^{\frac{1}{2}} = 1.02 \text{ GHz} \quad (20)$$

At this frequency, the angle of incidence θ_1 is determined by the expression [8]:

$$\theta_{1m}^* = \arcsin \left(\frac{n_e^2 - n_0^2}{2n_e^2 - n_0^2} \right)^{1/2} \quad (21)$$

Calculation using relation (21) gives the value of the angle of incidence θ_{1m}^* equal to $6^{\circ}13'$. Thus, in our experiments, the light fell at such small angle to the wave vector of the acoustic wave, changing slightly in the frequency region of 1 GHz, and diffracted at angles not exceeding $\sim 7^{\circ}$. The considered geometry of light diffraction ensures minimal influence of the diffraction divergence of the acoustic wave, since in this case in the light diffraction does not participate the part of the acoustic beam, whose polarization does not experience rotation due to a large deviation from the acoustic axis.

3. Results and Discussion

In the first series of experiments, the dependence of the intensity of diffracted light (I) on the distance (z) from the piezoelectric transducer along the direction of propagation of an acoustic wave with a frequency of 0.8 GHz was determined. Then the same measurements were carried out at other, higher frequencies. The values of these frequencies changed in steps equal to twice the natural frequency of the piezoelectric transducer. In our studies, this step was approximately 0.06 GHz.

The results of the experiment at frequencies of 1.0 and 1.06 GHz are shown in Fig. 3.

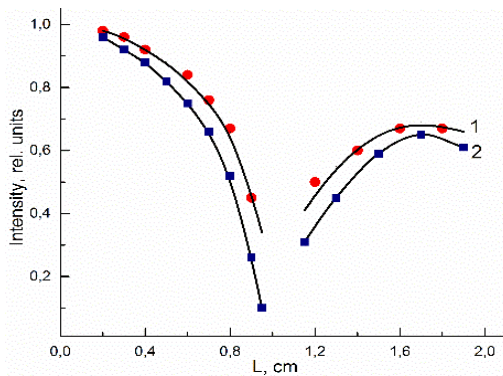


Fig. 3. Dependence of the intensity of diffracted light on the distance L from the piezoelectric transducer at frequencies of 1.01 GHz (1) and 1.06 GHz (2). The dots are the results of the experiment; curves – calculation according to equation (22).

It can be seen that the direction of change in the intensity of the diffracted light changes at the sample boundary (the length of the sample was 1.05 cm). This

effect can be easily explained by the fact that the acoustic wave is reflected from the free boundary of the sample and, accordingly, the direction of rotation of the plane of polarization changes to the opposite direction.

The results obtained on changes in the relative intensity of diffracted light at various points along the direction of propagation of an acoustically active wave can be used to find the characteristics of this wave. These include the initial phase of the displacement vector φ_0 , the attenuation coefficient α and the specific rotation of the wave polarization plane δ . To do this, you can use a set of k experimental values of diffracted light intensities for k independent measurements at different points of the sample.

The intensity of diffracted light can be represented by the following functional dependence of the relative intensity of diffracted light relative to the intensity of incident light (I/I_0) on the distance (z) over which the acoustic wave moves along the $[001]$ axis [9]:

$$I/I_0 \sim p_{44} \exp(-\alpha z) \cos^2(\delta z + \varphi_0) \quad (22)$$

Thus, the problem of finding three independent characteristics of an acoustically active wave is reduced to solving a system of k nonlinear equations of the form (22), when $k > 3$. For $k = 3$, the problem is purely algebraic, but its solution significantly depends on the choice of experimental points at which the intensity was measured, and the result will be noticeably affected by random measurement errors. On the contrary, the more measurement points, that is, the more the system is overdetermined, the more the random errors of individual measurements compensate each other and the solution becomes more reliable.

As is known, in nonlinear problems the solution can be obtained using various methods of linearizing the problem and using least squares estimates. The requirement for the set of required parameters is that, according to Legendre's principle, it minimizes the sum of squared residuals for the redundant system of equations. The described approach was implemented using an application program created in the Pascal language, with a special subprogram that minimizes the sum of squares of the difference function and works using the gradient descent method [21, 22].

The described procedure for determining the specific rotation of the polarization vector and the attenuation coefficient of a transverse acoustic wave propagating along the $[001]$ direction was carried out in the frequency range from 0.8 to 1.6 GHz. The calculation results are shown in Fig. 4, which also shows the values of the attenuation coefficient measured by the ultrasonic method. Solid lines are plotted assuming a quadratic dependence of these characteristics on frequency. It can be seen that the results follow these dependencies well.

In the second series of experiments, the dependence of the intensity of diffracted light on the frequency of acoustically active transverse waves at a

certain point of a crystalline sample used as a measuring transducer was studied. The results of the study are presented in Fig. 5, which shows the change in the frequency of the acoustic wave with a change in the intensity of the diffracted light. It can be seen that by observing a change in the intensity of diffracted light at a certain point of the sample, one can estimate the change in the frequency of the acoustic wave.

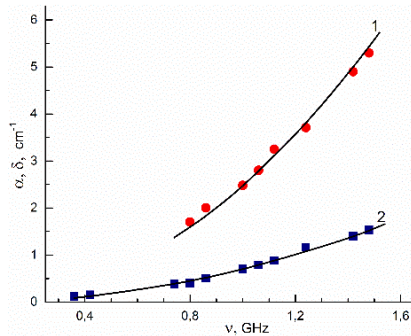


Fig. 4. Frequency dispersion of the specific rotation of the plane of polarization (1) and the attenuation coefficient (2) of transverse acoustic waves along the [001] direction in $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ crystals.

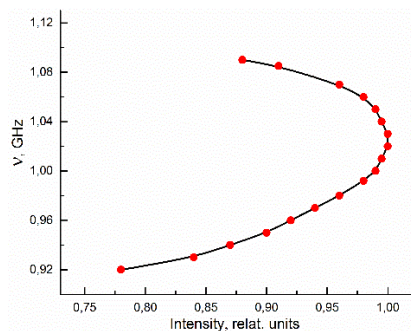


Fig. 5. Dependence of the frequency of transverse acoustic waves along the [001] axis on the intensity of diffracted light at a fixed point of a lanthanum gallosilicate sample.

Having previously calibrated the dependence of the intensity of diffracted light on frequency at a selected point of a crystalline sample used as a meter-sensor, this sample can be used to control the frequency of a transverse acoustic wave.

4. Conclusions

A comparison of the measured frequency values with the frequencies of the radio generator used showed that the proposed method makes it possible to determine the frequency of the acoustic wave with an accuracy of 1 %. Such a relatively large error of the proposed method is due to the influence of the internal conical refraction and the relatively low accuracy of determining the intensity of diffracted light. However, using the special comparison circuits for measuring

light intensity, it is possible to achieve a fairly high accuracy in determining the frequency of the acoustic wave.

The method can be useful for monitoring the stability or deviation of the frequency of the modulating signal in acousto-optic signal processing devices that use gyrotropic crystals. It should also be noted that it is necessary to ensure the broadband operation of the piezoelectric transducer in the controlled frequency range and, of course, the linear mode of operation of the photodetector.

Acknowledgments

The author expresses gratitude to U. A. Saidvaliev for the development and adjustment of special software that makes it possible to measure the intensities of diffracted light in automatic mode, as well as T. Sh. Mustafaev for sample preparation and assistance in measuring the attenuation coefficient of the acoustic wave using the ultrasonic method.

References

- [1]. A. A. Andronov, About acoustical activity, *Radio Physics Series*, Issue 3, 1960, pp. 645-648.
- [2]. D. L. Portigal, E. Burstein, Acoustical activity and other first-order spatial dispersion effects in crystals, *Phys. Rev.*, Vol. 170, Issue 3, 1968, pp. 673-679.
- [3]. K. Kumaraswamy, N. Krishnamurthy, Piezo-acoustic rotation in crystals, *J. Acoust. Soc. Am.*, Vol. 72, Issue 2, 1983, pp. 418-420.
- [4]. K. V. Bhagwat, R. Subramanian. Acoustical activity in the framework of the rotation-gradient theory of elasticity, *Phys. Rev. B*, Vol. 33, Issue 8, 1986, pp. 5795-5800.
- [5]. Y.-y. Li, L. Chen, Theory of acoustical activity, *Phys. Rev. B*, Vol. 36, Issue 18, 1987, pp. 9507-9513.
- [6]. T. P. Srinivasan, Description of acoustical activity using irreducible tensors, *J. Phys. C: Solid State Phys.*, Vol. 21, Issue 23, 1988, pp. 4207-4219.
- [7]. K. V. Bhagwat, R. Subramanian. Acoustical activity of crystals: comparative study of three descriptions, *Acta Cryst.*, Vol. A44, Issue 4, 1988, pp. 551-554.
- [8]. A. S. Pine, Direct observation of acoustical activity in quartz, *Phys. Rev. B*, Vol. 2, Issue 6, 1970, pp. 2049-2054.
- [9]. M. F. Bryzhina, S. Kh. Esayan, V. V. Lemanov, Acoustical activity in SiO_2 crystals, *Journal of Experimental and Theoretical Physics (SU)*, Vol. 25, Issue 11, 1977, pp. 513-515.
- [10]. M. F. Bryzhina, S. Kh. Esayan, Acoustical activity of trigonal crystals, *Solid State Physics (SU)*, Vol. 20, Issue 9, 1978, pp. 2628-2636.
- [11]. Q. Lin, F. Tao, T. Y. Zhang, S. W. Niu, C. Gou, Z. J. Shi, A study of acoustical activity of $\text{Bi}_{12}\text{GeO}_{20}$, *Solid State Communications*, Vol. 54, Issue 9, 1985, pp. 803-806.
- [12]. F. Tao, T. Y. Zhang, S. W. Niu, C. Gou, Z. J. Shi, Q. Lin, Inelastic neutron scattering study of acoustical activity of $\text{Bi}_{12}\text{GeO}_{20}$ and $\text{Bi}_{12}\text{SiO}_{20}$, *J. Acta Physica Sinica*, Vol. 35, Issue 2, 1986, pp. 196-202.

- [13]. F. R. Akhmedzhanov, Acoustical activity in $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ and $\text{Nd}_3\text{Ga}_5\text{SiO}_{14}$ crystals, in *Proceedings of the 14th Symposium on Thermophysical Properties*, USA, Colorado, 2000, p. 227.
- [14]. F. R. Akhmedzhanov, V. V. Lemanov, Acoustical activity in LiNbO_3 crystals at hypersonic frequencies. in *Proceedings of the 7th Symposium on Ferroelectricity*, Russia, St-Petersburg, 2002, p. 132.
- [15]. O. Buryy, N. M. Demyanyshyn, B. G. Mytsyk, A. S. Andrushchak, D. Sugak, Acousto-optic interaction in LGS and CTGS crystals, *Optics Continuum*, Vol. 1, Issue 6, 2022, pp. 1314-1323.
- [16]. F. R. Akhmedzhanov, T. Sh. Mustafaev, J. T. Nazarov, Attenuation of acoustic waves in langasite crystals, *Sensors & Transducers*, Vol. 254, Issue 7, 2021, pp. 38-42.
- [17]. F. R. Akhmedzhanov, Determination of frequency of acoustic waves by Bragg light diffraction method, in *Proceedings of the 5th IFSA Frequency & Time Conference (IFTC'23)*, 20-22 September 2023, Funchal (Madeira Island), Portugal, pp. 15-17.
- [18]. R. W. Dixon, Photoelastic properties of selected materials and their relevance for applications to acoustic light modulators and scanners, *J. Appl. Phys.*, 1967, Vol. 38, pp. 5149-5152.
- [19]. V. V. Lemanov, Acousto-optical study of ferroelectric crystal, *Ferroelectrics*, Vol. 7, Issue 1, 1974, pp. 11-16.
- [20]. R. W. Damon, W. T. Maloney, D. H. McMahon, *Physical Acoustics* (P. W. Mason, Ed.), Vol. 7, Elsevier, 1970.
- [21]. S. Boyd, L. Vandenberghe, *Convex Optimization*, University Press, Cambridge, 2004.
- [22]. A. V. Gasnikov, *Modern Numerical Optimization Methods. Method of Universal Gradient Descent*: Textbook, MIPT, 2018.



Published by International Frequency Sensor Association (IFSA) Publishing, S. L., 2024
(<http://www.sensorsportal.com>).

Universal Sensors and Transducers Interface (USTI-EXT) for extended temperature range

-55 °C ... +150 °C

26 measuring modes for all frequency-time parameters,
rotational speed, capacitance Cx, resistance Rx, resistive bridges
Frequency range, 0.05 Hz ... 7.5 MHz (120 MHz);
Programmable relative error, % 1 0.0005 %
Conversion speeds 6.25 us ... 12.5 ms
SPI, I2C, RS232 (master and slave, up to 76 800 baud rate)
Packages: 32-lead, 7x7 mm TQFP and 32-pad, 5x5 mm (QFN/MLF)

Applications: automotive industry, avionics, military, etc.

Call for Chapters

Your Chapter may be in the next Volume of the popular Book Series

ADVANCES IN SENSORS: REVIEWS

8

Open Access Book Series

Since 2012



- Physical, chemical and biosensors
- Sensor uncertainty, accuracy, calibration
- Smart sensors and systems
- Intelligent sensors and systems
- Artificial Intelligence enabled sensors
- Virtual and soft sensors
- Sensor instrumentation
- Sensor buses and networks
- Signal processing, sensor fusion
- Nanosensors
- Photonic sensors
- Sensor interfacing and signal conditioning
- Materials for sensors
- MEMS, BioMEMS and NEMS
- Remote Sensors and Telemetry
- Sensor packages and reliability

https://www.sensorsportal.com/HTML/IFSA_Publishing.htm

