

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

3 / 2022

**Ranging and Positioning Accuracy for Selected Correlators
under VHF Maritime Propagation Conditions**

K. Bronk et al.

Paper

3

**Analysis of an LSTM-based NOMA Detector Over Time
Selective Nakagami-m Fading Channel Conditions**

R. Shankar et al.

Paper

17

**An Approximate Evaluation of BER Performance for Downlink
GSVD-NOMA with Joint Maximum-likelihood Detector**

N. T. Hai and D. L. Khoa

Paper

25

**Performance Comparison of Optimization Methods for Flat-Top
Sector Beamforming in a Cellular Network**

P. Nandi and J. S. Roy

Paper

39

**An Extended Version of the Proportional Adaptive Algorithm
Based on Kernel Methods for Channel Identification with
Binary Measurements**

R. Fateh, A. Darif, and S. Safi

Paper

47

**Modeling of Microwave Cavities Based on SIBC-FDTD Method
for EM Wave Focalization by TR Technique**

Z. Li, Y. Aimer, and T. H. C. Bouazza

Paper

59

**Joint Optimization of Sum and Difference Patterns with
a Common Weight Vector Using the Genetic Algorithm**

J. R. Mohammed and D. A. Aljaf

Paper

67

(Contents Continued on Back Cover)

Editor-in-Chief

Adrian Kliks, *Poznan University of Technology, Poland*

Steering Editor

Pawel Pławiak, *National Institute of Telecommunications, Poland*

Editorial Advisory Board

Hovik Baghdasaryan, *National Polytechnic University of Armenia, Armenia*

Naveen Chilamkurti, *LaTrobe University, Australia*

Luis M. Correia, *Instituto Superior Técnico, Universidade de Lisboa, Portugal*

Luca De Nardis, *DIET Department, University of Rome La Sapienza, Italy*

Nikolaos Dimitriou, *NCSR "Demokritos", Greece*

Ciprian Dobre, *Politechnic University of Bucharest, Romania*

Filip Idzikowski, *Poznan University of Technology, Poland*

Andrzej Jajszczyk, *AGH University of Science and Technology, Poland*

Albert Levi, *Sabancı University, Turkey*

Marian Marciniak, *National Institute of Telecommunications, Poland*

George Mastorakis, *Technological Educational Institute of Crete, Greece*

Constantinos Mavromoustakis, *University of Nicosia, Cyprus*

Klaus Mößner, *Technische Universität Chemnitz, Germany*

Imran Muhammad, *King Saud University, Saudi Arabia*

Mjumo Mzyece, *University of the Witwatersrand, South Africa*

Daniel Negru, *University of Bordeaux, France*

Ewa Orłowska, *National Institute of Telecommunications, Poland*

Jordi Perez-Romero, *UPC, Spain*

Michał Pióro, *Warsaw University of Technology, Poland*

Konstantinos Psannis, *University of Macedonia, Greece*

Salvatore Signorello, *University of Lisboa, Portugal*

Adam Wolisz, *Technische Universität Berlin, Germany*

Tadeusz A. Wysocki, *University of Nebraska, USA*

Publications Staff

Content Editor: **Robert Magdziak**

Managing Editor: **Ewa Kapuściarek**

Technical Editor: **Włodzimierz Macewicz**

on-line: ISSN 1899-8852

© Copyright by National Institute of Telecommunications, Warsaw 2022

Ranging and Positioning Accuracy for Selected Correlators under VHF Maritime Propagation Conditions

Krzysztof Bronk, Magdalena Januszewska, Patryk Koncicki, Adam Lipka, Rafał Niski, and Błażej Wereszko

National Institute of Telecommunications, Warsaw, Poland

<https://doi.org/10.26636/jtit.2022.162422>

Abstract — The article presents an analysis of the features of selected correlators impacting the accuracy of determining the receiver's range and position in VHF marine environment. The paper introduces the concept of various correlators – including the double delta correlator – and describes the proposed measurement scenarios that have been designed to demonstrate the effectiveness of those components. The entire work was performed as part of the R-Mode Baltic and R-Mode Baltic 2 projects, with our goals including analyzing the impact of multi-path phenomena, changes in the sampling frequency or signal type on the determination of the received signal delay at the receiver. The measured data were processed in a signal correlation application and in a TOA-based tool in order to determine the receiver's position. This process made it possible to compare the selected correlating devices. The results presented in this article are to be used by IALA in developing a current version of the VHF data exchange system's (VDES) specification.

Keywords — correlators, e-navigation, maritime radiocommunications, positioning, ranging, R-Mode Baltic, VHF data exchange system

1. Introduction

Precise determination of location is a key factor in aquatic and marine environments. The process is based primarily on satellite systems. However, relying solely on satellite systems is risky due to GNSS jamming, spoofing or a potential global failure. Relying only on RTK or DGNS fixes is not sufficient as well. Hence the development of e-navigation services, i.e. an approach that combines modern navigation and communication technologies, thus creating an accurate, safe and effective system that is expected to be available for use by ships of all sizes [1]. The necessity of such a solution is essential in coastal areas, where availability of security systems is limited and where satellite signals are often disturbed. Therefore, if satellite and terrestrial systems are used simultaneously for location purposes, the number of ranging signals observed is increased and the geometry is improved. All this ensures better positioning accuracy and shortens the positioning lead time.

Ranging Mode for the Baltic Sea – R-Mode Baltic [2] and R-Mode Baltic 2 [3] – is one of the projects dedicated to the emerging e-navigation services. It is carried out by a consortium of 12 partners, from the Baltic Sea region, including the National Institute of Telecommunications, Poland. The scope and results of this project are presented in [4]. As part of the R-Mode Baltic and R-Mode Baltic 2 projects – whose main goal is to develop a non-GNSS marine positioning system – selected correlators have been researched and implemented, which allowed for the determination of more precise pseudoranges from terrestrial transmitting stations, and thus – for more accurate determination of the receiver's position at sea. Based on theoretical analyses and simulations, a software tool was developed and implemented in an R-Mode positioning system demonstrator. It was also tested in a laboratory and used during a measurement campaign performed in the marine environment – using the VHF radio maritime channel. The aim was to select a specific correlator that exhibits the best efficiency under multi-path propagation conditions or in a scenario in which the quality of the received useful signal is poor.

The article describes the entire process of researching the correlators known from [5]–[7], starting with a theoretical analysis, through the implementation stage, laboratory tests performed under various measurement scenarios, all the way to measurement campaign tests conducted with the use of a prototype transmitter and receiver.

2. Selection of the Optimum Correlator for R-mode Baltic System

Signal correlators are a very important component of radiocommunication and navigation systems. They are used to determine the delays of RF signals reaching the receiver, based on which the receiver's position can be determined. Such devices rely on correlation methods that incorporate numerous measuring approaches, including analysis of noisy signals. By using the autocorrelation, it is possible to de-

termine the extent to which signal values at a certain point in time will affect the signal at a given point in the future. These methods are also used for detecting and measuring parameters of periodic signals against the background of random interference, for detecting gravitational waves, as well as in space radar technology, in communicating with distant probes or in radio astronomy. Correlators are also used in transmittance and delay time measurements, in the prediction and filtering theory, in identifying energy and noise sources and in determining system properties based on specific input and output data. Their main advantages include the ability to analyze low-power signals which are additionally affected by noise from atmospheric disturbances or receiving devices [8]. The correlation technique is a relatively simple and effective method of detecting such signals. Due to very low power signals and the large amount of data that need to be collected and processed, this technique has specific features in radio navigation that distinguish it from other applications. In correlators, two measurement signals are subjected to cross-correlation. The cross-correlation function of two stationary random signals $x(t)$ and $y(t)$ can be expressed by:

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)y^*(t + \tau)dt, \quad (1)$$

where T – observation time and τ – time shift.

Such a function brings out the similarity between $x(t)$ and $y(t)$. However, if $x(t)$ and $y(t)$ are independent, then for each value of τ , function $R_{xy}(\tau)$ takes the value equal to zero, provided that $x(t)$ or $y(t)$ has the zero mean value. This enables the measurement of random signals occurring against the background of random independent disturbances. Let us now consider the cross-correlation function of two disturbed random signals. Suppose we have $z_1(t)$ and $z_2(t)$ containing a random useful signal $x(t)$ [8]:

$$z_1(t) = x(t) + n_1(t), \quad (2)$$

$$z_2(t) = x(t) + n_2(t), \quad (3)$$

where $n_1(t)$ and $n_2(t)$ are random disturbances present in $z_1(t)$ and $z_2(t)$, respectively. If $x(t)$ and $n_1(t)$, $x(t)$ and $n_2(t)$, as well as $n_1(t)$ and $n_2(t)$ are independent, then cross-correlation function of signals $z_1(t)$ and $z_2(t)$ is:

$$R_{z_1 z_2}(\tau) = R_{xx}(\tau), \quad (4)$$

where $R_{xx}(\tau)$ is the autocorrelation function of the useful signal. Function $R_{z_1 z_2}(\tau)$ reaches, at zero, a maximum equal to the mean square value $\overline{x^2(t)}$ of the useful signal. The measuring procedure based on formulas (1)–(4) consists in determining the cross-correlation function between two versions of the disturbed signal which was obtained, for example, as a result of amplification in two different paths of the measuring system. In real applications, the cross-correlation of signals hidden in noise and delayed against each other is the most interesting feature. Assume the transmitted signal $x(t)$ is a stationary random with a zero mean value. Let the received signal be a stationary random signal $y(t)$ with a zero mean value such that:

$$y(t) = ax(t - \tau_0) + n(t), \quad (5)$$

where: a – attenuation coefficient, τ_0 – delay, $n(t)$ – independent noise (disturbance) with zero mean value. The cross-correlation function of signals $x(t)$ and $y(t)$ is:

$$R_{xy}(\tau) = aR_{xx}(\tau - \tau_0). \quad (6)$$

The peak value of $R_{xy}(\tau)$ occurs at $\tau = \tau_0$ and is $\overline{ax^2(t)}$. This problem can be extended to a situation where each of the two signals is delayed, attenuated and disturbed, where the disturbances $n_1(t)$ and $n_2(t)$ are independent of the useful signal $x(t)$ and of one another [8]:

$$z_1(t) = a_1x(t - \tau_1) + n_1(t), \quad (7)$$

$$z_2(t) = a_2x(t - \tau_2) + n_2(t). \quad (8)$$

In such a case the cross-correlation function takes the following form:

$$R_{z_1 z_2}(\tau) = a_1 a_2 R_{xx}[\tau - (\tau_2 - \tau_1)]. \quad (9)$$

The peak values of $R_{z_1 z_2}(\tau)$ occurs for $\tau_0 = \tau_2 - \tau_1$ and is $a_1 a_2 \overline{x^2(t)}$.

A correlator based on the correlation function (9) is capable of obtaining appropriate synchronization. Under real life conditions, the main correlation peak of the correlation function may take a different shape (it may be either more jagged or smoother). This depends primarily on signal strength, noise power, multipath propagation, or the sampling frequency value used in the receiver. In order to obtain the most accurate information about the delay time of the signal reaching the receiver, three correlators were analyzed:

- basic correlator [5],
- narrow correlator [6],
- double delta correlator [7].

Figure 1 shows the baseband signal processing diagram for a single channel with particular emphasis placed on the correlators. Integrate and dump (I&D) blocks accumulate the correlators’ outputs and provide their in-phase (I) and quadrature (Q) components. The number of correlator pairs depends on the specific type of the correlator used in the analysis of the correlation function [9]:

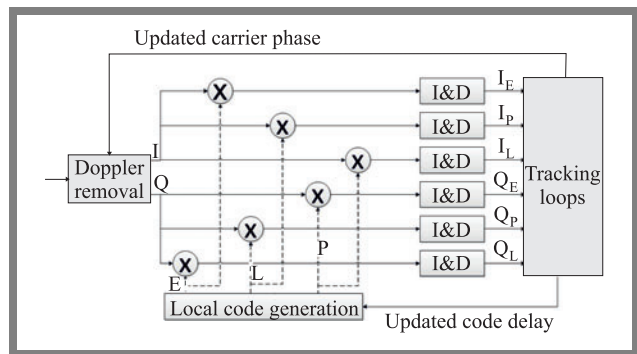


Fig. 1. Multicorrelators block diagram.

The basic correlator is the least complicated solution. Its output depends solely on the maximum sample value in main peak of the correlation function. In this solution, the correlator relies only on one locally generated code referred to as the prompt (P) replica [5]–[9]. When the code is correlated with

the matched replica of itself, the correlation result reaches the maximum – meaning a high degree of autocorrelation. When the code is correlated with a non-aligned replica of itself – the correlation result is low. Real-time systems are very much prone to noise and the conditions always vary. As a consequence, the autocorrelation peak seems to change constantly, resulting in the need of adding continuous phase and frequency (Doppler) tracking to match the code replica.

A more in-depth analysis of the correlation function is performed by the narrow correlator. It uses the sample with the maximum value and a pair of adjacent samples. The correlator is based on three replicas of the local code: prompt (P), early (E) and late (L) [6]–[10]. The idea behind the narrow correlator is to reduce the spacing between E and L correlators, so that interference immunity and accuracy of ranging can be improved. In the narrow correlator the effect of multipath signals on the correlation function is least significant at the peak value. Thus, designing the correlators in the vicinity of the maximum value can effectively reduce the influence of multipath. Each of those correlators estimates the correlation function value for a different sequence offset. The P correlator calculates the value of the correlation function for the current phase of the sequence, while E and L correlators use the accelerated and delayed sequence with respect to the P correlator. Relative time shift of the sequence between successive correlators is usually half the duration of the sequence’s elementary symbol (chip). The correlation function is shaped like a triangle within ± 1 chip around the maximum. The phases of the sequences in E , P and L correlators must lie within this triangle. If the values at either of the outputs of E or L correlators exceed the output of the P correlator, the phase of the locally produced sequence is updated (Fig. 2) [6].

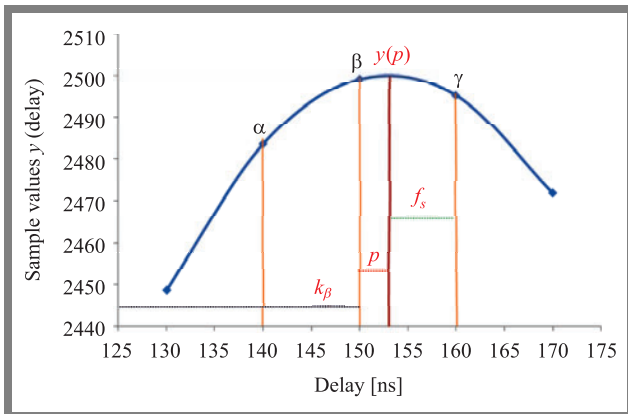


Fig. 2. Example of a correlation peak analyzed in a narrow correlator.

Having obtained the sample values of the correlation function at the output of the correlators, it is now possible to calculate the exact delay of the signals on the basis of the dependencies:

$$p = \frac{1}{2} \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \cdot \frac{1}{f_s}, \quad (10)$$

$$k^* \triangleq k_\beta + p, \quad (11)$$

$$y(p) = \beta - \frac{1}{4}(\alpha - \gamma)p, \quad (12)$$

where: f_s – sampling frequency, p – calculated delay in relation to the maximum sample where the point with the highest value is located, α, β, γ – values of three consecutive samples of the correlation function, k_β – delay of the sample with the highest value, k^* – total delay of the point with the highest value, $y(p)$ – maximum value of the correlation function.

The double delta correlator is a more advanced solution. It uses two pairs of correlators (instead of one) in order to estimate signal delay with high compensation of multipath effects [7]–[11]. The correlator is based on five replicas of the local code: prompt (P), early 1 (E_1), late 1 (L_1) as well as early 2 (E_2) and late 2 (L_2). The double delta correlator introduces a correction term that can compensate variations of the rising and falling edges caused by multipath. E_1 and L_1 are one pair of output correlator with an earlier and later correlation peak (d_1 spacing), while E_2 and L_2 are the second output pair with an earlier and later correlation peak (d_2 spacing) [2]. The distance can be calculated as:

$$D = (E_1 - L_1) - \frac{1}{2}(E_2 - L_2). \quad (13)$$

Since the differences $E_1 - L_1$ and $E_2 - L_2$ can be interpreted as “narrow correlators”, Eq. (13) can be:

$$D = \text{Narrow}(d_1) - \frac{1}{2}\text{Narrow}(d_2). \quad (14)$$

After the selection of the correlators, a full test campaign was performed on a $\pi/4$ -QPSK signal generated by the VHF data exchange system (VDES) [13], [14] software simulator [15]. For testing purposes, a second signal was generated by shifting half of the sample from the first signal and by adding low-power noise. This helps to determine the effectiveness of the implemented correlators. On the receiving side, the correlation function was obtained. The sampling frequency of 10 MHz was used, where one sample relates to 30 m of distance.

Figure 3 shows the error in estimating correlation peak delay in relation to the exact value of the delay in which the peak is located. This error is expressed in meters – the result is the distance between the determined correlation peak and the real signal delay value. The graph has been presented as a function of the signal-to-noise-ratio (SNR) in order to verify the efficiency of correlators with respect to signal strength.

The worst results were observed for the basic correlator. Therefore, the correlation function based on just one correlation peak is not capable of providing reliable data. A lot of information is lost regarding the exact delay of the received signal. Hence, this method is not suitable for beacon services due to the potentially significant pseudo-range errors. On the other hand, narrow and double delta correlators with the second pair spacing = 0.15 chip allow for a more precise delay estimation, as shown in Fig. 3. Better immunity to multipath propagation phenomena and the larger spread of the correlation function around the maximum value peak makes the double delta correlator a proposal that is most suitable for navigation applications.

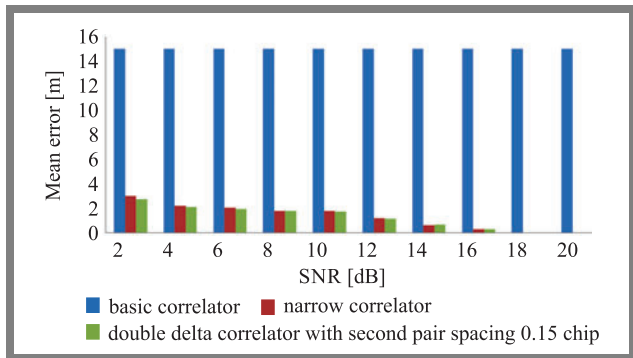


Fig. 3. Error of the determined correlation peak delay as a function of signal strength.

3. Multipath Propagation Impact on the Effectiveness of Correlators

The propagation conditions of a given radio channel depend on the properties of the wave itself, i.e. its length and polarization, as well as on the features of the environment in which the wave propagates. These include, for example, the topography and the type of the surface – radio waves propagate differently in areas covered with water, in forests, in urban or in open areas. Multipath transmission is a phenomenon that substantially impacts the signal in a given radio channel. The term “multipath” means that the signal – subjected to diffraction, refraction, scattering and reflection – reaches the receiver as a sum of many signals with different delays, phases, and amplitudes [16], [17]. Additionally, in the proximity of the receiver, each of the signal components is dispersed into other N components. If the receiver is in motion, the carrier frequency of each scattering component is shifted based on the Doppler effect.

The delay of multipath signals depends on the additional distance traveled by the reflected signal, with the said distance being longer than that of the direct path. During the measurement campaign organized by the National Institute of Telecommunications, the transmitter was installed in the Gdynia harbor, where, for the initial short-range measurements, the multipath effect was caused by nearby objects or buildings. On the other hand, for measurements at sea, the signal could also propagate to the receiver through reflections from maritime infrastructure, e.g. ships, ports or breakwaters. The presence of land (Hel Peninsula) along the line of sight also impacted the reflected signals reaching the receiver. The next important phenomenon is the tropospheric waveguide (duct) channel (Fig. 4) [18]. This is a specific type of radio propagation that tends to occur during periods of stable, anti-storm weather only.

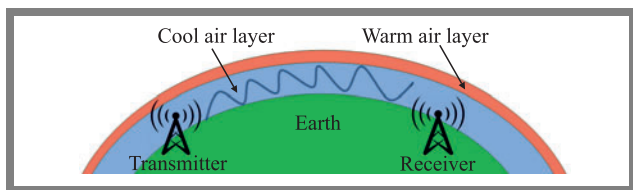


Fig. 4. Phenomenon of tropospheric ducts.

When instead of a normally expected drop, an RF signal encounters an increase in temperature at high altitudes (temperature inversion), the higher refractive index of the atmosphere will cause the signal to bend. Especially favorable conditions for the formation of tropospheric ducts occur mainly in the second half of the year [19]. Migrations of high-pressure areas observed in autumn, a large load of moisture in the atmosphere and the daily temperature fluctuations have a beneficial effect as well. Tropospheric ducts affect all radio frequencies, and signals amplified by this phenomenon are capable of reaching locations up to 1300 km away.

Radio waves that reach the receiver as multipath signals are superimposed (with a slight delay) on the main signal, which causes an abrupt unevenness in the shape of the correlation peak. If, on top of that, noise is also present along the propagation path, these distortions are even larger. Consequently, the calculation of the transmitter – receiver range can be difficult. The application of a correlator with an additional pair of correlators might be an efficient solution to that problem. Choosing appropriate spacing between additional correlators will cause the irregularities of the correlation peak to be covered along the entire range of the correlator’s operation and, consequently, the impact of the multipath phenomenon will be significantly reduced and the quality of pseudo-range determination will be improved.

As part of the research concerning the effectiveness of the proposed method, a simulation was carried out in which the signals reaching the receiver were delayed relative to each other by such a value that the distorted correlation peak was located between the first and second pair of the correlators.

The reflected signal reached the receiver 25 ns later and with less power compared to the one received directly. In Fig. 5, the effectiveness of the analyzed types of correlators for the above simulation test is shown. Signals with $\pi/4$ -QPSK modulation and a sampling frequency of 10 MHz were used.

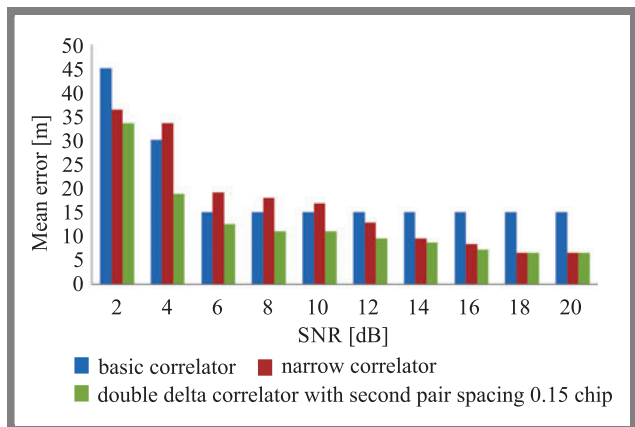


Fig. 5. Error of the correlation peak delay determined for the measurement scenario.

The results show that even for signals with a low SNR, the double delta correlator with a second pair spacing of 0.15 chip determines the delay of the transmitted signal more accurately compared to other solutions. It takes into account the distortion of the correlation peak due to the delay of the sig-

nal's replica. This means that it is an effective concept that can be used in a multipath environment to reduce the error in determining the time of arrival of a useful signal to the receiver.

4. Impact of Correlators on the Real R-mode Signal

In this section, correlators used under real conditions are presented and their influence on the accuracy of determining the receiver's position in maritime VHF channels is discussed. While many ships are equipped with sensors and assistance systems for positioning and navigation, collisions and groundings are still taking place. To reduce the risk of such incidents, a test stand was built under the R-Mode Baltic project for a ground-based positioning system called Ranging Mode (R-Mode) in the Baltic Sea [20]. This new system enables positioning even when Global Navigation Satellite Systems (GNSS) fail or are unavailable. The provision of reliable position, navigation and timing (PNT) information is key to safe navigation and is also essential to the development of new marine e-navigation applications. As part of this project, the effectiveness of selected correlators was evaluated as well.

During the measurement campaigns, the transmitting R-Mode station was located in the Gdynia harbor, while the receiving station was aboard the "Stena Baltica" ferry, with the exception of the stationary measurements when it was at a fixed location in the harbor of Jastarnia (as part of the R-Mode Baltic 2 project). All these prototype stations were developed by the authors of the article at the National Institute of Telecommunications.

Figure 6 presents a block diagram of the R-Mode demonstrator and the received signal processing path relied upon in order to obtain the most accurate receiver position. The purpose of operating and synchronizing the R-Mode VDES system has been described in detail in [4].

The R-mode system demonstrator presented in Fig. 6 consists of the following modules:

- RF module – which receives radio transmissions from the R-Mode stations and stores the I/Q samples,
- external rubidium oscillator,
- signal correlation application – which reads files with I/Q samples, correlates them and, thereafter having information about the coordinates of the stations, determines the pseudoranges,
- GNSS receiver – which provides the UTC time, GNSS position, speed, and track angle. The first two parameters are used for accuracy comparison purposes, while the other two are fed to the Kalman filter,
- R-Mode real-time positioning application – which determines the position based on the calculated pseudoranges to the R-Mode stations and additionally uses Kalman filtration. It also determines positioning errors with respect to GNSS position.

The deployment of this system made it possible research the correlators' effectiveness with respect to determining pseudoranges and position accuracy. The selection of the

best correlator and its configuration will allow to obtain more accurate location results in the upcoming measurement campaign, in the final R-Mode Baltic 2 demonstration.

4.1. Spacing Between Correlator Pairs, Sampling Frequency and their Impact on Efficiency

As part of the research, a number of modifications were introduced to the selected transmission signal and sampling frequency, which gave us the opportunity to increase the accuracy of the determined range. The aspect of changing the sampling frequency is particularly important due to the ability of building the target VDES R-Mode receiver with the use of software-defined radio technology, using off-the-shelf components. A number of scenarios have been verified during the measurement campaigns performed in the point-to-point mode, i.e. with the VDES R-mode base station located in the Gdynia harbor and with remote access to the receiving station installed in the harbor of Jastarnia.

Initially, all tests and measurements were carried out at a sampling frequency of 200 MHz and using a correlation sequence which consisted solely of the Gold sequence with a length of 1877 symbols. This resulted in a large number of files with the recorded samples and, consequently, longer time of processing the signal by correlators. The subsequent tests were conducted with lower frequencies and various correlators, which allowed to choose the optimal combination presented in Table 1. There is a summary of the root mean square error (RMS) of the determined distance accuracy in relation to the actual distance between the transmitter in Gdynia and the receiver in Jastarnia. In the case of the double delta correlator, the second pair spacing was set to 0.15 chip.

Tab. 1. Determined RMS error for selected correlators depending on the sampling frequency.

Sampling frequency [MHz]	RMS for the basic correlator [m]	RMS for the narrow correlator [m]	RMS for the double delta correlator with the second pair spacing of 0.15 chip [m]
200	98.17	93.31	92.22
100	90.32	91.81	90.19
50	94.35	93.12	91.04
10	90.74	84.25	81.41
1	100.53	82.2	79.65

The values presented in Table 1 have been obtained for the specified sampling frequency and for the Gold sequence signal. The tests for each sampling frequency were performed on a group of 1,000 recorded files with samples. The basic correlator utilizing a single correlation peak produced the worst results. With each decrease in the sampling frequency, the error resulting from the inaccuracy of one sample was higher. Narrow and double delta correlators produced the best results, with the latter being slightly more accurate. The high power of the received signals impacted the shape of the correlation peak, making its slopes smoother and, therefore,

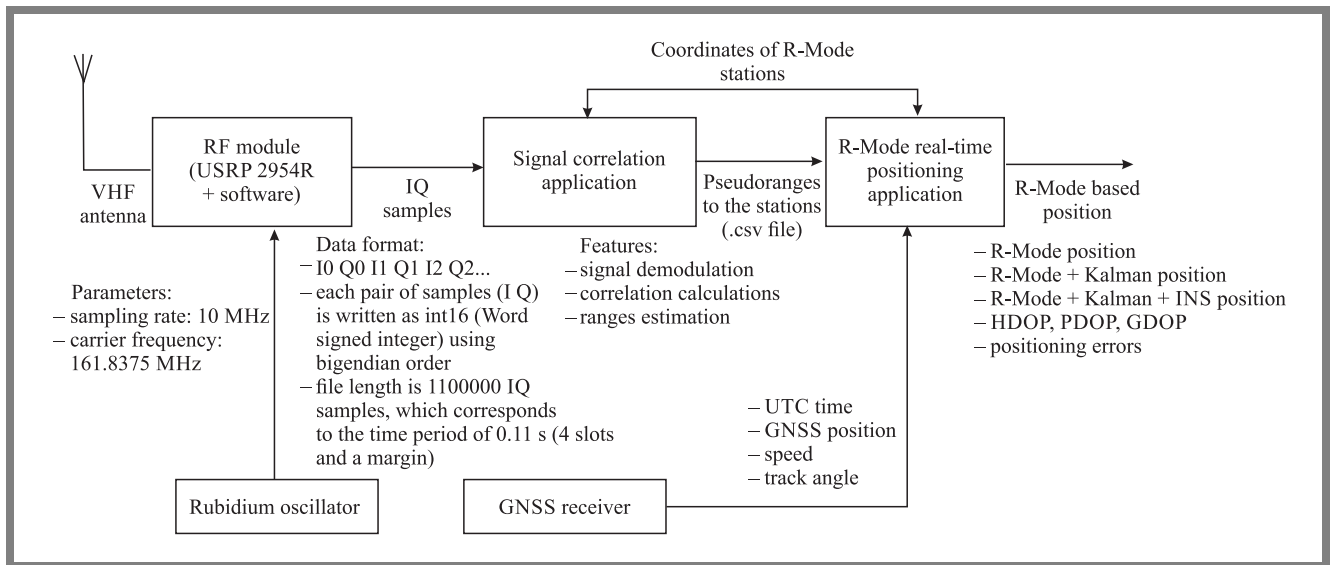


Fig. 6. Block diagram of the R-Mode positioning system demonstrator.

preventing the multicorrelators from exhibiting any major range determination deviations.

Next, we conducted the measurements, with varying transmitted signals. Each of the 4 emulated stations sent, once per second, a known ranging sequence as a part of the payload data [21]. In this case, the ranging sequence was a combination of two sequences (defined below) to customize the required performance based on two given scenarios:

- shorter distances between shore station and ship,
- longer distances between shore station and ship.

The first part of the ranging sequence was based on the $\pi/4$ -QPSK modulation alphabet with alternated constellation points (the so-called “alternating sequence”) [21]. The second part of the ranging sequence was a Gold code ($\pi/4$ -QPSK modulation). The length of each sequence is based on the weighting factor γ . This means that the length of the entire correlation sequence (1877 symbols) is multiplied by the appropriate weighting factor associated with the sequence type. Both sequences were weighted and merged with each other. To be more precise – the alternating sequence was multiplied by a weighting factor of $\gamma = 0.7$ for short distances (higher SNR) and $\gamma = 0.3$ for larger distances (lower SNR). The Gold code was multiplied by a weighting factor of 0.3 for short distances and 0.7 otherwise. Figure 7 shows the principle of creating such a correlation sequence.

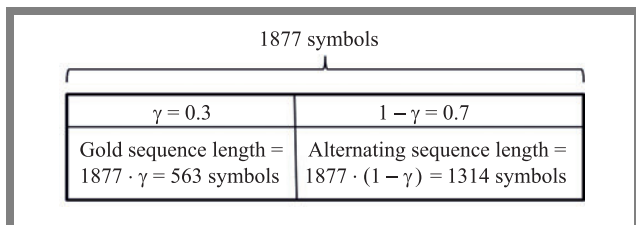


Fig. 7. A method of creating a correlation sequence consisting of two different types.

Table 2 shows the RMS values obtained for different γ coefficient and the correlator type for both transmitter and receiver

side. The obtained RMS was also analyzed in the context of the correlation of signals with different values of γ .

Tab. 2. RMS determined for selected correlators depending on the γ coefficient.

Factor γ	RMS for the based correlator [m]	RMS for the narrow correlator [m]	RMS for the double delta correlator [m]
$\gamma = 0.7$ (TX) correlated with $\gamma = 0.7$ (RX)	48.15	18.67	17.58
$\gamma = 0.3$ (TX) correlated with $\gamma = 0.3$ (RX)	50.37	20.88	20.54
$\gamma = 0.7$ (TX) correlated with $\gamma = 0$ (RX)	84.22	45.12	44.68
$\gamma = 0.3$ (TX) correlated with $\gamma = 0$ (RX)	70.7	45.01	44.52

The combination of the Gold signal and the alternating sequence is characterized by very good correlation properties. It can be noticed that this sequence improved the ranging accuracy by approximately 50 m compared to the scenario in which the Gold sequence was used solo. That is a significant improvement compared to previous studies. Again, the double delta correlator offered the best results, slightly better than those of the narrow correlator. The solution presented in [24] is one of the methods for evaluating the effectiveness of the correlator. This approach involves the calculation of the S-curve, the code multipath envelope and the thermal noise. Here, the authors propose a single universal model for the double delta correlator used in the marine environment. In order to select the most effective configuration, tests were carried out on a number of signals transmitted and received under marine environment conditions. For the purpose of fur-

the research, the effectiveness of the double delta correlator was tested depending on the spacing width of the second pair of correlators (Fig. 8).

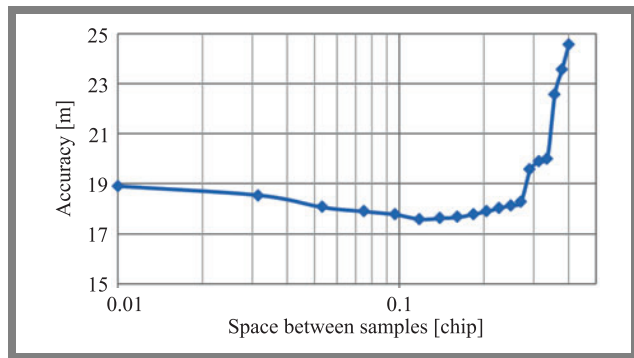


Fig. 8. Range estimation accuracy as a function of the spacing of the second pair of correlators in the double delta correlator.

Based on 1,000 received waveforms, the simulation of accuracy was performed by varying the spacing of the second pair of correlators. The most effective spacing width is approximately 0.1–0.2 symbol duration between the E_2 and L_2 correlators. Using a larger spacing is not rational, because samples that go beyond the correlation peak are taken into account and introduce additional errors.

To recapitulate, one may state that due to the possibility of analyzing a wider range of samples included in the correlation peak, double delta correlator is the best choice for large-scale applications in the marine environment. At the same time, the basic correlator was found not to be suitable for such purposes.

4.2. Correlator Type vs. Accuracy

After the selection of the best signal in terms of correlation properties, it was possible to research the selected correlators’ of pseudo-range accuracy. In order to finally calculate the receiver’s position using the calculated ranges, a measurement scenario was used where additional base stations were emulated¹. The coordinates of the emulated stations were chosen in such a way so that their respective distances from the receiver in Jastarnia were exactly the same as the distance between Jastarnia and the “real” transmitting station in Gdynia (Fig. 9). It is also worth noting that tests were performed in the Gulf of Gdańsk, in a line-of-sight (LOS) marine environment.

During the tests, a 4-slot message was transmitted and was recognized by the receiver as four signals from the emulated transmitters. Using the collected data sent via the VHF marine channel, the RMS factor was determined for:

- basic correlator,
- narrow correlator,
- double delta correlator with the second pair spacing of 0.1 chip,

¹Since only one “physical” base station exists so far (installed in the Gdynia harbor), the emulation of other stations was necessary. Otherwise, we would not be able to calculate the receiver’s position, because at least three stations are required.

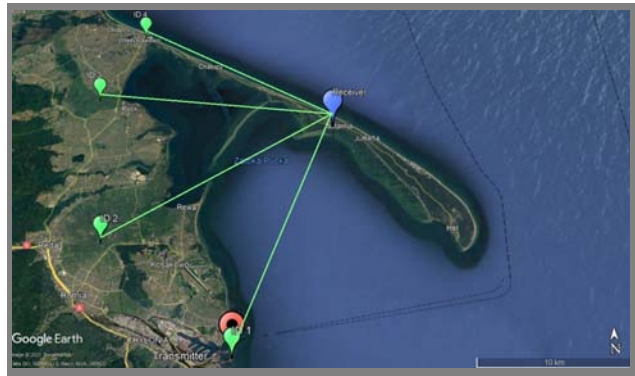


Fig. 9. Measurement scenario for four emulated broadcasting stations.

- double delta correlator with the second pair spacing of 0.2 chip.

All analyses, including digital signal processing, were carried out using a signal correlation application that was developed in-house.

The application includes various data processing modes: offline, useful for processing the collected data, or online, where samples are processed in real time, e.g. during the measurement campaign on a ship. The application allows to select the correlator to be used, with the possibility to set the required spacing in the double delta correlator. After measurements and signal processing, the accuracy of the determined distances of selected correlators was compared. Table 3 shows the RMS error values for each correlator based on approximately 3,000 measurements.

Tab. 3. RMS determined depending on the selected correlator.

		Emulated station			
		ID 1	ID 2	ID 3	ID 4
RMS [m]	Basic	31.798	33.816	38.122	33.771
	Narrow	24.228	24.324	24.352	24.264
	Double delta with second pair spacing of 0.2 chip	21.898	21.996	21.940	21.915
	Double delta with second pair spacing of 0.1 chip	21.580	21.675	21.593	21.597

As we can see, for the double delta correlator, the spread of the second pair of correlators translates into the achieved accuracy range. The narrow correlator generated narrow correlator generated less accurate results, which may be due to the fact that one pair of the correlators is not able to detect all the distortions that may have occurred in the correlation peak. Therefore, an additional pair of correlators allows to include all useful information contained in the correlation peak, which facilitates the determination of a more precise pseudo-range. With 0.2 chip spacing, however, the samples that were beyond the range of the correlation peak could potentially be included in the processing stage, which might have negatively impacted the results obtained. Uncorrelated noise that falls within the scope of the correlator analysis may

also impair range determination outcomes. In this case, due to the high signal level, the such differences were low, but under less favorable conditions i.e. with a higher noise level and with the shape of the correlation function not being as smooth, the impact may be significant. With theoretical research and the simulations performed taken into consideration, the optimum value of spacing between the second pair of samples is 0.1 chip. This reduces the multipath phenomenon and the impact of uncorrelated, additional noise. The results obtained with the use of the basic correlator were characterized by the lowest level of accuracy, as stated in the previous analyses. This confirms that for a lower sampling frequency, the error that results from the accuracy of one sample is too large, which makes this type of correlator unsuitable for use in navigation services.

In the next step, data from each of the four correlators was transferred to the software application for determining the position based on the obtained pseudo-ranges (this specific tool also was created by the authors, in-house). Based on the conducted research, it will be possible to determine the manner in which the selection of the correlator affects the determination of the receiver's position in a marine VHF channel.

The application allows to determine the position using the time of arrival (TOA) method [22], based on the measured pseudo-ranges from reference stations with known locations. The application displays the coordinates of the calculated position, positioning error, DOP coefficients, the number of visible reference stations, the calculated receiver clock bias, speed and course [23]. The map shows the calculated and actual positions of the receiver and the reference stations.

For each correlator, three scenarios were assessed in order to take into consideration different numbers of reference stations and varying geometries. Table 4 lists these scenarios along with the information on the number of reference stations used, HDOP values, with ID numbers referring to Fig. 9, while Fig. 10 visualizes the specific scenarios. The reference stations are marked in green, while the receiver is marked in blue. In each case, the position was calculated 2,741 times for each correlator. The same set of recorded samples was used as input for each of the correlators. This allows to assess the dependence of the positioning accuracy on the correlator applied, regardless of the noise level present in the channel.

Tab. 4. Positioning scenarios.

	Number of stations	HDOP	Emulated stations ID
Scenario 1	4	3.392	1, 2, 3, 4
Scenario 2	3	3.731	1, 2, 4
Scenario 3	3	9.199	2, 3, 4

For a clear visualization of the obtained results, Fig. 11 shows the detailed data for scenario 2 only, while Table 5 lists the RMS values obtained for each tested correlator for all scenarios.

The largest difference in terms of position determination accuracy was observed between the basic correlator and

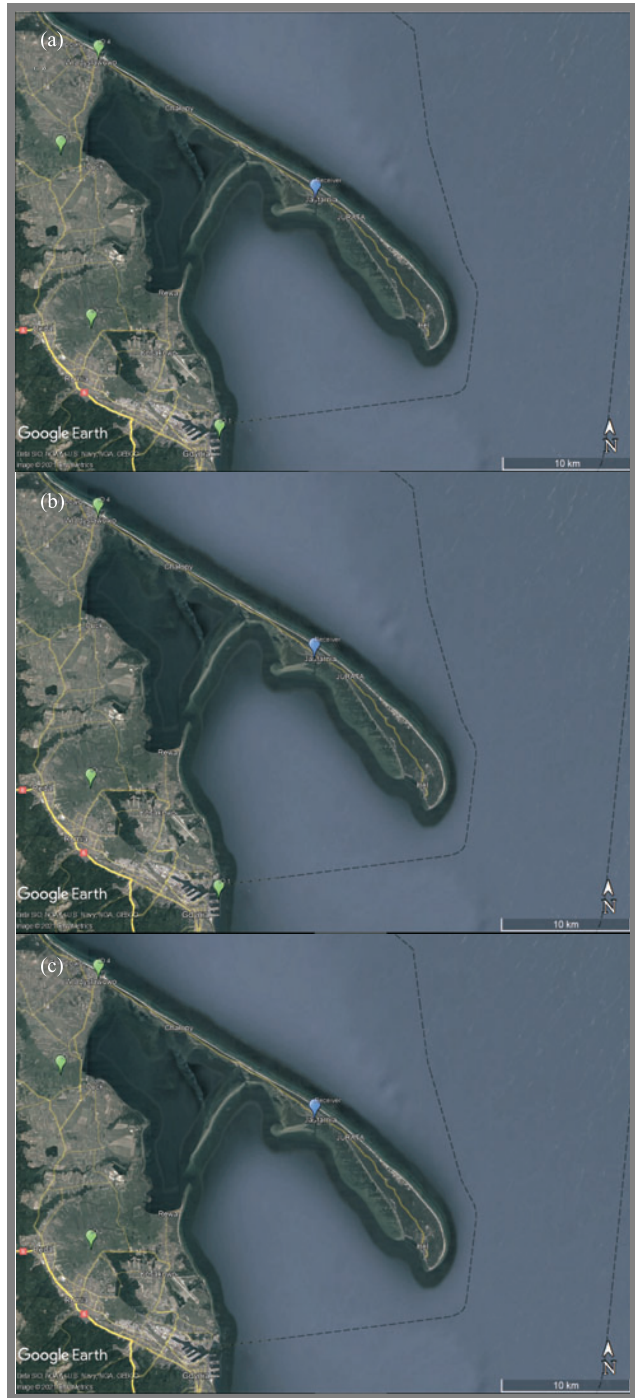


Fig. 10. Position reference stations used in the campaign: a) scenario 1 – four stations and good geometry, b) scenario 2 – three stations and good geometry, c) scenario 3 – three stations and poor geometry.

the remaining types. Regardless of the number of stations and the HDOP coefficient, all three correlators (narrow and two variants of double delta) performed significantly better than the basic correlator. As expected, the positioning error value increased for each of the correlators if the geometry of the reference stations deteriorated and if their number was decreasing.

Simultaneously, no significant differences in terms of position determination accuracy were observed between these three correlators. In each case, the RMS values were sim-

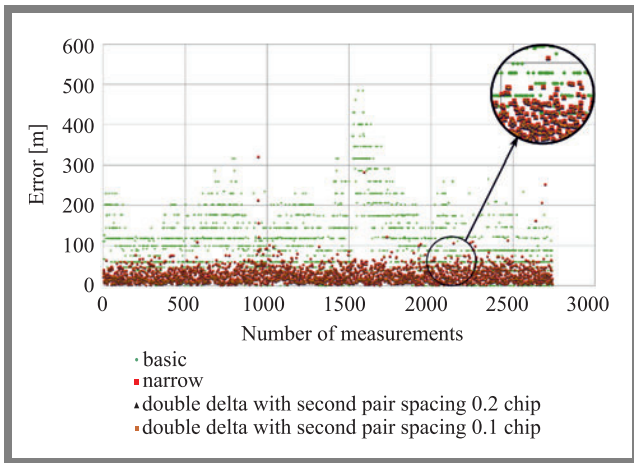


Fig. 11. TOA positioning error for scenario 2 for: basic correlator (green), narrow correlator (red), double delta correlator with second pair spacing of 0.1 chip (blue), and double delta correlator with second pair spacing of 0.2 chip (orange).

Tab. 5. Determined RMS of positioning error.

Correlator	RMS [m]		
	Scenario 1	Scenario 2	Scenario 3
Basic	130.443	133.053	271.791
Narrow	30.146	32.807	80.564
Double delta with second pair spacing of 0.2 chip	30.135	32.795	80.537
Double delta with second pair spacing of 0.1 chip	30.135	32.975	80.537

ilar. Nevertheless, since the differences in the accuracy of pseudo-range determination were small, the lack of ability to achieve an improvement in positioning accuracy may result from the emulation, because only one real reference station was used in the campaign. We can assume that if the reference signals from each of the stations are completely independent of each other (e.g. different propagation paths), the benefit of using a double delta correlator will become noticeable.

4.3. Type of the Correlator and the Accuracy of the Determined Range in the R-mode Campaign

Paper [4] presents a detailed analysis of the results obtained by the authors during the measurement campaign with the use of the VDES system and the basic correlator. As follow up, another measurement campaign was performed in June 2020 on the Gdynia-Karlskrona route, with a VHF transmitting antenna located in the Gdynia harbor and the receiver placed aboard a Stena Line ferry.

Figure 12 shows a graph presenting the ranging errors observed during the measurements. The red line shows the distance between the receiver and the transmitter. The blue points represent the error resulting from the difference between the correlator’s output and the reference measurement (by EGNOS + GNSS). For better visualization, the chart is di-

vided into three parts in which the measurements took place: LOS – i.e. the part of the measurements carried out in the Gdańsk Bay, mixed sea + land path – i.e. measurements that took place when the ship was behind the Hel Peninsula and measurements in the NLOS environment – i.e. with the ship outside the 50 km zone from the transmitter and out at sea.

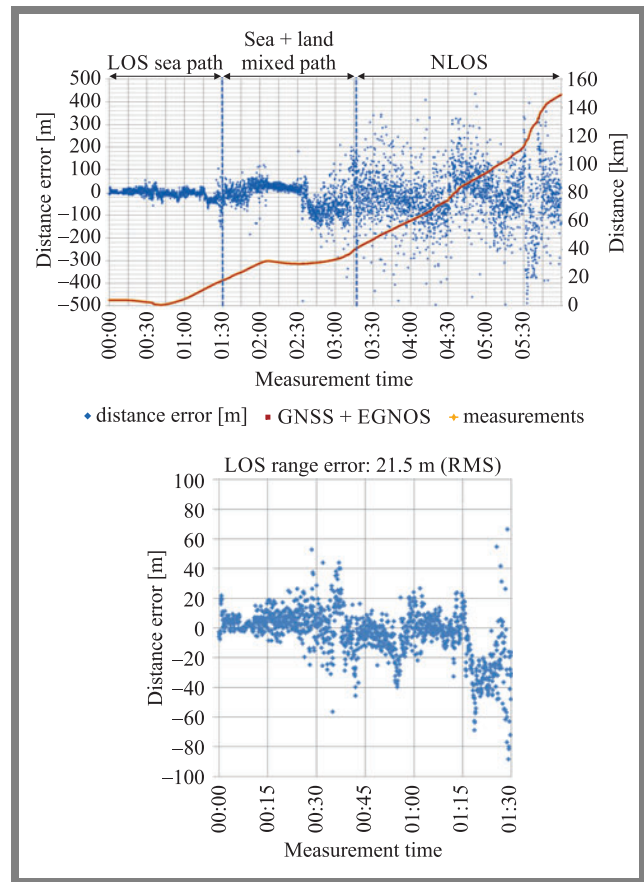


Fig. 12. Ranging errors in the VDES 2020 measurement campaign for the double delta correlator with second pair spacing of 0.1 chip.

Due to some modifications applied to the hardware configuration in the second measurement campaign, it is not possible to clearly assess the degree to which the double delta correlator contributed to the improvement of the calculated RMS. To be more specific, compared to the first VDES measurement campaign, the following changes were made, which could have an impact on the final results:

- using VDES signal without Hann window – to improve transmitted signal power,
- setup transmission without physical VHF filter – to improve transmitted signal power,
- using amplifier with higher output power levels,
- incorporating different LNA and filter configuration on receiver side.

For example, in the case of the LOS environment, an RMS of 20 m was obtained with the double delta, whereas in the previous measurement campaign these values were over 30 m. Figure 13 shows a map presenting the measurement campaign’s route with the applied errors.

For short distances, it can be seen that the ranges provided by GNSS almost coincide with those calculated by the sig-

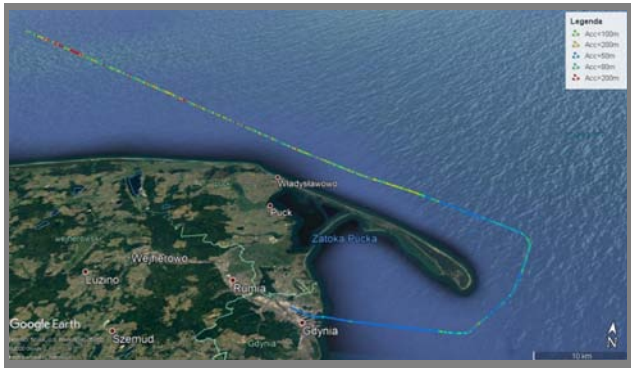


Fig. 13. Map of the 2020 Gdynia – Karlskrona measurement campaign.

nal correlation application. Such results allow for calculation of the exact position of the receiving station. This is very important, especially in ports where accurate, precise navigation is critical. For long distances (over 66 km) the error of positioning is approximately 180 m. This is a very good result compared to previous measurement campaigns. It should also be added that at a distance of approximately 120 km, these accuracies were in the order of 300 meters, which is still a satisfactory value. The RMS curve for the basic and double delta correlator for the second VDES measurement campaign is shown in Fig. 14.

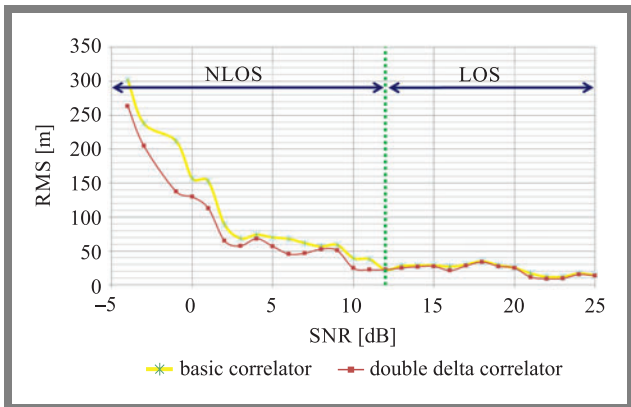


Fig. 14. RMS curve for second measurement campaign.

In another step, a comparison of the efficiency of basic and double delta correlators in the second VDES measurement campaign was researched using samples recorded at a sampling frequency of 200 MHz. The difference between the obtained RMS is presented in Fig. 15.

For the LOS zone, the differences in the calculated distances are imperceptible, due to the high sampling frequency. In contrast, for NLOS, the double delta correlator has a higher RMS value resulting from multipath propagation. e.g. in the area behind the Hel Peninsula, where there are many obstacles. In relation to the basic correlator, this improves the determination of distances, allowing to achieve accuracy of up to 30 m. Future plans assume that four transmitting stations will be set up in the Baltic Sea. Then, it will be possible to fully verify the effectiveness of the double delta correlator in determining the receiver’s position.

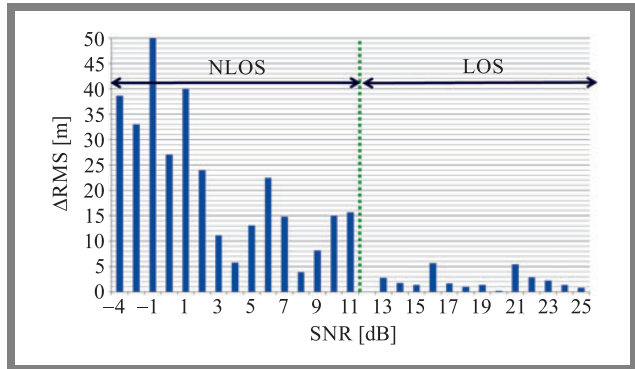


Fig. 15. Difference of the RMS between the double delta and the basic correlators.

5. Long-term Stationary Measurements Using a Best Correlator

As a follow up to the R-Mode Baltic project (ended on 03/31/2021), the authors continued their research within the framework of the Ranging Mode Baltic Sea test bed evaluation project (R-Mode Baltic 2). The main goal of the R-Mode Baltic 2 project is a long-term evaluation of the R-Mode Baltic test stand and additional testing of new R-Mode concepts. To achieve that goal, the project consortium will increase the monitoring capabilities of the R-Mode Baltic test stand and equip vessels with R-Mode-ready receivers and marine applications from the R-Mode Baltic project. This expanded network of static and dynamic monitoring stations will be used for extensive R-Mode performance studies over a project period of nine months. The results are essential for the further development of the proposed solution and will facilitate its ultimate transformation into a reliable and internationally recognized backup maritime navigation system.

Such an approach allows to study the features of the double delta correlator and check its effectiveness and stability by means of long-term measurements conducted in various weather conditions. The preliminary results that were obtained from the 11-day campaign are presented here. The transmitter was installed in the Gdynia harbor. The EIRP power was 25 W and the antenna height was 28 m above sea level. The receiver was located in the harbor of Jastarnia – approx. 20 km away, with its antenna positioned at 17 m above sea level. The transmission between the stations took place under LOS conditions and entirely over a sea-covered area. Both stations were equipped with rubidium oscillators. The receiver was also equipped with a low noise amplifier (LNA) (noise figure of 0.6 dB) and a VHF bandpass 3 dB filter. Fig. 16 shows the results obtained at the beginning of the measurements.

The RMS is presented for two cases: when the mean error from the measurements was subtracted every day and when it was subtracted just once on the first day of the campaign. From Fig 16, the potential accumulation of the mean error could be observed, which confirms the stability of the measurements. The graph indicates the atmospheric factors (wind direction,

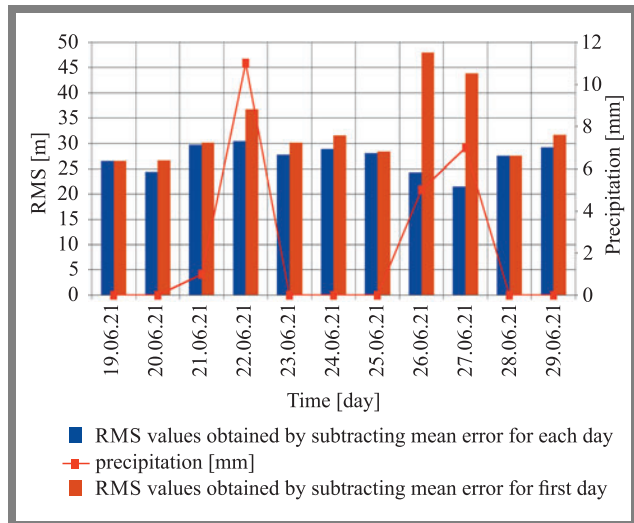


Fig. 16. Analysis of the observed RMS depending on the weather conditions.

precipitation) that could affect the results. The phenomenon of ducts, introduced in Section 3, can be observed in Fig. 17.

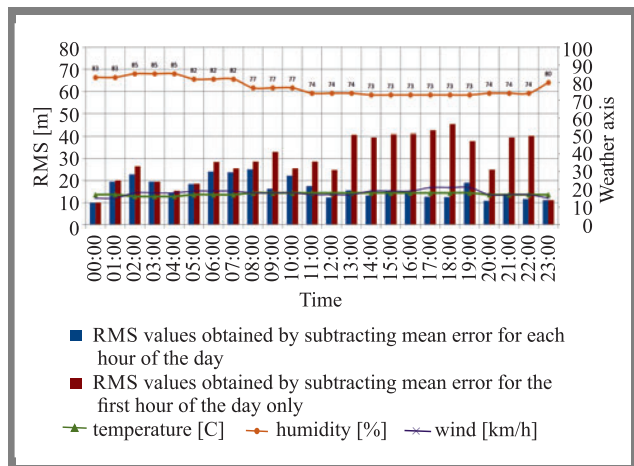


Fig. 17. The phenomenon of radio ducts visible during the measurements conducted on 26th June 2021.

Additionally, Fig. 16 presents the impact of rain on signal attenuation resulting in the ducts phenomenon, which shows a clear temporary increase in the accumulated RMS. One may also see, however, that it did not affect the RMS values obtained by subtracting the mean error for each hour. By the end of the project, it will be possible to collect a large amount of data, thanks to which it will be possible to analyze the measurements obtained from the double delta correlator in terms of RMS changes depending on the season of the year or time of the day. In addition, the National Institute of Telecommunications is preparing a research focusing on time and frequency synchronization of the R-Mode system with optical fibers. Thanks to such an approach, an opportunity would arise to compare stationary measurements using a rubidium oscillator with measurements obtained in a scenario in which synchronization is achieved by means of a fiber optic solution. This would ensure the elimination of time error sources, thus greatly enhancing the quality of data.

6. Conclusions

It is the double delta correlator (with the second pair spacing = 0.1 chip), selected for signal propagation on the VHF channel, that offers the highest level of accuracy in terms of range and position determination in the marine environment. The analysis conducted has shown the sheer number of factors that influence the results of the studied correlators. These included multipath propagation, sampling frequency, selection of signal modulation and its structure, spacing between pairs of correlators, and weather conditions prevailing during the measurements.

The research campaign was divided into a theoretical phase, followed by a series of simulation tests and concluded with measurement campaigns relying on the VDES R-Mode system demonstrator. The double delta correlator displayed the best properties as far as the analysis of the correlation function was concerned. For signals received along the Gdynia-Jastarnia route (with the distance equaling approx. 20 km), where the SNR of the received signal was about 2 dB, the accuracy of the calculated distance was 21.58 m. The results obtained with the use of the double delta correlator, with the spread of the second pair of samples equaling 0.1 chip, was 30.135 m. Because the measurements were static with an almost constant SNR value, the differences in accuracy between the correlator with sample spacing of 0.1 and 0.2 were hardly visible and amounted to approx. 0.3 m. This is due to the fact that positioning errors result from the superposition of distance errors. We assume that the advantages resulting from applying the double delta correlator with the spread of the second pair of samples = 0.1 chip will be noticeable in real conditions, i.e. for different propagation paths and for different SNR values. However, confirmation of this assumptions requires that another measurement campaign be conducted. Furthermore, during the measurements, the phenomenon of tropospheric ducts could be observed. The double delta correlator showed an RMS that was increased even by 20 m compared with the measurements performed with the duct phenomenon not being present. The software implemented by the authors allowed to conduct an in-depth analysis. It included an application for the correlation of signals, determining pseudo-ranges, and software determining positions based on distance measurements from several R-Mode reference stations using the TOA method in LOS and NLOS environments.

7. Future Work

The selected optimal correlator will be used primarily in the next measurement campaign, the purpose of which will be to demonstrate the capability of the VDES R-Mode system and its usefulness for the PNT sector in marine conditions. The received data will be processed on the survey ship in real time and the highest achievable level of accuracy will be required. These tests will show the effectiveness of the selected double delta correlator in marine navigation applications. The data that has been already collected with the use of the Gdynia-Jastarnia setup will allow to check the effectiveness of the

double delta correlator, but also to evaluate the operation of the entire VDES R-Mode positioning system in the long term. It will also allow to assess the dependence of the calculated RMS on the various weather conditions. Long term plans assume that the time and reference clock will be synchronized with the use of optical fibers connecting the transmitter with a common central time standard. This will allow to separate the system from sources of time errors and will help compare the positioning errors with the results obtained in the course of previous campaigns, where time synchronization was based on rubidium oscillators. All of these factors will help improve the R-Mode test bench and eventually transform it into the most accurate positioning and navigation system that will provide reliable data, even in a scenario in which the GNSS is not available. These activities contribute to increasing the level of protection and safety in the Baltic Sea by improving the technical capabilities of the broadly understood maritime sector.

8. Acknowledgments


The R-Mode Baltic 2 (Ranging Mode for the Baltic Sea) project is cofounded from the Interreg Baltic Sea Region Programme 2014–2022 and the funds reserved for science in the years 2021–2022 granted for the purpose of a co-financed international project.

References

- [1] –, EfficienSea 2 project official website: <https://www.iala-aism.org/technical/e-nav-testbeds/efficienssea-2/>.
- [2] –, Ranging Mode for the Baltic Sea project official website: <http://r-mode-baltic.eu/>.
- [3] –, R-Mode Baltic 2 project official website: <https://projects.interreg-baltic.eu/projects/r-mode-baltic-2-253.html>.
- [4] K. Bronk, P. Koncicki, A. Lipka, R. Niski, and B. Wereszko, "Concept, signal design, and measurement studies of the R-Mode Baltic system", *Navigation*, vol. 68, no. 3 pp. 465-483 2021 (DOI: 10.1002/navi.443).
- [5] –, European Space Agency, Baseband Processing, (https://gssc.esa.int/navipedia/index.php/Baseband_Processing).
- [6] M. Irsigler, "Multipath Propagation, Migration and Monitoring in the Light of Galileo and the Modernized GPS", 2008 (<https://athene-forschung.unibw.de/doc/86276/86276.pdf>).
- [7] Q. Bo, L. Longlong, L. Bian, X. Wang, and M. Yansong, "An Unambiguous Multipath Mitigation Method Based on Double-Delta Correlator for BOC Modulation Signal", *China Satellite Navigation Conference (CSNC) 2016 Proceedings*, pp. 7–8, 2016 (DOI: 10.1007/978-981-10-0934-149).
- [8] J. Lal-Jadziak, "Pomiar szumu w szumie metodą korelacyjną (Noise in noise measurement by means of correlation method)", *Przegląd Elektrotechniczny*, vol. 92, no. 11, pp. 179–182, 2016 (ISSN 0033-2097, DOI: 10.15199/48.2016.11.44) [in Polish].
- [9] –, European Space Agency, Multicorrelator, (<https://gssc.esa.int/navipedia/index.php/Multicorrelator>).
- [10] V. Dierendonck, P. Fenton, and T. Ford, "Theory and Performance of Narrow Correlator Spacing in a GPS Receiver", *Navigation*, 1992.
- [11] K. Benachenhou, E. Sari, and Hammadouche, "Multipath Mitigation in GPS/Galileo Receivers with Different Signal Processing Techniques", *SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications*, 2009 (<http://www.setit.rnu.tn/CDs%20SETIT/SETIT%202009/Telecom%20and%20Network/216.pdf>).
- [12] M. Tamazin, A. Noureldin, M. Korenberg, and A. Kamel, "A New High-Resolution GPS Multipath Mitigation Technique Using Fast Orthogonal Search", *The Journal of Navigation*, pp. 794–814, 2016 (DOI: 10.1017/S0373463315001022).
- [13] ITU-R Recommendation M.2092-0. Technical characteristics for a VHF data exchange system in the VHF maritime mobile band, 2015 (<https://www.itu.int/rec/R-REC-M.2092>).
- [14] ITU-R Recommendation M.1371-5. Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band, 2014 (<https://www.itu.int/rec/R-REC-M.1371-5-201402-I/en>).
- [15] K. Bronk, P. Koncicki, A. Lipka, D. Rutkowski, and B. Wereszko, "Simulation and measurement studies of the VDES system's terrestrial component", *Polish Maritime Research*. 2019; 1(101), vol. 26, pp. 95–106, 2019 (DOI: 10.2478/pomr-2019-0011).
- [16] D. Egea-Roca, "Change Detection Techniques for GNSS Signal-Level Integrity", *Department of Telecommunications and Systems Engineering*, 2017 (<http://www.tesisenred.net/bitstream/handle/10803/458425/der1de1.pdf>).
- [17] A. Broumandan, G. Lachapelle, and J. Nielsen, "TOA Estimation Enhancement based on Blind Calibration on Synthetic Arrays", *2008 IEEE 68th Vehicular Technology Conference*, 2008 (DOI: 10.1109/VETEFC.2008.86).
- [18] L. Rogers, "Likelihood estimation of tropospheric duct parameters from horizontal propagation measurements", *Radio Science*, vol. 32, no. 1, pp. 79–92, 1997 (DOI: 10.1029/96RS02904).
- [19] –, ITU-R Recommendation P.834-7. Effects of tropospheric refraction on radiowave propagation, 2015 (<https://www.itu.int/rec/R-REC-P.834-7-201510-S/en>).
- [20] K. Bronk, M. Januszewska, P. Koncicki, R. Niski, and B. Wereszko, "The concept of the R-Mode Baltic System using AIS Base Stations", *Przegląd Telekomunikacyjny – Wiadomości Telekomunikacyjne*. vol. 6, pp. 257–260, 2019 DOI: 10.15199/59.2019.6.27.
- [21] –, IALA Guideline G1158. The Technical Specification of VDES, Edition 1.0, 2020 (<https://www.iala-aism.org/product/g1158/>).
- [22] A. Kupper, "Location-Based Services: Fundamentals and Operation", *Wiley*, 2005 (ISBN: 9780470092316).
- [23] P. Groves, "Principles of GNSS Inertial, and Multisensor Integrated Navigation Systems", Artech House, 2008 (ISBN: 9781580532556).
- [24] T. Pany, M. Irsigler, and B. Eissfeller, "S-Curve Shaping: A New Method For Optimum Discriminator Based Code Multipath Mitigation", *Institute of Geodesy and Navigation*, pp. 2139–2154, 2005 (<https://www.researchgate.net/publication/238623083>).



Krzysztof Bronk Ph.D. (2010), is an Assistant Professor in National Institute of Telecommunications. He is an author or co-author of more than 40 reviewed scientific articles and publications and about 20 R&D technical documents and studies. His research is mainly centered on the field of radio-communication systems and networks designing and planning, software defined and cognitive radio systems development, multi-antenna technology, cryptography, propagation analysis, transmission and coding techniques as well as positioning systems and techniques. His interests include also multithread and object-oriented applications, devices controlling applications, DSP algorithms and quality measurement solutions.

 <https://orcid.org/0000-0002-3594-8462>

E-mail: K.Bronk@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland



Magdalena Januszewska graduated Electronics and Telecommunications at the Faculty of Electronics, Telecommunications and Informatics of the Gdańsk University of Technology. She started working at the Institute of Telecommunications after completing student internship there. During last years she participated in the R-Mode Baltic and R-Mode Baltic 2

projects. She was involved in the creation and development of a navigation simulator for studying the impact of the distance measurement error on the positioning error in sea conditions. She specializes in mathematical modeling and high-level programming.

<https://orcid.org/0000-0002-4802-6379>

E-mail: M.Januszewska@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland



Patryk Koncicki graduated from the Faculty of Electronics, Telecommunications and Informatics of the Gdańsk University of Technology, receiving an M.Sc. in radio communications. He is professionally involved in the subject of maritime and satellite navigation systems. He deals with it on a daily basis implementation

of advanced digital signal processing algorithms. Develops your skills programming with particular emphasis on the C++ language.

<https://orcid.org/0000-0003-2618-1594>

E-mail: P.Koncicki@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland



Adam Lipka received the M.Sc. and Ph.D. degrees in Telecommunication from Gdansk University of Technology in October 2005 and June 2013, respectively. Since January 2006, he has been working in the National Institute of Telecommunications in its Wireless Systems and Networks Department in Gdansk (currently as an Assistant Professor).

His scientific interests include contemporary transmis-

sion techniques, MIMO systems and radio waves propagation. He is an author or co-author of over 50 scientific papers and publications

<https://orcid.org/0000-0002-2919-4270>

E-mail: A.Lipka@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland



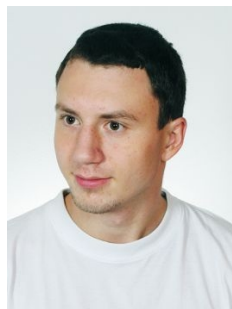
Rafał Niski in 2001 graduated from Gdansk University of Technology and received M.Sc. in the field of radiocommunications. Since that time he has been working in the National Institute of Telecommunications in Gdansk, firstly as an Assistant Professor, and after receiving the Ph.D. degree in 2006 as an Associate Professor. From 2005 till 2012 he was

the Head of the Wireless Systems and Networks Department and since 2016 he is the Head of Network and Equipment Measurement Section. His scientific research concerns the theory and techniques of mobile communication, radio networks design and planning and measurements of transmission and quality parameters in radio networks. He is the author or co-author of nearly 100 scientific publications. From 2007 till 2021 was a member of Scientific Council of the National Institute of Telecommunications.

<https://orcid.org/0000-0002-5106-9046>

E-mail: R.Niski@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland



Błażej Wereszko received the M.Sc. in Electronics and Telecommunications from Gdańsk University of Technology in 2011. Since 2010 he has been working in the Wireless Systems and Networks Department of the National Institute of Telecommunications. His scientific interests focus on wireless communications, radio waves propagation, radiolocation

techniques and satellite communications. He is an author or co-author of over 20 scientific papers and publications in the field of radiocommunication.

<https://orcid.org/0000-0001-7474-692X>

E-mail: B.Wereszko@il-pib.pl

National Institute of Telecommunications, Warsaw, Poland

Analysis of an LSTM-based NOMA Detector Over Time Selective Nakagami- m Fading Channel Conditions

Ravi Shankar¹, Jyoti L. Bangare², Ajay Kumar³, Sandeep Gupta⁴, Haider Mehraj⁵,
and Shriram S. Kulkarni⁶

¹Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India,

²Cummins College of Engineering for Women, Pune, India,

³Department of Computer Science and Engineering, Jecrc University, Jaipur, India,

⁴Electrical & Electronics Engineering Department, Eklavya University, Sagar Road, Damoh, India,

⁵Department of Electronics and Communication Engineering, Baba Ghulam Shah Badshah University, Rajouri, J&K, India,

⁶Department of Information Technology, Sinhgad Academy of Engineering, Pune, India

<https://doi.org/10.26636/jtit.2022.161222>

Abstract — This work examines the efficacy of deep learning (DL) based non-orthogonal multiple access (NOMA) receivers in vehicular communications (VC). Analytical formulations for the outage probability (OP), symbol error rate (SER), and ergodic sum rate for the researched vehicle networks are established using i.i.d. Nakagami- m fading links. Standard receivers, such as least square (LS) and minimum mean square error (MMSE), are outperformed by the stacked long-short term memory (S-LSTM) based DL-NOMA receiver. Under real time propagation circumstances, including the cyclic prefix (CP) and clipping distortion, the simulation curves compare the performance of MMSE and LS receivers with that of the DL-NOMA receiver. According to numerical statistics, NOMA outperforms conventional orthogonal multiple access (OMA) by roughly 20% and has a high sum rate when considering i.i.d. fading links.

Keywords — deep learning (DL), multiple-input multiple-output (MIMO), non orthogonal multiple access (NOMA), orthogonal multiple access (OMA).

1. Introduction

Nowadays, vehicles are capable of exchanging, in real time, data about their speed, position, and driving directions using vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communications [1]. Vehicles may now also receive notifications from many directions thanks to the technology supporting V2I communication, giving them a clear 360° picture of every other car in their surroundings [2], so that they are able to identify potential threats. The V2V device then alerts drivers via tactile, audible, or visual alarms, [3]–[4] (Fig. 1). The main drivers of VC applications are multimedia and safety. While traffic management and multimedia applications require increased energy efficiency (EE), spectrum efficiency (SE), and high connectivity in V2I and V2V wireless communications, safety messages need an exceptional end-to-end dependability and exceptionally low latency [5]. Unfortunately, existing VC technologies, such as wireless access in

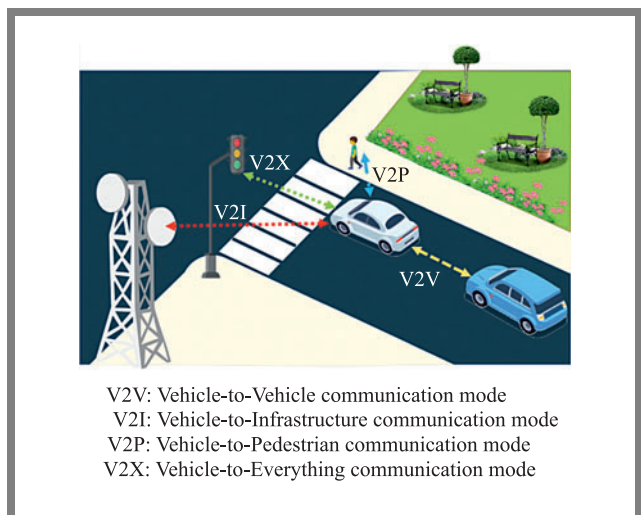


Fig. 1. Schematic representation of V2I and V2V networks.

vehicular environments, 4G, and LTE-A, are based on orthogonal frequency division multiple access (OFDMA) and are unable of providing the high SE and end-to-end reliability rates required for enhancing VC.

Non-orthogonal multiple access (NOMA) systems have gained a lot of interest in recent years due to the advancement of 5G cellular networks [6]–[8]. The high throughput of NOMA, allowing it to serve large numbers of users utilizing the same time and frequency resources, is the major rationale for its adoption in 5G [9]. NOMA approaches are divided into two categories: power-domain and code-domain [10]–[11]. In the power domain variety, NOMA accomplishes multiplexing, but in the code domain, NOMA achieves multiplexing. The focus of this paper is on the power-domain NOMA which will be hereinafter referred to simply as NOMA.

NOMA is an approach that is considered of being capable of meeting data rates and user access needs associated with multimedia applications and the Internet of Things (IoT). NOMA

is a viable approach for meeting 5G wireless communications objectives, such as high SE, extremely low latency, and massive connectivity. It has been often utilized in conjunction with the MIMO technique, relaying communications, cognitive cooperative systems, millimeter-wave communications, and other technologies to maximize sum-rate and user fairness under fading channel conditions (Fig. 2).

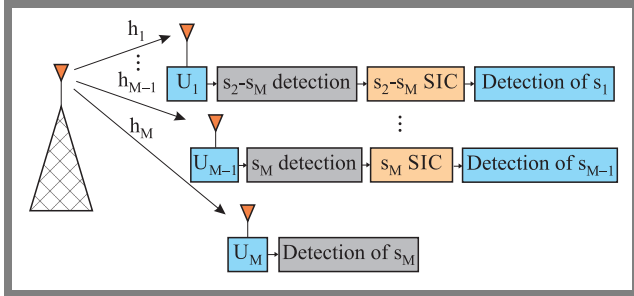


Fig. 2. Downlink for multiple user NOMA for with different fading channels conditions.

The rollout of 5G is associated with new features and technologies allowing operators to take advantage of new infrastructure capabilities. Artificial intelligence/machine learning (AI/ML), a prospect approach for developing adaptive and predictive systems, has evolved in both vehicles and traditional wireless networks. ML can handle highly dynamic vehicular network challenges that traditional solutions, such as classical control loop design and optimization techniques, cannot cope with by relying on data-centric methodologies [12]–[13].

V2V and V2X connectivity are the next paradigms in connected vehicle research. Existing V2X concepts, rely on classical OMA, which employs orthogonal resources. This makes it difficult to deploy NOMA, since its performance is strongly dependent on a large channel gain differential existing between users. As a result, OMA-based V2X may not be able to satisfy V2X criteria in high-traffic areas. NOMA provides multiplexing in the power domain to serve several users at the same time or to share frequency resources, thus offering a considerable increase in SE over OMA [14]–[15].

2. Related Work

The SER and OP performance of cooperative NOMA was examined in [16] and the findings were compared with non-cooperative NOMA in terms of data throughput, OP, and diversity gain, considering i.i.d. Nakagami- m fading links.

In [17], the authors investigated a DL-aided NOMA system and presented the applications of DL in other wireless technologies. The authors employed the recurrent neural network (RNN) algorithm for identifying fading channel coefficients. In paper [18], the authors investigated an LSTM NOMA receiver under the frequency flat Rayleigh fading channel scenario. The LSTM algorithm was employed for obtaining the optimal receiver. In article [19], the authors investigated a ubiquitous bidirectional LSTM-based NOMA receiver under the imperfect successive interference cancellation (SIC) scenario. Simulation results demonstrated that the

DL-based NOMA receiver performs better than the traditional SIC MIMO-NOMA techniques. However, the authors of papers [14]–[19] did not consider the time-varying channel or the node mobility scenario.

In paper [20], the authors investigated a multiple user NOMA system under frequency flat Rayleigh fading channel conditions. The NOMA approach was used by the BS to provide connectivity, user fairness, and a high SE for multiusers under time-selective fading channel conditions. In addition, at the BS, an optimal power allocation mechanism was used to share the available power by assigning a power allocation factor to each of the users.

The authors of [21] investigated channel capacity of a DL MIMO-NOMA system by considering different multi-antenna scenarios over generalized fading channels in the presence of perfect and imperfect SIC schemes. The authors looked at a broad architecture for numerous NOMA users using TAS-assisted Alamouti space-time codeword transmission. At the output of the maximum-ratio combiner of the NOMA users, accurate formulations of the probability density function of the TAS-OSTBC processed signal-to-noise ratio (SNR) were generated. The authors also looked at the impacts of power coefficients and fading factors on the error performance of TAS-OSTBC-assisted NOMA users.

The authors of [22] explored a NOMA VC network under time selective independent but not necessarily identically distributed (i.n.i.d.) Nakagami- m fading channel conditions. When a BS communicates with vehicles travelling away from the BS using single-input multiple-output technology (SIMO), diversity combining techniques, such as maximum ratio combining (MRC) and selection combining (SC) are used at the receiver of each vehicle to fusion the signals received at the antennas. Analytical formulas of the OP and ergodic sum rate are obtained in this context for the examined vehicle networks under the assumption of independent but not necessarily identically distributed (i.n.i.d.) Nakagami- m fading channels.

In this paper, we consider DL-based NOMA, assuming that the channel will become time-selective due to node mobility conditions. A performance comparison is provided between a conventional NOMA receiver and a S-LSTM based NOMA receiver for various shape parameters values and node mobility scenarios.

3. Signal and Channel Model

3.1. Time Selective Nakagami- m Fading Channel Model

Due to the presence of node mobility, the channel will become time selective in nature. The first order autoregressive process, written as in [23]–[24], is:

$$d(k) = \rho d(k-1) + \sqrt{1-\rho^2} e(k), \quad (1)$$

where k and $k-1$ denote the two neighboring time instants and may be used to construct the time selective channel model. The term $e(k)$ denotes a random process, modeled as $\text{CN}(0, \sigma^2)$.

ρ represents the correlation coefficients that develop as a result of the node's mobility and Doppler spread expressed as:

$$\rho = J_0 \frac{2\pi f_c v}{R_S c}$$

where f_c represents the carrier frequency of the radio wave, v is the relative velocity between two communicating cellular users, c denotes the speed of light, $J_0(\cdot)$ denotes the Bessel function of the zeroth order and first kind, and R_S represents the data transmission rate.

3.2. Signal Model

In our analysis, we have considered i.i.d. Nakagami- m time selective channel fading connections, with a fading severity parameter m and the average fading link gain of Ω_i , $i \in \{SD, SR, RD\}$. The channel is no longer frequency flat and due to the Doppler spread, it will become the frequency selective, causing inter symbol interference (ISI). In order to mitigate the effect of ISI CP is used in the orthogonal frequency division multiplexing (OFDM) system. Channel impulse response length should be longer than CP length to obtain lower SER performance. Due to reflection, refraction, and scattering, the receiver receives numerous copies of the signal due to multipath propagation:

$$\left\{ \sum_{n=0}^{K-1} d(n) \right\}.$$

The signal received after the transmission of the OFDM symbol $s(n)$ is [12]–[18]:

$$r(n) = x(n) \otimes d(n) + \eta(n), \quad (2)$$

where $d(n)$ represents time selective i.i.d. Nakagami- m faded random samples, \otimes denotes the circular convolution $\eta(n)$ represents the channel noise with the expected value of 0 and standard deviation of $\sqrt{N_0/2}$, i.e. $\mathbf{CN}(0, N_0/2)$.

After performing the Fourier transform and removing the CP at the receiver, the resulting signal is [12]–[18]:

$$R(k) = X(k)D(k) + \tilde{N}(k), \quad (3)$$

where $R(k)$, $X(k)$, $D(k)$, and $\tilde{N}(k)$ are the discrete Fourier transform (DFT) of $r(n)$, $x(n)$, $d(n)$, and $\eta(n)$, respectively. In an uplink (UL) NOMA transmission, the composite signal at the BS is [12]–[18]:

$$R(k) = \sum_{t=1}^M \sqrt{P_t(n)} X_t(k) D_t(k) + \tilde{N}(k), \quad (4)$$

where $R(k)$ denotes the received signal corresponding to the transmission of $X_t(k)$ and $\tilde{N}(k)$ represents channel noise. $P_t(n)$ represents the power allocated to user t on the k -th subcarrier. For M subcarriers, the total power is expressed as P . The optimal power allocation factor is:

$$\beta_t(k) = \frac{P_t(k)}{P},$$

for user t .

The total available power is expressed as $\sum_{t=1}^M \beta_t(k) = 1$. The channel is essentially a multitap type due to multipath

propagation. Channel impulse response $d_t(n)$ for user t is:

$$d_t(n) = \sum_{l=1}^K d_{t,l} \delta(k - k_{t,l}),$$

where $d_{t,l}$ represents the complex channel gain and $k_{t,l}$ represents time delay of the l -th multipath. DFT of $d_t(n)$ is given as $d_t(k)$. The total number of resolved paths is equal to 50 and fading links are i.i.d. time selective Nakagami- m distributed.

4. DL-based NOMA Receiver

4.1. S-LSTM Basics

Numerous tasks that former learning algorithms for recurrent neural networks (RNNs) were not capable of accomplishing may be solved by LSTMs. In a 5G NOMA network, LSTM may be used for such tasks as channel estimation, SER computation, optimal power allocation, and OP calculation. Time-series forecasts may also be successfully made with LSTMs. Based on real-time wireless propagation data sets that are studied using different parameters, including the number of fading channel instances, the authors of [24] explore a LSTM network for fading channel coefficients of the DL NOMA system. Currently, S-LSTMs are a reliable method for resolving complex sequence prediction issues.

An S-LSTM architecture is an LSTM standard composed of numerous LSTM layers. The model becomes deeper as LSTM hidden layers are stacked, more appropriately qualifying the method as DL. A multilayer perceptron neural network may become deeper by including more hidden layers. It is known that the additional hidden layers integrate the learnt representation from the earlier layers to produce new representations with a high degree of abstraction, taking lines, forms, and things as examples. Instead of sending a single value, an LSTM layer located above transmits a set of values to another LSTM layer positioned below. One output time step is utilized for each input time step, rather than one output time step for all input time steps [9]. The primary distinction between LSTM and S-LSTM is that in a S-LSTM-based system, time slots are essentially sub-carriers, and after considering the single time step in the S-LSTM architecture, DL training may be performed by utilizing the multiple user identification method for a specific sub-carrier.

4.2. Model Training

OFDM data symbols have the form of packets, with a total of 84 carriers. An OFDM data packet consists of 4 symbols. For channel estimation, two pilots are assigned. Each OFDM symbol consists of 2 bits per subcarrier. Because we are dealing with complicated data symbols, the next step is to create a feature vector (FV).

At the training stage, the complex data symbol consists of both real and complex components. The dimension of the FV is determined by the number of features per sample. The FV has a size of $84 \times 4 \times 2 = 672$ for 84 sub-carriers. The S-LSTM NOMA channel estimator is trained to understand

the signal associated with the k -th subcarrier by incorporating the necessary label in the training. The label is a number that indicates the combination of both users' transmitted symbols. Because both users are transmitting quadrature amplitude modulation (QAM) or 4-phase shift keying (PSK) signals, there will be 20 combinations/labels. In Matlab software, deep neural networks (DNNs) are developed by connecting DL layers to the DL Toolbox. Users may construct DL models and track their development using this tool. The dimension of the real-valued FV, which is 672, governs the size of the input to the input layer. The S-LSTM layer has 250 hidden units, followed by a fully linked layer with an output size of 25-bits. The classification layer generates an estimated label to map both users' transmitted signals simultaneously, and the softmax layer applies the softmax function to the input.

5. Simulation Results

The suggested S-LSTM-based NOMA detector is trained using simulation data and its performance is compared to that of the classic SIC receiver method. The prior channel state information (CSI) increases SER performance, allowing the MMSE and LS techniques to estimate the fading channel coefficients, respectively. SER is obtained per sub-carrier for various SNR regimes. For both offline and online training stages, the channel is assumed to be time selective or fast fading to minimize the influence of ISI and Doppler spread.

To analyze even minor fading channel variations, each OFDM packet provides a noticeable random phase shift to the fading channel of each cellular user. For both cellular users, the target signal-to-interference noise ratio (SINR) is 16 dB. For optimal or maximum likelihood receivers, which are used to test the accuracy of S-LSTM-based receivers, the entire CSI scenario is considered. 520,000 OFDM samples and 250 epochs were used to train this algorithm. When employing some training pilots that, remove CP or encounter non-linear clipping noise, S-LSTM-based receivers are more accurate than standard receivers used in the simulation.

In the simulated scenario, there are 84 subcarriers and a 30-second long CP. There are 35 multipaths and the carrier frequency is 3 GHz. To support sophisticated 4PSK and QAM modulation, the maximum delay spread is set to 30 symbols.

5.1. Investigation of OP for Node Velocity and Shape Parameters

Simulation findings for NOMA-based 5G vehicle networks validate analytical formulations of OP and the average sum rate. A DL V2A environment is analyzed in which 3 users are travelling away from the BS at 55 km/h. With a transmission symbol rate of $R_s = 20$ Mbps and a carrier frequency of $f_c = 6$ GHz, the BS connects with user 1, user 2, and user 3. User 1 is farthest from the BS and has the poorest channel conditions. The channel state is inversely proportional to the distance according to:

$$H_{n,1} = \frac{H_{dn}}{\sqrt{1 + d_n^\varepsilon}}.$$

Therefore, user 2 is travelling via the best channel. $\varepsilon = 3$ is the path loss exponent. At $t = 1$, performance factors $\alpha_1(1)$, $\alpha_2(1)$ and $\alpha_3(1)$ for mobile users 1, 2, and 3 are 0.6, 0.27, and 0.13, respectively. The order of the power coefficients is altered at t -th time instant in accordance with the channel order of the mobile users at that time instant. The minimum detection rate for each mobile user is $R_t = 1$ bps/Hz, resulting in a threshold SNR $\psi_{th,1} = \psi_{th,2} = \psi_{th,3} = 1$ for NOMA mobile users, $\psi_{th} = 7$ is the SNR threshold for traditional OMA, which can be calculated from [9]:

$$\frac{1}{3} \sum_{n=1}^N \log_2(1 + \psi_{th,n}) = \frac{1}{3} \log_2(1 + \psi_{th}). \quad (5)$$

The time selective fading channel can be modelled using the autoregressive process with variance of $\sigma_{en}^2 = 0.01$ at the point in time of $t = 3$. For single input multiple output NOMA, the receiver at each vehicle uses optimal combining and zero forcing schemes, whereas for MIMO-NOMA, it uses singular value decomposition (SVD). The average SNR received at each link is separated using an exponential power decay profile since all diversity branches at each vehicle are

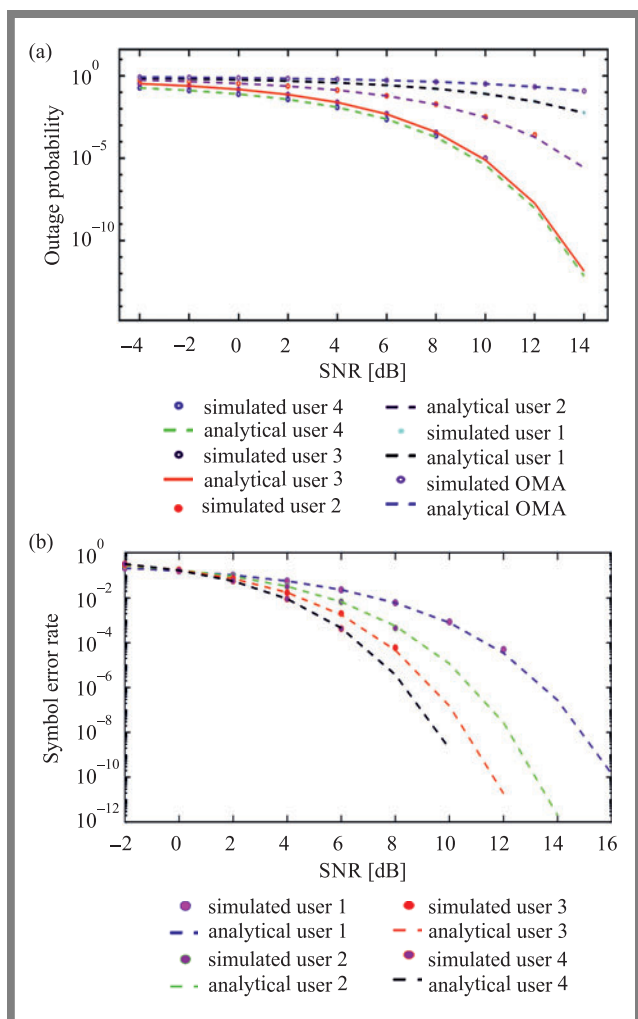


Fig. 3. OP vs. SNR for single SISO NOMA. (a) $m = 2$ and (b) $m = 3$.

i.i.d. We use the maximum received average SNR of $\Omega_l = 3$ and a fading factor of $\delta = 0.30$ in the simulations.

By assuming a perfect CSI, the outcomes of i.i.d. considerations are contrasted with those of the i.i.d. channel consideration. For $m = 2$, which represents the Rayleigh fading channel, Fig. 3a shows the outage performance of three NOMA vehicles and a standard OMA (non-line of sight condition). The findings reveal that, despite being allocated with the lowest power coefficient from the BS, the user with the best channel conditions (user 3) surpasses all three vehicles in terms of outage performance. Since they are provided with a higher power coefficient in NOMA than in OMA, the user with the poorest channel conditions (user 1) performs badly when compared to others. However, they outperform the classical OMA scheme.

The outage performance of the users with NOMA and OMA with $m = 2$ is shown in Fig. 3b. When compared to $m = 3$, performance is better, since the diversity benefit for $m = 2$ is bigger. For $m = 2$, user 1 of NOMA outperforms OMA by 2 dB. However, in the case of $m = 2$, as opposed to $m = 1$, performance decreases owing to i.i.d. considerations being greater. This indicates that in non-line of sight situations, the impact of i.i.d. considerations is reduced.

5.2. Effect of the Number of Pilots and Node Mobility

Both LS and MMSE techniques may yield reliable forecasts when 110 pilots are used, as illustrated in Fig. 4. Nevertheless, S-LSTM-based NOMA receivers are superior to other traditional NOMA receivers. A reduction in the number of pilots (to 30) for both user 1 and user 2 greatly reduces the decoding accuracy of LS and MMSE algorithms to SNR = 14 dB. The channels are time-selective, and it has been shown that as the communicating node's velocity increases, SER power decreases.

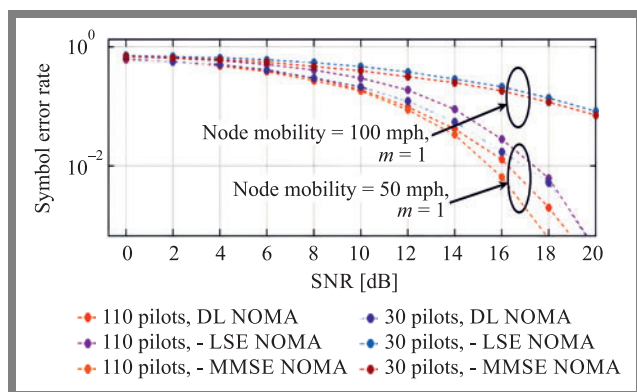


Fig. 4. SER vs. SNR of an S-LSTM-based DL NOMA receiver with 110 and 30 pilots over time selective Nakagami- m fading channel conditions.

In contrast, the DL NOMA receiver can achieve the performance of the 110 pilots example, demonstrating that S-LSTM-based receivers are more robust for several pilots and can achieve higher performance with fewer pilots.

5.3. Analysis of End-to-end System Performance

DL NOMA works considerably better when CP length is greater than impulse response. It has been discovered that neither LS nor MMSE receivers are capable of accurately estimating CSI. When exposed to severe ISI effects, even with excellent channel estimation, an optimum ML-based NOMA receiver can no longer offer the best response.

Time selective fading is used to test robustness of the DL NOMA receiver. SER performance of the DL NOMA receiver is comparable to that of an ideal ML-based NOMA receiver when the impact of node mobility is neglected. Additionally, as fading severity increases, SER performance improves. Furthermore, the DL NOMA receiver is resilient to the signal strength of the SLSTM-based DL NOMA receiver for user 2 (low channel gain user or far user), as shown in Fig. 5 and has a traditional error estimation effect. Furthermore, the DL NOMA receiver is resilient to the signal strength of the S-LSTM-based DL NOMA receiver in the case of user 2 (poor channel gain user or far user), as shown in Fig. 5, and propagates the estimated effect of flaws in the standard SIC scheme.

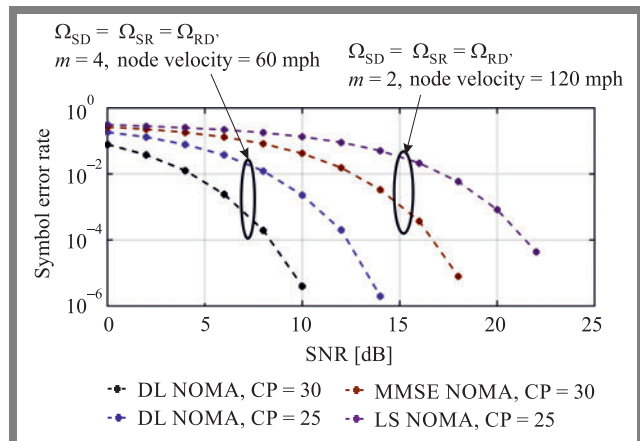


Fig. 5. SER vs. SNR for S-LSTM-based DL-NOMA receivers for various CP lengths under time selective Nakagami- m fading links.

The DL receiver has been shown to be resistant to random phase shifts and offers equal performance to its counterpart under ideal conditions, when used in a high mobility situation with a time varying channel. It has been proved through simulation that lower node velocity enhances end-to-end system performance.

5.4. SER Investigation Considering the Non-linear CN Problem

Due to the presence of the nonlinear noise results, higher backoff from peak output is required to maintain linearity in the power amplifier. Figure 6 shows the error performance of MMSE, and an S-LSTM-based NOMA receiver when the DNN receiver is facing non-linear noise, considering 4QAM complex modulated symbols. When the clipping ratio is equal the SER performance the DL NOMA receiver is much better than that of MMSE for SNR > 12 dB. The S-LSTM receiver outperforms the standard NOMA receiver, as shown in Fig. 7.

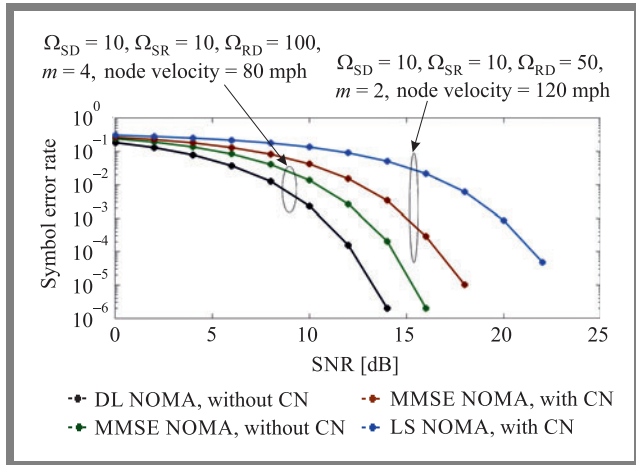


Fig. 6. SER vs. SNR of the S-LSTM-based NOMA system with and without CN for various node mobility and fading severities.

However, its detection performance varies depending on the node’s mobility situation.

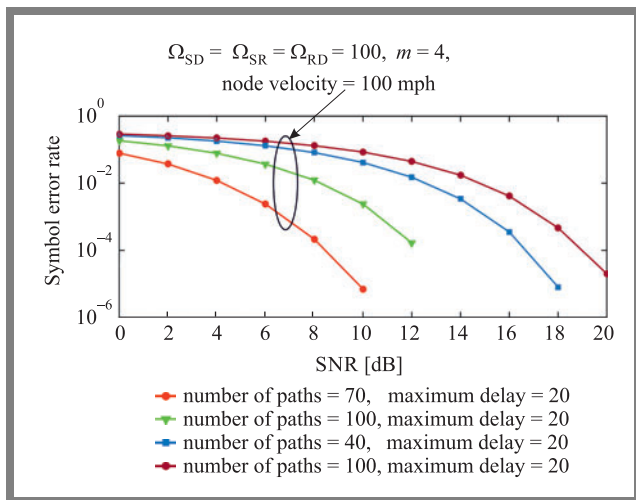


Fig. 7. Error probability vs. SNR considering gaps between testing and training phases.

5.5. Robustness Investigation over Time Selective Fading

In the online training step, CSI is calculated using data sets that are identical to those used in the offline training stage, and 4QAM complex modulated symbols are employed. The gap between online and offline deployments exists in real-time propagation situations. Furthermore, for the trained model to work, these differences must be stable. Figure 8 shows the effect of changing the fading relationship statistics used throughout the training and testing stages.

5.6. Effect of the LR on SER Performance

Here, the DL NOMA detector’s error probability performance is examined, and the error rate plots for the two mobile users are shown in Fig. 9 under time varying channel conditions. It has been observed that lower LRs yield lower SERs, implying that greater LRs will result in fast neural network weight up-

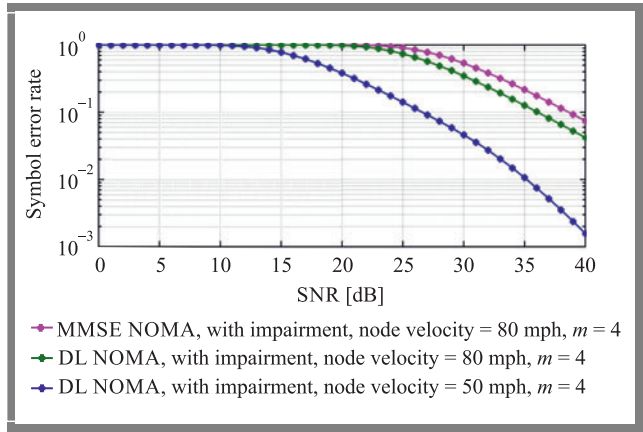


Fig. 8. SER vs. SNR under time selected Nakagami-*m* fading connections vs. considering all impairments.

dates and larger validation errors when using 4QAM complex modulated symbols.

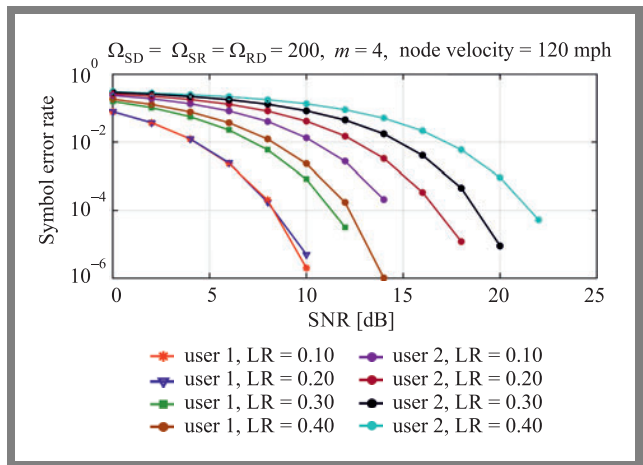


Fig. 9. SER graphs of the DL NOMA detector under the time selective Nakagami-*m* fading channel for various values of LR.

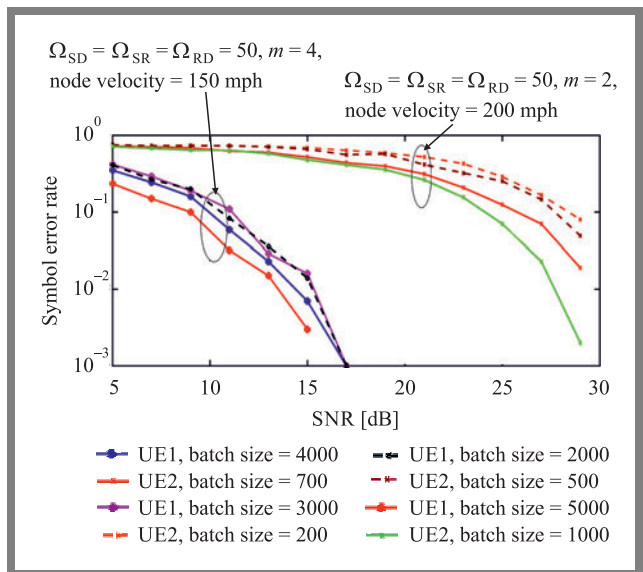


Fig. 10. SER plots of DL NOMA over time selective Nakagami-*m* fading channel settings trained with varying batch sizes.

5.7. Impact of Batch Size Considering Node Mobility Conditions

In this step, the training OFDM symbols are separated into packets, and iteration occurs throughout the training stage. The full dataset for this study takes 50 iterations to finish the epoch. Figure 10 depicts the effect of various batch sizes on DL system performance, demonstrating that bigger batches improve the SER. Small batches take much less time to converge compared with large batches in the training phase. Therefore, validation accuracy is the same. Smaller batches, on the other hand, result in less accurate testing.

6. Conclusion

Despite being assigned the lowest power coefficient by the BS, the users with the best channel conditions outperform all other users in terms of outage performance. Under time varying channel conditions, it has been observed that lower LR's yield the lower SERs, implying that greater LR's will result in fast neural network weight updates and larger validation errors when using complex modulated symbols. Larger batches need fewer iterations and DL fading channel coefficients change rapidly due to time selective fading, but each update uses more data to build a more accurate gradient estimate. Consequently larger batch sizes significantly improve spectral efficiency.

References

- [1] T. Xu, C. Xu, and Z. Xu, "An efficient three-factor privacy-preserving authentication and key agreement protocol for vehicular ad-hoc network", in *China Communications*, vol. 18, no. 12, pp. 315–331, 2021 (DOI: 10.23919/JCC.2021.12.020).
- [2] L.-L. Wang, J.-S. Gui, X.-H. Deng, F. Zeng, and Z.-F. Kuang, "Routing Algorithm Based on Vehicle Position Analysis for Internet of Vehicles", in *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11701–11712, 2020 (DOI: 10.1109/JIOT.2020.2999469).
- [3] F. Zhu, *et al.*, "Parallel Transportation Systems: Toward IoT-Enabled Smart Urban Traffic Control and Management", in *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4063–4071, 2020 (DOI: 10.1109/TITS.2019.2934991).
- [4] P. K. Singh, S. K. Nandi, and S. Nandi, "A tutorial survey on vehicular communication state of the art, and future research directions", *Vehicular Communications*, vol. 18, Article ID 100164, 2019 (ISSN 2214–2096, DOI: 10.1016/j.vehcom.2019.100164).
- [5] A. Kumar, S. Majhi, and H.-C. Wu, "Physical-Layer Security of Underlay MIMO-D2D Communications by Null Steering Method Over Nakagami- m and Norton Fading Channels", in *IEEE Transactions on Wireless Communications* (DOI: 10.1109/TWC.2022.3178758).
- [6] B.P. Chaudhary, R. Shankar, and R.K. Mishra. "A tutorial on cooperative non-orthogonal multiple access networks", *The Journal of Defense Modeling, and Simulation*, 2021 (DOI: 10.1177/1548512920986627).
- [7] L. Bhardwaj, R.K. Mishra, and R. Shankar, "Investigation of low-density parity check codes concatenated multi-user massive multiple-input multiple-output systems with imperfect channel state information", *The Journal of Defense Modeling, and Simulation*, vol. 19, no. 3, pp. 539–550, 2022 (DOI: 10.1177/1548512920968639).
- [8] M.K. Beuria, R. Shankar, and S. S. Singh, "Analysis of the energy harvesting non-orthogonal multiple access technique for defense applications over Rayleigh fading channel conditions", *The Journal of Defense Modeling, and Simulation*, 2021 (DOI: 10.1177/15485129211021168).
- [9] R. Tiwari and S. Deshmukh, "Prior information-based Bayesian MMSE estimation of velocity in HetNets", *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 81–84, 2018 (DOI: 10.1109/LWC.2018.2857805).
- [10] R. Tiwari and S. Deshmukh, "Analysis and design of an efficient handoff management strategy via velocity estimation in HetNets", *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 3, 2022 (DOI: 10.1002/ett.3642).
- [11] R. Tiwari and S. Deshmukh, "Handover count based MAP estimation of velocity with prior distribution approximated via NGSIM data-set", *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4352–4361, 2021 (DOI: 10.1109/TITS.2020.3043888).
- [12] S. Wong, *et al.*, "Traffic forecasting using vehicle-to-vehicle communication", *3rd Annual Conference on Learning for Dynamics and Control*, pp. 917–929, 2021 (<https://arxiv.org/pdf/2104.05528>).
- [13] C. Lin, Q. Chang, and X. Li, "A Deep Learning Approach for MIMO-NOMA Downlink Signal Detection", *Sensors*, vol. 19, p. 2526, 2019 (DOI: 10.3390/s19112526).
- [14] J. M. Kang, I. M. Kim, and C. J. Chun, "Deep Learning-Based MIMO-NOMA With Imperfect SIC Decoding", in *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020 (DOI: 10.1109/JSYST.2019.2937463).
- [15] R. Malladi, M. K. Beuria, R. Shankar, and S. S. Singh, "Investigation of the fifth generation non-orthogonal multiple access technique for defense applications using deep learning", *The Journal of Defense Modeling, and Simulation*, 2021 (DOI: 10.1177/15485129211022857).
- [16] X. Gong, X. Yue, and F. Liu, "Performance Analysis of Cooperative NOMA Networks with Imperfect CSI over Nakagami- m Fading Channels", *Sensors*, vol. 20, no. 2, p. 424, 2021 (DOI: 10.3390/s20020424).
- [17] Narengerile and J. Thompson, "Deep Learning for Signal Detection in Non-Orthogonal Multiple Access Wireless Systems", 2019 *UK/China Emerging Technologies (UCET)*, pp. 1–4, 2019 (DOI: 10.1109/UCET.2019.8881888).
- [18] R. Shankar, T. V. Ramana, P. Singh, S. Gupta, and H. Mehraj, "Examination of the Non-Orthogonal Multiple Access System Using Long Short Memory Based Deep Neural Network", *Journal of Mobile Multimedia*, vol. 18, no. 2, pp. 451–474, 2021 (DOI: 10.13052/jmm1550-4646.18214).
- [19] M. AbdelMoniem, S. M. Gasser, M. S. El-Mahallawy, M. W. Fakhri, and A. Soliman. 2019. "Enhanced NOMA System Using Adaptive Coding and Modulation Based on LSTM Neural Network Channel Estimation", *Applied Sciences*, vol. 9, no. 15, Article ID 3022, 2019 (DOI: 10.3390/app9153022).
- [20] M. A. Ahmed, A. Baz, and C. C. Tsimenidis, "Performance analysis of NOMA systems over Rayleigh fading channels with successive-interference cancellation", *IET Communications* 14, no. 6 pp. 1065–1072, 2020 (DOI: 10.1049/iet-com.2019.0504).
- [21] S. Mukhtar and G.R. Begh, "Error analysis of TAS-OSTBC assisted downlink NOMA system over generalized $\eta - \mu \eta - \mu$ fading Channel", *International Journal of Communication Systems*, e5234 (DOI: 10.1002/dac.5234).
- [22] D. K. Patel, *et al.*, "Performance Analysis of NOMA in Vehicular Communications Over i.n.i.d. Nakagami- m Fading Channels", in *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6254–6268, 2021 (DOI: 10.1109/TWC.2021.3073050).
- [23] A. Saxena Vehicle-to-Vehicle Communication: Let the car message while driving, not you! eInfochips, an Arrow company, (<https://www.einfochips.com/blog/vehicle-to-vehiclecommunication-let-the-car-message-while-driving-not-you/>).
- [24] S. Barmounakis, *et al.*, "LSTM-based QoS prediction for 5G-enabled Connected and Automated Mobility applications", *IEEE 4th 5G World Forum (5GWF)*, pp. 436–440, 2021 (DOI: 10.1109/5GWF52925.2021.00083).



Ravi Shankar received his B.E. in Electronics and Communication Engineering from Jiwaji University, Gwalior, India, in 2006. He received his M. Tech. degree in Electronics and Communication Engineering from GGSIPU, New Delhi, India, in 2012. He received a Ph.D. in Wireless Communication from the National Institute of Technology Patna, Patna, India, in 2019. From 2013 to 2014 he was an Assistant Professor at MRCE Faridabad, where he was engaged in researching wireless communication networks. His current research interests cover cooperative communication, D2D communication, IoT/M2M networks and networks protocols. He is a student member of IEEE.

 <https://orcid.org/0000-0001-7532-3275>

E-mail: ravishankar.nitp@gmail.com

Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India



Jyoti L. Bangare is working as an Assistant Professor at the Department of Computer Engineering at Cummins College of Engineering for Women, Pune, India. She received her B.Eng. in Computer Science from Savitribai Phule Pune University, Pune, Maharashtra, India. She obtained an M.Eng. Degree in Computer Science from Savitribai Phule Pune University, Pune, India. She is pursuing her Ph.D. in Computer Engineering at Savitribai Phule Pune University, Pune, India. Her main areas of interest includes machine learning, data analytics, artificial intelligence, and IoT.

Cummins College of Engineering for Women, Pune, India



Ajay Kumar received an M.Tech. (CSE) from Rajasthan Technical University, Kota in 2009 and B.Tech. (CSE) from Dr. B.R.A. University, Agra in 2001. He is currently working as an Assistant Professor at JECRC University and is pursuing a Ph.D (CSE) from JECRC University, Jaipur. His research interests include mobile communications, wireless networks and machine learning.

E-mail: ajay.kumar@jecrcu.edu.in

Department of Computer Science and Engineering, Jecrc University, Jaipur, India



Sandeep Gupta received his B.Tech. degree in Electrical and Electronics Engineering from UCER Niani, Allahabad, India in 2006. He has completed a Ph.D. in control & power systems. He is an Associate Professor at EKLAVYA University, Sagar Road, Damoh. His areas of interest cover application of artificial intelligence in power system control design, FACTS devices, renewable energy, power electronics and stability of power systems with machine learning.

 <https://orcid.org/0000-0002-3734-3723>

E-mail: jecсандеep@gmail.com

Electrical & Electronics Engineering Department, Eklavya University, Sagar Road, Damoh, India



Haider Mehraj received his B.Tech. in Electronics and Communication Engineering from the Guru Nanak Dev University, Amritsar, India in 2009 and MTech in Communication and Information Technology from National Institute of Technology, Srinagar, India in 2011. He is currently pursuing Ph.D. in Biometrics at the National Institute of Technology, Srinagar, India and working as Assistant Professor in BGSB University, Rajouri, India. He has several national and international publications to his credit. His research interests include Biometrics, Image Processing, Deep Learning, and Pattern Recognition.

 <https://orcid.org/0000-0002-2215-8373>

E-mail: haidermehraj@bgsbu.ac.in

Department of Electronics and Communication Engineering, Baba Ghulam Shah Badshah University, Rajouri, J&K, India



Shriram S. Kulkarni is working as an Associate Professor and HOD of the Department of Information Technology in Sinhgad Academy of Engineering, Pune, India. He completed his M.E. and Ph.D. in E&TC Engineering. His main areas of interest include wireless communication, machine learning, and IoT.

E-mail: sskulkarni.sae@sinhgad.edu

Department of Information Technology, Sinhgad Academy of Engineering, Pune, India

An Approximate Evaluation of BER Performance for Downlink GSVD-NOMA with Joint Maximum-likelihood Detector

Ngo Thanh Hai and Dang Le Khoa

Department of Telecommunications and Networks, University of Science, VNU-HCM, District 5, Ho Chi Minh City, Vietnam

<https://doi.org/10.26636/jtit.2022.160922>

Abstract — Generalized Singular Value Decomposition (GSVD) is the enabling linear precoding scheme for multiple-input multiple-output (MIMO) non-orthogonal multiple access (NOMA) systems. In this paper, we extend research concerning downlink MIMO-NOMA systems with GSVD to cover bit error rate (BER) performance and to derive an approximate evaluation of the average BER performance. Specifically, we deploy, at the base station, the well-known technique of joint-modulation to generate NOMA symbols and joint maximum-likelihood (ML) to recover the transmitted data at end user locations. Consequently, the joint ML detector offers almost the same performance, in terms of average BER as ideal successive interference cancellation. Next, we also investigate BER performance of other precoding schemes, such as zero-forcing, block diagonalization, and simultaneous triangularization, comparing them with GSVD. Furthermore, BER performance is verified in different configurations in relation to the number of antennas. In cases where the number of transmit antennas is greater than twice the number of receive antennas, average BER performance is superior.

Keywords — *generalized singular value decomposition (GSVD), joint maximum-likelihood, joint modulation, MIMO, non-orthogonal multiple access (NOMA)*

1. Introduction

Non-orthogonal multiple access (NOMA) has emerged as a promising technology for the next generation of wireless networks (5G and beyond). This is due to the fact that NOMA is capable of improving spectrum efficiency, providing better fairness, as well as reducing latency in serving users all those factors are necessary for intelligent and dynamic next generation wireless networks [1], [2]. In conventional orthogonal multiple access (OMA), multiple users are assigned to different radio resources, such as frequency and time, meaning that the number of users served is limited. However, NOMA can provide massive connections by simultaneously serving multiple users using the same spectrum resources [3], but this is done at the expense of increased intra-cell interference. To mitigate intra-cell interference, NOMA exploits successive interference cancellation (SIC) at receivers to detect desired signals [4]. Therefore, the key principle of NO-

MA is based on superposition coding (SC) at the transmitter and SIC at the receiver.

Recently, the combination of multiple-input multiple-output with NOMA (MIMO-NOMA) has received a lot of attention in wireless communication due to its high spectral efficiency. In [5], ergodic capacity maximization was studied for the Rayleigh fading channel in MIMO-NOMA with statistical channel state information at the transmitter. The authors of [6] have investigated problems affecting the downlink MIMO-NOMA system with regards to clustering, beamforming, and power allocation. Many works have shown that the performance of MIMO-NOMA is superior to that of MIMO-OMA. However, MIMO-NOMA with a precoder scheme was realized and offered potential performance gains [7]. The authors of [7] have proposed a signal alignment based framework with precoding that is not only general and applicable to both uplink and downlink MIMO-NOMA systems, but also achieves a significant performance gain compared to MIMO-NOMA without precoding.

Precoding schemes are usually classified into two categories: nonlinear precoding and linear precoding. Nonlinear precoding is commonly known as dirty paper precoding (DPC) [8], [9], which can reach the maximum capacity region of MIMO channels if the transmitters perfectly estimate channel state information. However, DPC is difficult to implement due to computational complexity of the detection process. In order to reduce decoding complexity on the user side, linear precoding is necessary. The key principle of precoding consist in transforming channel matrices into diagonal matrices in the process of zero-forcing (ZF)-based precoding, block diagonalization (BD)-based precoding [10], and generalized singular value decomposition (GSVD) [11]. All of the above methods are referred to as simultaneous diagonalization (SD). In addition, the channel matrices, after being detected at the users', may have the form of triangular matrices when simultaneous triangularization (ST)-based precoding [12] is applied by relying on QR decomposition. The authors of [12] have revealed that the performance of ST precoding is close to that of the upper bounds of DPC and outperforms SD precoding as GSVD in terms of total system capacity. As far as

antenna configurations are concerned, ZF precoding and BD precoding are valid only when the total number of receive antennas of all users is lower than that of transmit antennas at the base station. Furthermore, ST precoding is capable of achieving better total system capacity if the number of transmit antennas is greater than that of receive antennas of each user. GSVD precoding, meanwhile, may apply to all antenna configurations.

As mentioned above, GSVD is a simple tool for linear precoding schemes for MIMO-NOMA implementations. In essence, GSVD can be extended to a point-to-point MIMO channel, where singular value decomposition (SVD) is applied during the conversion process. In [13], the authors proposed a transmission protocol combining GSVD and NOMA and evaluated the system's performance based on the expected data rates. Here, the scheme was considered in the asymptotic regime and the number of transmit antennas and receive antennas approached the infinite value. Moreover, the authors came up with limiting the distribution of the squared generalized singular value of the two users' channel matrices. The authors of [13] continued to make important contributions regarding GSVD-NOMA by achieving some new results on the distribution of the squared generalized singular value, as shown in [14]. In this paper, we take advantage of the joint density probability function of GSVD singular values in [14] to derive the average BER performance. In [15], the GSVD-NOMA scheme has been considered with a channel estimation error. This research has proposed three models of uncertainty and realized power allocation to balance signal-to-interference-plus-noise ratio (SINR).

Distribution of the squared generalized singular value function presented in [14] is only applicable for average results computations. However, in some research schemes concerned with secure transmission analysis and channel power allocation, the marginal probability density function (PDF) is necessary. Hence, the authors of [16] have obtained the distribution characteristics of the ordered GSVD singular values. The theoretical analysis of GSVD-based security transmission has first been presented in [14], where performance of a GSVD-based MIMO-OMA system was investigated for secrecy outage probability. Focusing on security of transmission in GSVD-based MIMO-NOMA schemes, the authors of [17], [18] analyzed theoretical secrecy outage probability. The results they obtained revealed the superiority of GSVD-NOMA in terms of efficiency and security, compared to GSVD-OMA.

As far as BER performance of NOMA is concerned, a relatively small number of studies has been carried out. In [19], the exact closed-form BER expression of the QPSK constellation for an uplink NOMA system was expressed over an additive white Gaussian noise (AWGN) channel. In [20], an exact closed-form BER expression under SIC error for downlink NOMA over Rayleigh fading channels was derived. Besides, the authors have also derived one-degree integral form exact expression and closed-form approximate BER expression for uplink NOMA. Moreover, over the Nakagami- m flat fading channel, the exact BER of downlink NOMA systems with

SIC was derived for two and three user systems [21]. However, the performance of MIMO-NOMA, has been only studied in terms of overall system capacity and outage probability [5], [6]. The BER performance of the system has not been studied extensively. Recently, in [22], BER performance of an uplink NOMA was investigated with the use of the joint maximum-likelihood detector, where the base station was assumed to be equipped with N antennas. Apart from the SIC technique at the receivers, the authors in [23] came up with a technique to detect desired signals at the receivers, known as log-likelihood ratios (LLRs). For a downlink NOMA, the LLRs are characterized by almost the same error probability performance as ideal SIC probability.

In this paper, we consider a MIMO-NOMA system with GSVD, consisting of two users communicating with a base station (BS). The BS modulates the data of the two users using quadrature phase shift keying (QPSK) and superposes the said data by joint-modulation or multi-user superposition transmission case 2 (MUST-2) [24] to generate their respective NOMA symbols. For each user, we use the joint maximum-likelihood to recover data on each parallel GSVD-MIMO channel. The main contribution of this paper is that we derive the approximate expression of the average BER performance for the near user and the far user in downlink MIMO-NOMA systems with GSVD, as well as verify the correctness of the approximate expression obtained in the course of the Monte Carlo simulation. By relying on the approximate expression and simulation results, precoding schemes are compared with each other in order to choose the suitable precoding method for each antenna configuration. Moreover, by evaluating BER performance of GSVD, applicable antenna configurations are determined that may be designed.

The paper is organized as follows. Section 2 presents the system's model, the fundamental theory of GSVD, signal processing (MUST-2) at BS, and the joint maximum-likelihood (ML) detector to decode signals at the end users. In Section 3, we analyze numerical average BER performance, as well as derive its approximate closed-form expressions. In Section 4, works related to other precoding schemes, such as ZF, BD, and ST, are presented. In Section 5, numerical results are obtained to verify the precision of the analysis performed, the approximate expressions, and the simulation results. This section also shows the comparison with detection techniques and different precoding schemes. Finally, conclusion are presented in Section 6. Lemma and Theorem proofs are given in the Appendix.

2. System Model and Signal Processing

2.1. System Model

In this paper, we consider a MIMO-NOMA downlink system with one BS and two users: near user (NU) and far user (FU). The BS is equipped with N antennas and M antennas for each user (Fig. 1): \mathbf{H}_n and \mathbf{H}_f are $M \times N$ channel matrices from BS to NU and FU, respectively. Each element of the channel matrices is a mutually independent and identically distributed

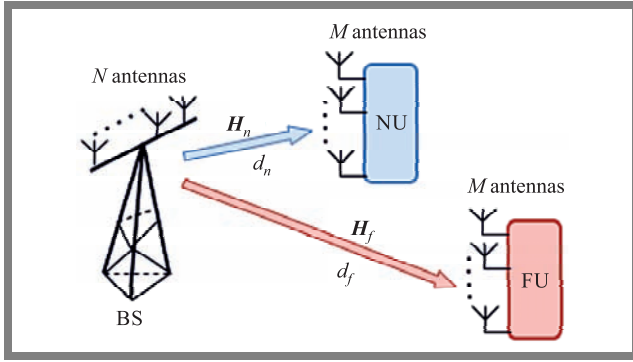


Fig. 1. A two-users MIMO downlink system model.

(i.i.d.) complex Gaussian random variable with zero mean and unit variance $\mathcal{CN}(0, 1)$. The user channel is assumed to be constant in terms of the transmission duration of one codeword and changes independently from one codeword to the next. As such, it is viewed as a quasi-static channel. Moreover, to apply GSVD to the linear precoding scheme, we assume that channel state information (CSI) is known fully at both the base station and the users. d_n and d_f denote distances between the base station and NU and FU, respectively. α is the path loss exponent.

Let $\mathbf{S} \in \mathbb{C}^{L \times 1}$ be the transmit signal vector with the length L . The transmit signal is precoded with the linear precoder matrix $\mathbf{V} \in \mathbb{C}^{N \times L}$. The precoded signal vector is used to transmit the result as:

$$\mathbf{S}_p = \frac{1}{t} \mathbf{V} \mathbf{S}, \quad (1)$$

where t denotes the power normalization factor. Assuming that the average transmit power at BS is P , the value of t is chosen that need, to satisfy the following condition:

$$P = \frac{1}{t^2} E [\text{trace} (\mathbf{V} \mathbf{S} \mathbf{S}^H \mathbf{V}^H)]. \quad (2)$$

At the near user and the far user, the received signal is presented, respectively, as:

$$\begin{aligned} \tilde{\mathbf{Y}}_n &= \frac{d_n^{-\frac{\alpha}{2}}}{t} \mathbf{H}_n \mathbf{V} \mathbf{S} + \mathbf{N}_n, \\ \tilde{\mathbf{Y}}_f &= \frac{d_f^{-\frac{\alpha}{2}}}{t} \mathbf{H}_f \mathbf{V} \mathbf{S} + \mathbf{N}_f, \end{aligned} \quad (3)$$

where $\mathbf{N}_j \sim \mathcal{CN}(0, N_0 \cdot \mathbf{I}_M)$, $j \in \{n, f\}$ is the additive white Gaussian noise (AWGN) vector and \mathbf{I}_M denotes the identity matrix of size M . Moreover, at each user signals $\tilde{\mathbf{Y}}_j$ are detected with the linear matrices $\mathbf{U}_j^H \in \mathbb{C}^{K \times M}$, leading to:

$$\begin{aligned} \mathbf{Y}_n &= \frac{d_n^{-\frac{\alpha}{2}}}{t} \mathbf{U}_n^H \mathbf{H}_n \mathbf{V} \mathbf{S} + \tilde{\mathbf{N}}_n, \\ \mathbf{Y}_f &= \frac{d_f^{-\frac{\alpha}{2}}}{t} \mathbf{U}_f^H \mathbf{H}_f \mathbf{V} \mathbf{S} + \tilde{\mathbf{N}}_f, \end{aligned} \quad (4)$$

where $\tilde{\mathbf{N}}_j$ denotes AWGN after the detection process. The choice of \mathbf{U}_j^H and \mathbf{V} needs to satisfy diagonalization or triangularization conditions. In this paper, we apply GSVD to diagonalization. Then the product of three matrices \mathbf{U}_j^H , \mathbf{H}_j and \mathbf{V} is the diagonal matrix \mathbf{D}_j .

2.2. GSVD and the Joint PDF of Squared Generalized Singular Values

GSVD is found in [25] under the assumption of the same number of columns in two channel matrices and is presented in more detail in [13], [16]. By applying GSVD, \mathbf{H}_n , \mathbf{H}_f are decomposed as follows:

$$\mathbf{H}_n = \mathbf{U}_n \mathbf{D}_n \mathbf{V}^{-1} \quad \text{and} \quad \mathbf{H}_f = \mathbf{U}_f \mathbf{D}_f \mathbf{V}^{-1}, \quad (5)$$

where \mathbf{U}_n , $\mathbf{U}_f \in \mathbb{C}^{M \times M}$ are two unitary matrices, $\mathbf{V} \in \mathbb{C}^{N \times N}$ is an invertible matrix. \mathbf{D}_n , $\mathbf{D}_f \in \mathbb{C}^{M \times N}$ are two non-negative diagonal matrices whose structure depends on the choices of M and N .

a) *The case when $M \geq N$.*

\mathbf{D}_n , \mathbf{D}_f are given by:

$$\mathbf{D}_n = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{O}_{(M-N) \times N} \end{bmatrix} \quad \text{and} \quad \mathbf{D}_f = \begin{bmatrix} \mathbf{O}_{(M-N) \times N} \\ \mathbf{S}_2 \end{bmatrix}, \quad (6)$$

where $\mathbf{O}_{(M-N) \times N}$ denotes the zero matrix of size $(M - N) \times N$, $\mathbf{S}_1 = \text{diag}(\alpha_1, \dots, \alpha_N)$, $\mathbf{S}_2 = \text{diag}(\beta_1, \dots, \beta_N)$ satisfying $1 \geq \alpha_1 \geq \dots \geq \alpha_N \geq 0$, $1 \geq \beta_N \geq \dots \geq \beta_1 \geq 0$ and $\alpha_i^2 + \beta_i^2 = 1$, $i = 1, \dots, N$. The generalized singular values are defined as α_i . Based on proposition 1.2 from [26], the unordered generalized singular values, squared, of the pair \mathbf{H}_n , \mathbf{H}_f ($X_i = \alpha_i^2$) follow the law of the beta-Jacobi ensemble. Moreover, by combining them with the introduction from [27], we achieve the following joint probability density function of $X_i \in [0, 1]$:

$$\begin{aligned} f_{X_1, \dots, X_N}(x_1, \dots, x_N) &= c_{J1} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \\ &\times \prod_{i=1}^N x_i^{M-N} (1 - x_i)^{M-N}, \end{aligned} \quad (7)$$

where $c_{J1} = \prod_{j=1}^N \frac{\Gamma(2M-N+j)}{\Gamma(1+j)[\Gamma(M-N+j)]^2}$. Let $Y_i = \beta_i^2$, due to $\beta_i^2 = 1 - \alpha_i^2$, so the joint probability density function of $Y_i \in [0, 1]$ can be:

$$\begin{aligned} f_{Y_1, \dots, Y_N}(y_1, \dots, y_N) &= c_{J1} \prod_{1 \leq i < j \leq N} (y_i - y_j)^2 \\ &\times \prod_{i=1}^N y_i^{M-N} (1 - y_i)^{M-N}. \end{aligned} \quad (8)$$

b) *The case when $M < N < 2M$.*

Put $q = 2M - N$ and $r = N - M$, \mathbf{D}_n and \mathbf{D}_f are written as follows:

$$\begin{aligned} \mathbf{D}_n &= \begin{bmatrix} \mathbf{I}_r & \mathbf{O}_{r \times q} & \mathbf{O}_{r \times r} \\ \mathbf{O}_{q \times r} & \mathbf{S}_1 & \mathbf{O}_{q \times r} \end{bmatrix}, \\ \mathbf{D}_f &= \begin{bmatrix} \mathbf{O}_{q \times r} & \mathbf{S}_2 & \mathbf{O}_{q \times r} \\ \mathbf{O}_{r \times r} & \mathbf{O}_{r \times q} & \mathbf{I}_r \end{bmatrix}, \end{aligned} \quad (9)$$

where $\mathbf{S}_1 = \text{diag}(\alpha_1, \dots, \alpha_q)$, and $\mathbf{S}_2 = \text{diag}(\beta_1, \dots, \beta_q)$, satisfying $1 \geq \alpha_1 \geq \dots \geq \alpha_q \geq 0$, $1 \geq \beta_q \geq \dots \geq \beta_1 \geq 0$ and $\alpha_i^2 + \beta_i^2 = 1$, $i = 1, \dots, q$.

Lemma 1. When $M < N < 2M$, the joint probability density function of the unordered generalized singular values, squared, of the pair $\mathbf{H}_n, \mathbf{H}_f \in \mathbb{C}^{M \times N}$ ($X_i = \alpha_i^2$) is given by:

$$f_{X_1, \dots, X_q}(x_1, \dots, x_q) = c_{J2} \prod_{1 \leq i < j \leq q} (x_i - x_j)^2 \times \prod_{i=1}^q x_i^r (1 - x_i)^r, \quad (10)$$

with $c_{J2} = \frac{1}{q!} \prod_{i=1}^q \frac{\Gamma(2M-i+1)}{\Gamma(q-i+1)[\Gamma(M-i+1)]^2}$.

Proof: see Appendix A.

The joint probability density function of $Y_i \in [0, 1]$ can be easily concluded as:

$$f_{Y_1, \dots, Y_q}(y_1, \dots, y_q) = c_{J2} \prod_{1 \leq i < j \leq q} (y_i - y_j)^2 \times \prod_{i=1}^q y_i^r (1 - y_i)^r, \quad (11)$$

c) **The case when $N \geq 2M$.**

\mathbf{D}_n and \mathbf{D}_f are expressed as follows:

$$\mathbf{D}_n = \begin{bmatrix} \mathbf{I}_M & \mathbf{O}_{M \times (N-M)} \end{bmatrix}, \quad (12)$$

$$\mathbf{D}_f = \begin{bmatrix} \mathbf{O}_{M \times (N-M)} & \mathbf{I}_M \end{bmatrix}.$$

The structure of \mathbf{D}_n and \mathbf{D}_f in Eq. (12) is completely independent of small-scale fading properties.

2.3. Modulation MUST-2 at BS

Clearly, from the GSVD diagonalization for two channel matrices, we get the length of the transmit signal vector $\mathbf{S} \in \mathbb{C}^{N \times 1}$. The precoding matrix and detection matrices, respectively, are $\mathbf{V} \in \mathbb{C}^{N \times N}$ and $\mathbf{U}_j^H \in \mathbb{C}^{M \times M}$. The received vectors \mathbf{Y}_n and \mathbf{Y}_f at NU and FU are expressed as:

$$\text{NU: } \mathbf{Y}_n = \frac{d_n^{-\frac{\alpha}{2}}}{t} \mathbf{D}_n \mathbf{S} + \tilde{\mathbf{N}}_n, \quad (13)$$

$$\text{FU: } \mathbf{Y}_f = \frac{d_f^{-\frac{\alpha}{2}}}{t} \mathbf{D}_f \mathbf{S} + \tilde{\mathbf{N}}_f,$$

where $\tilde{\mathbf{N}}_j = \mathbf{U}_j^H \mathbf{N}_j$. Due to the fact that \mathbf{U}_j is a unitary matrix, $\tilde{\mathbf{N}}_j \sim \mathcal{CN}(0, N_0 \cdot \mathbf{I}_M)$.

At the BS, we consider three types of symbols. The first type is the QPSK symbol of NU's signal denoted as s_i^n , $E(|s_i^n|^2) = 1$. Next, s_i^f is the QPSK symbol of FU's signal, $E(|s_i^f|^2) = 1$. Finally, the NOMA symbol for the two users is denoted as s_i . The NOMA symbol has a generic form of:

$$s_i = \sqrt{\phi P} s_i^n + \sqrt{\theta P} s_i^f, \quad (14)$$

where ϕ, θ are the power allocation coefficients satisfying $\phi + \theta = 1$, $\phi < \theta$, for efficient SIC at NU. The modulation in Eq. (14) is referred to as multi-user superposition transmission case 1 (MUST-1) [24]. Due to independent modulation in conventional NOMA, the constellation of s_i does not follow the Gray mapping rule. Therefore, we modulate NOMA

symbols using MUST-2 or joint-modulation, which means that bits from different users are mapped to one symbol taking into account the allocated power and the number of bits of each user. In this paper, we use a 16-QAM Gray-mapped constellation for joint mapping since 2 bits are assigned for NU and 2 bits are assigned for FU. The allocated power for NU and FU are respectively ϕP and θP , we have (Fig. 2):

$$d_1 = \frac{\sqrt{\theta P} - \sqrt{\phi P}}{\sqrt{2}} \quad \text{and} \quad d_2 = \frac{\sqrt{\theta P} + \sqrt{\phi P}}{\sqrt{2}},$$

It turns out that the constellation of MUST-2 is generated by permuting the position of points in the MUST-1 constellation satisfying the Gray mapping rule. Therefore, if users have different modulation orders according to their constellation's IQ, MUST-2 is valid for modulation at BS. Consider, for instance, a joint symbol at BS that has a single bit for FU and two bits for NU. Then, their constellation IQ will be the 8-QAM mapped Gray rule, with the positions of points arranged based on the power allocated to each user.

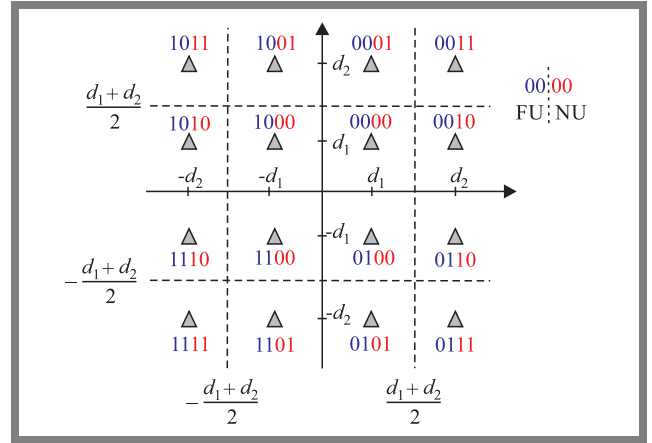


Fig. 2. Constellation of MUST-2 with 2 bits for NU and 2 bits for FU.

Based on the structure of $\mathbf{D}_n, \mathbf{D}_f$ and the received vector in Eq. (13), we formulate the forms of the transmitted vector \mathbf{S} at BS as:

a) **The case when $M \geq N$.**

The structure of \mathbf{S} is expressed as:

$$\mathbf{S} = (s_1, s_2, \dots, s_N)^T, \quad (15)$$

where \mathbf{S} comprises N NOMA symbols $s_i, i = 1, \dots, N$. The received symbol decomposed into parallel channels is written as:

$$y_i^n = \frac{d_n^{-\frac{\alpha}{2}}}{t_1} \alpha_i s_i + \tilde{n}_i^n, \quad (16)$$

$$y_i^f = \frac{d_f^{-\frac{\alpha}{2}}}{t_1} \beta_i s_i + \tilde{n}_i^f.$$

The power normalization factor is given by:

$$t = t_1 = \sqrt{\frac{N}{2M-N}} \quad [13] \quad \text{and} \quad \tilde{n}_i^n, \tilde{n}_i^f \sim \mathcal{CN}(0, N_0).$$

b) **The case when $M < N < 2M$.**

Assume that BS transmits the symbols vector \mathbf{S} with N NOMA symbols. At the receiver, FU receives only the first M symbols, whereas, NU only gets the last M symbols.

So, we formulate a structure of symbols \mathcal{S} transmitted at BS as:

$$\mathcal{S} = \left(\underbrace{\sqrt{P}s_1^n, \dots, \sqrt{P}s_r^n}_{r \text{ NU's symbols}}, \underbrace{s_1, \dots, s_q}_{q \text{ NOMA symbols}}, \underbrace{\sqrt{P}s_1^f, \dots, \sqrt{P}s_r^f}_{r \text{ FU's symbols}} \right)^T \quad (17)$$

The signals received at NU and FU are represented as:

$$\begin{aligned} y_i^n &= \frac{d_n^{-\frac{\alpha}{2}}}{t_2} \sqrt{P}s_i^n + \tilde{n}_i^n, \quad i = 1, \dots, r, \\ y_i^n &= \frac{d_n^{-\frac{\alpha}{2}}}{t_2} \alpha_i s_i + \tilde{n}_i^n, \quad i = r+1, \dots, M, \\ y_i^f &= \frac{d_f^{-\frac{\alpha}{2}}}{t_2} \beta_i s_i + \tilde{n}_i^f, \quad i = 1, \dots, q, \\ y_i^f &= \frac{d_f^{-\frac{\alpha}{2}}}{t_2} \sqrt{P}s_i^f + \tilde{n}_i^f, \quad i = q+1, \dots, M, \end{aligned} \quad (18)$$

where: $t_2 = t_1$ [13],

$$\begin{aligned} \alpha_i |_{i=r+1, \dots, M} &= \alpha_i |_{i=1, \dots, q}, \\ s_i |_{i=r+1, \dots, M} &= s_i |_{i=1, \dots, q}, \\ s_i^f |_{i=q+1, \dots, M} &= s_i^f |_{i=1, \dots, r}. \end{aligned}$$

c) The case when $N > 2M$.

The form of the symbol's vector at transmitted at BS is represented by:

$$\mathcal{S} = \left(\underbrace{\sqrt{P}s_1^n, \dots, \sqrt{P}s_M^n}_{M \text{ NU's symbols}}, \underbrace{0, \dots, 0}_{N-2M \text{ symbols } 0}, \underbrace{\sqrt{P}s_1^f, \dots, \sqrt{P}s_M^f}_{M \text{ FU's symbols}} \right)^T \quad (19)$$

NU and FU obtain the received symbols as:

$$\begin{aligned} y_i^n &= \frac{d_n^{-\frac{\alpha}{2}}}{t_3} \sqrt{P}s_i^n + \tilde{n}_i^n, \\ y_i^f &= \frac{d_f^{-\frac{\alpha}{2}}}{t_3} \sqrt{P}s_i^f + \tilde{n}_i^f, \quad i = 1, \dots, M, \end{aligned} \quad (20)$$

where $t_3 = \sqrt{\frac{2M}{N-2M}}$ [13].

2.4. Joint Maximum-likelihood Detector at NU and FU

As far as the QPSK symbols are concerned, users demodulate them easily by means of the maximum likelihood decision [28] on the parallel channels for $M < N < 2M$ and $2M < N$ scenarios. However, with NOMA symbols, we apply the joint maximum-likelihood (ML) detector to the estimation of NU signals and FU signals. This approach is mentioned in [22] to analyze BER performance of the uplink NOMA system with multiple receive antennas over the Rayleigh fading channel. This means that each user estimates firstly the joint symbols (NOMA symbols) on the 16-QAM

constellation and, after that, based on their correct-order bits, obtains their own symbols. The detection of joint symbols is:

$$r_i^* = \arg \min_{r_i \in \mathcal{X}} |z_i^j - r_i|^2, \quad (21)$$

where:

$$z_i^n = \frac{t_1}{\alpha_i d_n^{-\frac{\alpha}{2}}} y_i^n, \quad z_i^f = \frac{t_1}{\beta_i d_f^{-\frac{\alpha}{2}}} y_i^f$$

and $j \in \{n, f\}$ on the NOMA symbol channels. \mathcal{X} is a set of the constellation point coordinates. The users use r_i^* for bit mapping and obtain decoded bits for NU and FU. Here, the first two bits of a joint symbol correspond to FU, and the remaining two bits are represented as two bits of NU.

3. BER Performance Analysis

In this section, we derive the approximate expression of the average BER performance of NU and FU.

Let us define the generic form of the constellation point as $\overline{b_1 b_2 b_3 b_4}$, where b_1, b_2 are two bits of FU corresponding to the blue bits in Fig. 2 and b_3, b_4 represent two red bits shown in this figure, being the two bits of NU. First, we investigate BER performance in one codeword. After that, the average BER performance is calculated for the overall fading domain.

3.1. BER of NU

a) The case when $M \geq N$.

NU receives signals on N parallel NOMA symbol channels, so the average BER is:

$$P_{n1} = \frac{1}{N} \sum_{i=1}^N P_i^{n1}, \quad (22)$$

where P_i^{n1} is the average BER in the i -th parallel channel. P_i^{n1} is represented by the error probability for bit b_3 as P_{b3} and the error probability for bit b_4 as P_{b4} in the form of $\overline{b_1 b_2 b_3 b_4}$ as follows:

$$P_i^{n1} = \frac{1}{2} (P_{b3} + P_{b4}). \quad (23)$$

Bit $b_3 = 1$ when the real part of the transmitted symbol s_{iI} equals either $-d_2$ or d_2 and $b_3 = 0$ implies that $s_{iI} = -d_1$ or $s_{iI} = d_1$. Then, P_{b3} is:

$$P_{b3} = \frac{1}{4} (P_{b3|s_{iI}=-d_2} + P_{b3|s_{iI}=-d_1} + P_{b3|s_{iI}=d_1} + P_{b3|s_{iI}=d_2}), \quad (24)$$

where $P_{b3|s_{iI}=x}$ is the error probability for bit b_3 when the real part of the transmitted symbol assumes the value of x . From Eq. (21), z_i^n is written as: $z_i^n = s_i + w_i^n$ and $w_i^n = \frac{t_1}{\alpha_i d_n^{-\frac{\alpha}{2}}} \tilde{n}_i^n$. By investigating the constellation in Fig. 2, $P_{b3|s_{iI}=-d_2}$ can be defined by:

$$\begin{aligned} P_{b3|s_{iI}=-d_2} &= \Pr \left(-\frac{d_1 + d_2}{2} < -d_2 + w_{iI}^n < \frac{d_1 + d_2}{2} \right) \\ &= \Pr \left(-\frac{d_1 - d_2}{2} < w_{iI}^n < \frac{d_1 + 3d_2}{2} \right), \end{aligned} \quad (25)$$

where w_{iI}^n is the real part of w_i^n and $w_{iI}^n \sim \mathcal{N}\left(0, \frac{t_1^2 N_0}{2\alpha_i^2 d_n^{-\alpha}}\right)$. Putting $\rho = \frac{P}{N_0}$, $a_1 = d_n^{-\alpha} \phi$, $a_2 = d_n^{-\alpha} (2\sqrt{\theta} - \sqrt{\phi})^2$ and $a_3 = d_n^{-\alpha} (2\sqrt{\theta} + \sqrt{\phi})^2$. By integrating the probability density function of w_{iI}^n over the value domain in Eq. (25), we obtain $P_{b3|s_{iI}=-d_2}$ as:

$$P_{b3|s_{iI}=-d_2} = Q\left(\sqrt{\frac{a_1\rho}{t_1^2}\alpha_i^2}\right) - Q\left(\sqrt{\frac{a_3\rho}{t_1^2}\alpha_i^2}\right). \quad (26)$$

Similarly as in $P_{b3|s_{iI}=-d_2}$, $P_{b3|s_{iI}=-d_1}$ is:

$$\begin{aligned} P_{b3|s_{iI}=-d_1} &= \Pr\left(w_{iI}^n < \frac{d_1 - d_2}{2}\right) \\ &\quad + \Pr\left(w_{iI}^n > \frac{3d_1 + d_2}{2}\right) \\ &= Q\left(\sqrt{\frac{a_1\rho}{t_1^2}\alpha_i^2}\right) + Q\left(\sqrt{\frac{a_2\rho}{t_1^2}\alpha_i^2}\right). \end{aligned} \quad (27)$$

Additionally, we also show that $P_{b3|s_{iI}=d_1} = P_{b3|s_{iI}=-d_1}$ and $P_{b3|s_{iI}=d_2} = P_{b3|s_{iI}=-d_2}$. Due to the symmetrical property of the constellation in Fig. 2, we obtain $P_{b3} = P_{b4}$. Therefore, the average BER for one codeword on the i -th parallel channel is:

$$\begin{aligned} P_i^{n1} &= \frac{1}{2} \left[2Q\left(\sqrt{\frac{a_1\rho}{t_1^2}\alpha_i^2}\right) + Q\left(\sqrt{\frac{a_2\rho}{t_1^2}\alpha_i^2}\right) \right. \\ &\quad \left. - Q\left(\sqrt{\frac{a_3\rho}{t_1^2}\alpha_i^2}\right) \right]. \end{aligned} \quad (28)$$

Next, in the overall fading domain we evaluate the average BER of NU:

$$\begin{aligned} \bar{P}_{n1} &= \int_0^1 \dots \int_0^1 \frac{1}{N} \sum_{i=1}^N P_i^{n1}(x_i) \\ &\quad \times f_{X_1, \dots, X_N}(x_1, \dots, x_N) dx_1 \dots dx_N. \end{aligned} \quad (29)$$

Theorem 1. The average BER of NU in the overall fading domain can be approximated as:

$$\begin{aligned} \bar{P}_{n1} &\simeq \frac{c_{J1}}{2N} \left[\frac{1}{2} \sum_{\sigma \in \mathcal{S}_N} \sum_{j=1}^N B(p_{j1} + 1, q_1 + 1) G(t_1, p_{j1}, q_1) \right. \\ &\quad \times \prod_{\substack{i=1 \\ i \neq j}}^N B(p_{i1} + 1, q_1 + 1) + \sum_{\sigma_1, \sigma_2 \in \mathcal{S}_N} \text{sgn}(\sigma_1) \text{sgn}(\sigma_2) \\ &\quad \times \sum_{j=1}^N B(p_{j2} + 1, q_1 + 1) G(t_1, p_{j2}, q_1) \\ &\quad \left. \times \prod_{\substack{i=1 \\ i \neq j}}^N B(p_{i2} + 1, q_1 + 1) \right], \end{aligned} \quad (30)$$

where \mathcal{S}_N is the set of the permutations of $\{1, 2, \dots, N\}$ and $\text{sgn}(\sigma)$ denotes the sign of the permutations σ , $q_1 = M - N$, $p_{k1} = M - N + 2\sigma(k) - 2$, $p_{k2} = M - N + \sigma_1(k) + \sigma_2(k) - 2$, $k \in \{i, j\}$. $B(x, y)$ is the Beta

function defined in [29]. $G(t, x, y)$ is:

$$\begin{aligned} G(t, x, y) &= \frac{1}{3} F\left(\frac{a_1}{2}\right) + \frac{1}{6} F\left(\frac{a_2}{2}\right) - \frac{1}{6} F\left(\frac{a_3}{2}\right) \\ &\quad + F\left(\frac{2a_1}{3}\right) + \frac{1}{2} F\left(\frac{2a_2}{3}\right) - \frac{1}{2} F\left(\frac{2a_3}{3}\right), \end{aligned}$$

where $F(u) = {}_1F_1(x + 1; x + y + 2; -\frac{u\rho}{i^2})$ and ${}_1F_1(a; b; z)$ is the generalized hypergeometric function [30].

Proof: See Appendix B.

b) The case when $M < N < 2M$.

NU receives symbols on $N - M$ QPSK symbol channels and $2M - N$ NOMA symbol channels, so the average BER is:

$$P_{n2} = \frac{1}{M} \left(r P_n^{\text{QPSK}} + \sum_{i=1}^q P_i^{n2} \right), \quad (31)$$

where P_n^{QPSK} is the average BER performance on the QPSK symbol channel [28]:

$$P_n^{\text{QPSK}} \approx Q\left(\sqrt{\frac{P d_n^{-\alpha}}{t_2^2 N_0}}\right). \quad (32)$$

Considering the NOMA symbol channels, only the number of channels differs between $M \geq N$ and $M < N < 2M$ cases, so P_i^{n2} can be expressed similarly as P_i^{n1} :

$$\begin{aligned} P_i^{n2} &= \frac{1}{2} \left[2Q\left(\sqrt{\frac{a_1\rho}{t_1^2}\alpha_i^2}\right) + Q\left(\sqrt{\frac{a_2\rho}{t_1^2}\alpha_i^2}\right) \right. \\ &\quad \left. - Q\left(\sqrt{\frac{a_3\rho}{t_1^2}\alpha_i^2}\right) \right]. \end{aligned} \quad (33)$$

The average BER of NU is evaluated in the overall fading domain:

$$\bar{P}_{n2} = \frac{1}{M} \left(r P_n^{\text{QPSK}} + \sum_{i=1}^q \bar{P}_i^{n2} \right). \quad (34)$$

By applying **Lemma 1** and the same argument as in **Theorem 1**, we obtain the approximate expression of

$T_n = \sum_{i=1}^q \bar{P}_i^{n2}$ as:

$$\begin{aligned} T_n &\simeq \frac{c_{J2}}{2} \left[\frac{1}{2} \sum_{\sigma \in \mathcal{S}_q} \sum_{j=1}^q B(p'_{j1} + 1, q_2 + 1) G(t_2, p'_{j1}, q_2) \right. \\ &\quad \times \prod_{\substack{i=1 \\ i \neq j}}^q B(p'_{i1} + 1, q_2 + 1) + \sum_{\sigma_1, \sigma_2 \in \mathcal{S}_q} \text{sgn}(\sigma_1) \text{sgn}(\sigma_2) \\ &\quad \times \sum_{j=1}^q B(p'_{j2} + 1, q_2 + 1) G(t_2, p'_{j2}, q_2) \\ &\quad \left. \times \prod_{\substack{i=1 \\ i \neq j}}^q B(p'_{i2} + 1, q_2 + 1) \right], \end{aligned} \quad (35)$$

where \mathcal{S}_q is the set of the permutations of $\{1, 2, \dots, q\}$ and $p'_{k1} = N - M + 2\sigma(k) - 2$, $p'_{k2} = N - M + \sigma_1(k) + \sigma_2(k) - 2$, $k \in \{i, j\}$ and $q_2 = N - M$.

Substituting Eqs. (32) and (35) into Eq. (34), we achieve the approximate average BER for NU.

c) The case when $N > 2M$.

The user's channel is decomposed into M complex Gaussian channels. From Eq. (20), the average BER for NU is:

$$\bar{P}_{n3} \approx Q \left(\sqrt{\frac{Pd_n^{-\alpha}}{t_3^2 N_0}} \right). \quad (36)$$

3.2. BER of FU

a) The case when $M \geq N$.

In this case, FU also receives signals on N parallel NOMA symbol channels, so the average BER is expressed as:

$$P_{f1} = \frac{1}{N} \sum_{i=1}^N P_i^{f1}, \quad (37)$$

where P_i^{f1} is the average BER on the i -th parallel channel. FU's signals can be identified by the first two bits b_1, b_2 of the transmitted symbol. This generates the result of $P_i^{f1} = \frac{1}{2}(P_{b1} + P_{b2})$. P_{bj} is the error probability of j -th bit $j = 1, 2$. Along similar lines, in the NU case, we also get $P_{b1} = P_{b2}$ and:

$$P_i^{f1} = \frac{1}{2} \left[Q \left(\sqrt{\frac{c_1 \rho}{t_1^2} \beta_i^2} \right) + Q \left(\sqrt{\frac{c_2 \rho}{t_1^2} \beta_i^2} \right) \right], \quad (38)$$

where:

$$c_1 = d_f^{-\alpha} \left(\sqrt{\theta} - \sqrt{\phi} \right)^2, \quad c_2 = d_f^{-\alpha} \left(\sqrt{\theta} + \sqrt{\phi} \right)^2.$$

Due to the similarity of Joint-PDF of X_i and Y_i in Eqs. (7) and (8), it can be shown that:

$$\begin{aligned} \bar{P}_{f1} \approx & \frac{c_{J1}}{4N} \left[\frac{1}{2} \sum_{\sigma \in S_N} \sum_{j=1}^N \mathbf{B}(p_{j1} + 1, q_1 + 1) \mathbf{H}(t_1, p_{j1}, q_1) \right. \\ & \times \prod_{\substack{i=1 \\ i \neq j}}^N \mathbf{B}(p_{i1} + 1, q_1 + 1) + \sum_{\sigma_1, \sigma_2 \in S_N} \text{sgn}(\sigma_1) \text{sgn}(\sigma_2) \\ & \times \sum_{j=1}^N \mathbf{B}(p_{j2} + 1, q_1 + 1) \mathbf{H}(t_1, p_{j2}, q_1) \\ & \left. \times \prod_{\substack{i=1 \\ i \neq j}}^N \mathbf{B}(p_{i2} + 1, q_1 + 1) \right], \quad (39) \end{aligned}$$

where $p_{i1}, p_{i2}, p_{j1}, p_{j2}$ and q_1 are defined in **Theorem 1**. $\mathbf{H}(t, x, y)$ is expressed through $F(u)$ as:

$$\mathbf{H}(t, x, y) = \frac{1}{3} \mathbf{F} \left(\frac{c_1}{2} \right) + \frac{1}{3} \mathbf{F} \left(\frac{c_2}{2} \right) + \mathbf{F} \left(\frac{2c_1}{3} \right) + \mathbf{F} \left(\frac{2c_2}{3} \right).$$

b) The case when $M < N < 2M$.

By the same argument as in the NU case, we get the average BER of FU on the overall fading domain:

$$\bar{P}_{f2} = \frac{1}{M} \left(r P_f^{\text{QPSK}} + \sum_{i=1}^q \bar{P}_i^{f2} \right). \quad (40)$$

Similarly as T_n , let $T_f = \sum_{i=1}^q \bar{P}_i^{f2}$. We can prove that

$$\begin{aligned} T_f \approx & \frac{c_{J2}}{4} \left[\frac{1}{2} \sum_{\sigma \in S_q} \sum_{j=1}^q \mathbf{B}(p'_{j1} + 1, q_2 + 1) \mathbf{H}(t_2, p'_{j1}, q_2) \right. \\ & \times \prod_{\substack{i=1 \\ i \neq j}}^q \mathbf{B}(p'_{i1} + 1, q_2 + 1) + \sum_{\sigma_1, \sigma_2 \in S_q} \text{sgn}(\sigma_1) \text{sgn}(\sigma_2) \\ & \times \sum_{j=1}^q \mathbf{B}(p'_{j2} + 1, q_2 + 1) \mathbf{H}(t_2, p'_{j2}, q_2) \\ & \left. \times \prod_{\substack{i=1 \\ i \neq j}}^q \mathbf{B}(p'_{i2} + 1, q_2 + 1) \right], \quad (41) \end{aligned}$$

where $p'_{i1}, p'_{i2}, p'_{j1}, p'_{j2}$ and q_2 are defined in Eq. (35). P_f^{QPSK} is the average BER in the QPSK symbol channel given by:

$$P_f^{\text{QPSK}} \approx Q \left(\sqrt{\frac{Pd_f^{-\alpha}}{t_2^2 N_0}} \right). \quad (42)$$

By substituting Eqs. (41) and (42) into (40), we derive the approximate expression for the average BER of FU.

c) The case when $N > 2M$.

The average BER for FU is evaluated as:

$$\bar{P}_{f3} \approx Q \left(\sqrt{\frac{Pd_f^{-\alpha}}{t_3^2 N_0}} \right). \quad (43)$$

The summary theoretical analysis BER performance is shown in Table 1.

4. Works Relating to ZF, BD, and ST

In this section, we briefly mention the precoding techniques as ZF, BD, and ST. This serves as a basis for comparing them with GSVD in terms of the BER performance.

4.1. ZF Based Precoding

Zero-forcing based precoding [10] is valid only when $2M \leq N$. In this case, the detection matrices at the users are $\mathbf{U}_n = \mathbf{I}_M$, and $\mathbf{U}_f = \mathbf{I}_M$. The precoding matrix is:

$$\mathbf{V} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}, \quad (44)$$

where $\mathbf{H} \in \mathbb{C}^{2M \times N}$ is denoted as $\mathbf{H} = [\mathbf{H}_n^H \ \mathbf{H}_f^H]^H$. The precoding and detection processes are presented through the following equations:

$$\begin{aligned} \mathbf{U}_n^H \mathbf{H}_n \mathbf{V} &= [\mathbf{I}_M \ \mathbf{0}], \\ \mathbf{U}_f^H \mathbf{H}_f \mathbf{V} &= [\mathbf{0} \ \mathbf{I}_M]. \end{aligned} \quad (45)$$

Using $E(\mathbf{S}\mathbf{S}^H) = \mathbf{I}_N$, from Eq. (2), we easily obtain the power normalization factor given by:

$$t_{ZF} = \sqrt{\frac{2M}{N - 2M}}. \quad (46)$$

Tab. 1. Analysis of BER performance of NU and FU for GSVD based precoding.

BER	Antenna configurations		
	$M \geq N$	$M < N < 2M$	$N > 2M$
NU	$\bar{P}_{n1} = E_{X_1, \dots, X_N} \left(\frac{1}{N} \sum_{i=1}^N P_i^{n1} \right)$ $\bar{P}_{n1} \simeq \text{Eq. (30)}$	$\bar{P}_{n2} = \frac{1}{M} \left[rQ \left(\sqrt{\frac{Pd_n^- \alpha}{t_2^2 N_0}} \right) + E_{X_1, \dots, X_q} \left(\sum_{i=1}^q P_i^{n2} \right) \right]$ $\bar{P}_{n2} \simeq \frac{1}{M} \left[rQ \left(\sqrt{\frac{Pd_n^- \alpha}{t_2^2 N_0}} \right) + T_n \right], T_n \simeq \text{Eq. (35)}$	$\bar{P}_{n3} \approx Q \left(\sqrt{\frac{Pd_n^- \alpha}{t_3^2 N_0}} \right)$
FU	$\bar{P}_{f1} = E_{Y_1, \dots, Y_N} \left(\frac{1}{N} \sum_{i=1}^N P_i^{f1} \right)$ $\bar{P}_{f1} \simeq \text{Eq. (39)}$	$\bar{P}_{f2} = \frac{1}{M} \left[rQ \left(\sqrt{\frac{Pd_f^- \alpha}{t_2^2 N_0}} \right) + E_{Y_1, \dots, Y_q} \left(\sum_{i=1}^q P_i^{f2} \right) \right]$ $\bar{P}_{f2} \simeq \frac{1}{M} \left[rQ \left(\sqrt{\frac{Pd_f^- \alpha}{t_2^2 N_0}} \right) + T_f \right], T_f \simeq \text{Eq. (41)}$	$\bar{P}_{f3} \approx Q \left(\sqrt{\frac{Pd_f^- \alpha}{t_3^2 N_0}} \right)$

4.2. BD Based Precoding

Similarly to ZF based precoding [10], BD based precoding is valid only when $2M \leq N$. Carrying out SVD of \mathbf{H}_n and \mathbf{H}_f , can be obtained as:

$$\begin{aligned} \mathbf{H}_n &= \tilde{\mathbf{U}}_n \begin{bmatrix} \tilde{\mathbf{D}}_n^{(1)} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_n^{(1)} & \tilde{\mathbf{V}}_n^{(0)} \end{bmatrix}^H, \\ \mathbf{H}_f &= \tilde{\mathbf{U}}_f \begin{bmatrix} \mathbf{O} & \tilde{\mathbf{D}}_f^{(1)} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_f^{(0)} & \tilde{\mathbf{V}}_f^{(1)} \end{bmatrix}^H, \end{aligned} \quad (47)$$

where $\tilde{\mathbf{D}}_n^{(1)}, \tilde{\mathbf{D}}_f^{(1)} \in \mathbb{C}^{M \times M}$ and $\tilde{\mathbf{V}}_n^{(0)}, \tilde{\mathbf{V}}_f^{(0)} \in \mathbb{C}^{N \times (N-M)}$. Next, using SVD to $\mathbf{H}_n \tilde{\mathbf{V}}_f^{(0)}$ and $\mathbf{H}_f \tilde{\mathbf{V}}_n^{(0)}$, the results of the analyses are presented as:

$$\begin{aligned} \mathbf{H}_n \tilde{\mathbf{V}}_f^{(0)} &= \mathbf{U}_n \begin{bmatrix} \mathbf{D}_n^{(1)} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{V}_n^{(1)} & \mathbf{V}_n^{(0)} \end{bmatrix}^H, \\ \mathbf{H}_f \tilde{\mathbf{V}}_n^{(0)} &= \mathbf{U}_f \begin{bmatrix} \mathbf{O} & \mathbf{D}_f^{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_f^{(0)} & \mathbf{V}_f^{(1)} \end{bmatrix}^H, \end{aligned} \quad (48)$$

with $\mathbf{D}_n^{(1)}, \mathbf{D}_f^{(1)} \in \mathbb{C}^{M \times M}$ and $\mathbf{V}_n^{(1)}, \mathbf{V}_f^{(1)} \in \mathbb{C}^{(N-M) \times M}$. The precoding matrix \mathbf{V} can be obtained by concatenation of the precoding matrices as:

$$\mathbf{V} = \begin{bmatrix} \tilde{\mathbf{V}}_f^{(0)} \mathbf{V}_n^{(1)} & \tilde{\mathbf{V}}_n^{(0)} \mathbf{V}_f^{(1)} \end{bmatrix}. \quad (49)$$

The strategies of precoding and detection at BS and the users respectively can be performed by:

$$\begin{aligned} \mathbf{U}_n^H \mathbf{H}_n \mathbf{V} &= \begin{bmatrix} \mathbf{D}_n^{(1)} & \mathbf{O} \end{bmatrix}, \\ \mathbf{U}_f^H \mathbf{H}_f \mathbf{V} &= \begin{bmatrix} \mathbf{O} & \mathbf{D}_f^{(1)} \end{bmatrix}. \end{aligned} \quad (50)$$

By using $E(\mathbf{S}\mathbf{S}^H) = \mathbf{I}_N$, from Eq. (2) the power normalization factor can be:

$$t_{BD} = \sqrt{2M}. \quad (51)$$

4.3. ST Based Precoding

ST based precoding is mentioned in [12] and is valid when $M \leq N$.

a) The case when $M \leq N < 2M$.

From (47), we concatenate $\tilde{\mathbf{V}}_n^{(0)}$ and $\tilde{\mathbf{V}}_f^{(0)}$, the matrix \mathbf{H} is:

$$\mathbf{H} = \begin{bmatrix} \tilde{\mathbf{V}}_n^{(0)} & \tilde{\mathbf{V}}_f^{(0)} \end{bmatrix}^H. \quad (52)$$

Next, realizing SVD decomposition \mathbf{H} , we obtain:

$$\mathbf{H} = \tilde{\mathbf{U}} \tilde{\mathbf{D}} \begin{bmatrix} \tilde{\mathbf{V}}^{(1)} & \tilde{\mathbf{V}}^{(0)} \end{bmatrix}^H, \quad (53)$$

where $\tilde{\mathbf{V}}^{(0)} \in \mathbb{C}^{N \times (2M-N)}$. Let, QR decomposition be:

$$\begin{aligned} \mathbf{Q}_n \mathbf{R}_n &= \mathbf{H}_n \begin{bmatrix} \tilde{\mathbf{V}}^{(0)} & \tilde{\mathbf{V}}_f^{(0)} \end{bmatrix}, \\ \mathbf{Q}_f \mathbf{R}_f &= \mathbf{H}_f \begin{bmatrix} \tilde{\mathbf{V}}^{(0)} & \tilde{\mathbf{V}}_n^{(0)} \end{bmatrix}, \end{aligned} \quad (54)$$

where $\mathbf{Q}_n, \mathbf{Q}_f \in \mathbb{C}^{M \times M}$. By setting the precoding matrix $\mathbf{V} = \begin{bmatrix} \tilde{\mathbf{V}}^{(0)} & \tilde{\mathbf{V}}_f^{(0)} & \tilde{\mathbf{V}}_n^{(0)} \end{bmatrix}$ and choosing $\mathbf{U}_n = \mathbf{Q}_n, \mathbf{U}_f = \mathbf{Q}_f$, the simultaneous triangularization of \mathbf{H}_n and \mathbf{H}_f is:

$$\begin{aligned} \mathbf{U}_n^H \mathbf{H}_n \mathbf{V} &= \begin{bmatrix} \mathbf{R}_n & \mathbf{O} \end{bmatrix}, \\ \mathbf{U}_f^H \mathbf{H}_f \mathbf{V} &= \begin{bmatrix} \mathbf{R}'_f & \mathbf{O} & \mathbf{R}''_f \end{bmatrix}, \end{aligned} \quad (55)$$

where $\mathbf{R}_n \in \mathbb{C}^{M \times M}$, $\mathbf{R}'_f \in \mathbb{C}^{M \times (2M-N)}$ and $\mathbf{R}''_f \in \mathbb{C}^{M \times (N-M)}$. Moreover, \mathbf{R}_n and $\mathbf{R}_f = \begin{bmatrix} \mathbf{R}'_f & \mathbf{R}''_f \end{bmatrix}$ are upper-triangular matrices with real-valued entries on their main diagonals.

From Eq. (2), the power normalization factor in ST based precoding case is given as:

$$t_{ST} = \sqrt{N}. \quad (56)$$

b) The case when $N \geq 2M$.

From Eq. (47), we get the precoding matrix such as:

$$\mathbf{V} = \begin{bmatrix} \tilde{\mathbf{V}}_f^{(0)} & \tilde{\mathbf{V}}_n^{(0)} \end{bmatrix}. \quad (57)$$

Realizing triangularization $\mathbf{H}_n \tilde{\mathbf{V}}_f^{(0)}$ and $\mathbf{H}_f \tilde{\mathbf{V}}_n^{(0)}$ by QR decomposition:

$$\begin{aligned} \mathbf{Q}_n \mathbf{R}_n &= \mathbf{H}_n \tilde{\mathbf{V}}_f^{(0)}, \\ \mathbf{Q}_f \mathbf{R}_f &= \mathbf{H}_f \tilde{\mathbf{V}}_n^{(0)}, \end{aligned} \quad (58)$$

with $\mathbf{R}_n, \mathbf{R}_f \in \mathbb{C}^{M \times (N-M)}$. Letting $\mathbf{U}_n = \mathbf{Q}_n$ and $\mathbf{U}_f = \mathbf{Q}_f$ the simultaneous triangularization of \mathbf{H}_n and \mathbf{H}_f is:

$$\begin{aligned} \mathbf{U}_n^H \mathbf{H}_n \mathbf{V} &= \begin{bmatrix} \mathbf{R}_n^{(1)} & \mathbf{R}_n^{(0)} & \mathbf{O} \end{bmatrix}; \\ \mathbf{U}_f^H \mathbf{H}_f \mathbf{V} &= \begin{bmatrix} \mathbf{O} & \mathbf{R}_f^{(1)} & \mathbf{R}_f^{(0)} \end{bmatrix}, \end{aligned} \quad (59)$$

with $\mathbf{R}_n^{(0)}, \mathbf{R}_f^{(0)} \in \mathbb{C}^{M \times (N-2M)}$. The two upper-triangular matrices are $\mathbf{R}_n^{(1)}, \mathbf{R}_f^{(1)} \in \mathbb{C}^{M \times M}$ with real-valued entries on their main diagonals.

From Eq. (2), we can easily obtain the power normalization factor as:

$$t_{ST} = \sqrt{2(N - M)}. \quad (60)$$

5. Numerical Results

For numerical simulations, we carry out the Monte Carlo simulation over 10^5 independent trials to verify the correctness of derived theoretical and approximate expressions of BER performance for NU and FU in GSVD-NOMA.

Suppose that the path-loss factor $\alpha = 2$, the distances $d_n = 1, d_f = 3$, the range of the power allocation ratio $\theta = 0.55 : 0.05 : 0.95$, and $\text{SNR} = \frac{P}{N_0} = 0 : 2.5 : 35$ dB. The simulation parameters are given in Table 2. We analyze the BER performance as a function of the transmission's SNR and θ using precoding schemes and different antenna configurations scenarios.

Tab. 2. Simulation parameters.

Path-loss factor	$\alpha = 2$
Distances	$d_n = 1, d_f = 3$
Power allocation ratio	$\theta = 0.55 : 0.05 : 0.95$
Number of transmit antennas	$N = 2, 3, 5, 7, 9$
Number of receive antennas	$N = 2, 4, 7$
Signal-to-noise ratio [dB]	$\text{SNR} = 0 : 2.5 : 35$
Number of trials	10^5

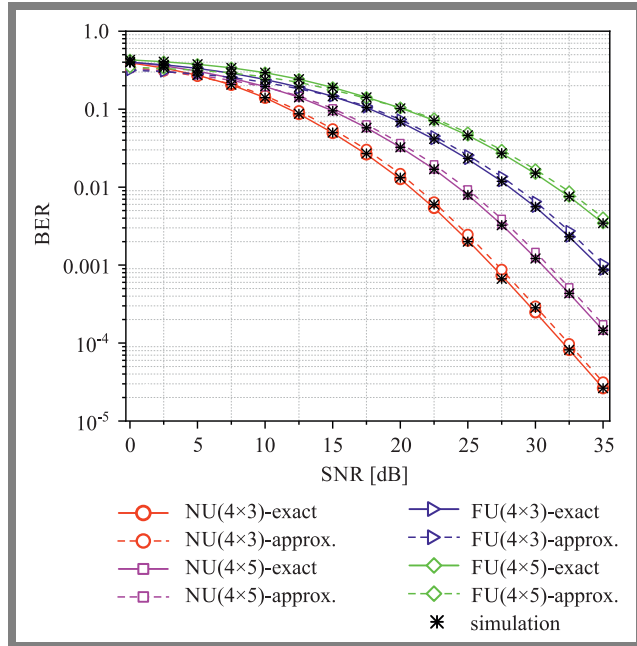


Fig. 3. BER performance of the near user and the far user, system performance curves are shown for two scenarios: $M \geq N$ (4×3) and $2M > N > M$ (4×5).

In Fig. 3, we calculate numerically the average BER performance of NU and FU, using Eqs. (30), (34), (39), and (40). Then, we validate the derived results by means of simulations under scenarios with the transmit and receive antenna configurations of 4×3 and 4×5 . The examples under consideration correspond with the $N < 2M$ scenario. As a result, the relatively high number of trials makes the simulation results more precise and closer to the theoretical results. Moreover, the approximate derivations agree quite well with the actual analysis.

Figure 4 shows the comparison of ST and GSVD precoding for $M = 4$ and $N = 5$, in the $M < N < 2M$ scenario. One may clearly observe that BER performance of ST is superior to that of GSVD for NU and FU. For example, the loss in BER performance of two users for GSVD, when compared to ST precoding, equals approx. 5 dB for BER of 0.016. This is due to the fact that the values of fading channel coefficients α and β in the decomposition process performed by GSVD are lower than or equal to 1, as mentioned in Section 2.2, whereas the entries of ST diagonal matrices do not apply to all conditions. Therefore, if an antenna configuration is chosen that belongs to the $M < N < 2M$ case, ST precoding should be taken into consideration. Furthermore, this choice is completely relevant due to the outperformance of ST precoding in terms of the ergodic rate region compared to GSVD [12].

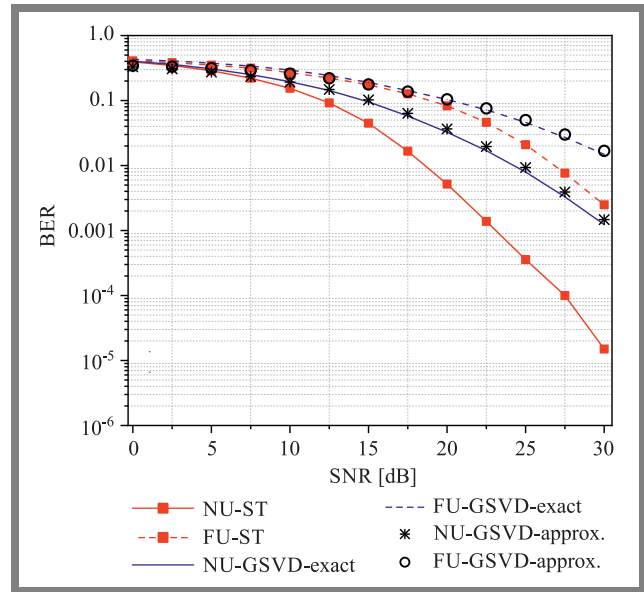


Fig. 4. Comparison of BER performance for ST based precoding and GSVD based precoding in the case of $2M > N > M$ for the (4×5) antenna configuration.

BER performance gain continues to be investigated in the case $N \geq 2M$ for ZF, BD, ST, and GSVD based precoding. Specifically, the number of transmit antennas is $N = 9$ and the users are equipped with $M = 4$ antennas, as shown in Fig. 5. We observed that BER performance of ST is dropped significantly compared to ZF, BD, and GSVD. This problem can be interpreted in such a way that based on triangular channel matrices in ST, the detection at the users' is undertaken in reverse order of the transmit signal vector. Moreover, for each subsequent

symbol, self-interference caused by the previously detected symbols needs to be eliminated. The self-interference cancellation process is usually imperfect, meaning that the system’s performance is negatively impacted. As the above analysis shows, the power normalization factor is equal for GSVD and ZF. Moreover, after decomposition, the channel matrices have diagonal entries equal to 1. As a result, we can observe that BER performance gain of ZF is the same as in GSVD.

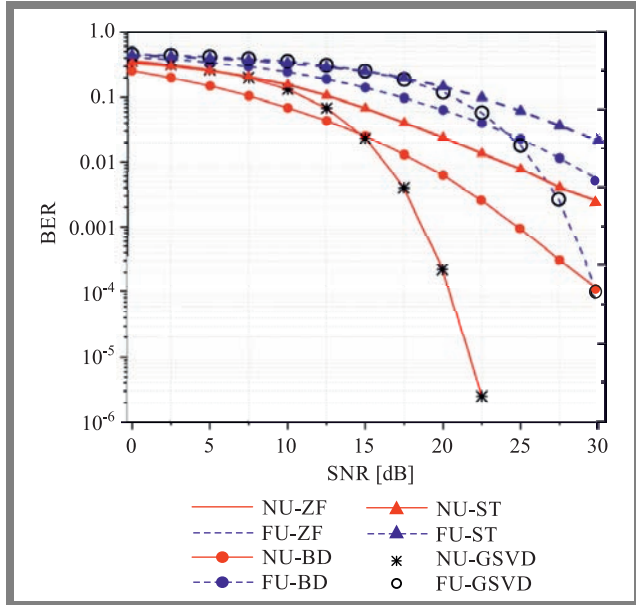


Fig. 5. BER performance comparison of ZF, BD, ST and GSVD based precoding schemes in the case of $N \geq 2M$ for antenna configurations.

In the low SNR regime, BD precoding performs better than ZF and GSVD. However, when transmit SNR is in the higher regime, ZF and GSVD precoding dramatically outperform BD precoding in terms of BER. In this antenna configuration, the parallel SISO channel in BD is dependent on small-scale fading elements, whereas in the case of ZF and GSVD the MIMO-NOMA channel is decomposed completely into the parallel AWGN channel. Therefore, when the average transmit power increases at the transmitters, BER performance of ZF and GSVD is considerably more superior. In scenarios in which the number of transmit antennas is greater than twice the number of receive antennas, GSVD and ZF based precoding schemes are chosen to improve the system’s BER performance.

Figure 6 shows the result of a comparison of two detection techniques applied to NU, namely joint ML and symbol-level SIC (SL-SIC) with the ideal SIC. SL-SIC is the technique studied in [23]. In SL-SIC, NU demodulates the FU’s signals and a hard decision is made, with channel coding not being performed. After that, NU regenerates the signals of FU and uses SIC to cancel them. With the ideal SIC, we assume that signals from FU are completely cancelled by NU. It is observed that joint ML significantly outperforms SL-SIC and offers almost the same performance as ideal SIC in terms of average BER. Performance of SL-SIC depends on power allocation coefficients θ and significantly degrades

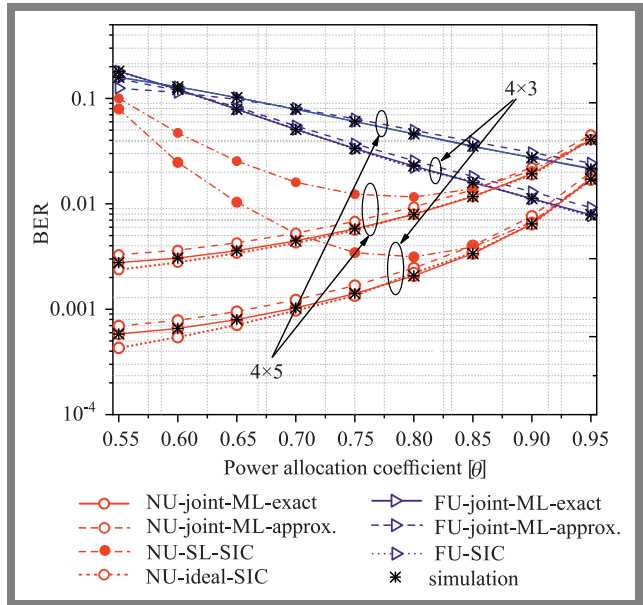


Fig. 6. BER performance comparison of the detector schemes: joint maximum-likelihood (ML), symbol-level SIC (SL-SIC) and ideal SIC under effect of power allocation coefficient θ .

BER for NU at small θ . A decrease in power allocated to FU symbols causes FU symbols to be detected erroneously by implementing SIC at NU. When θ is high, interference from NU may exert a weak impact on FU performance. However, at high θ , detection at NU is problematic due to low power allocation to NU. If the SL-SIC detector is applied to NU, power allocation should be considered.

In Fig. 7, we plot the average BER performance achieved by the joint ML detector versus transmit SNR when the number of transmit antennas changes, i.e. $N = 2, 3, 5, 7, 9$ and $M = 4$ for receive antennas. Based on the results shown, an increase in N decays BER performance of NU and FU be-

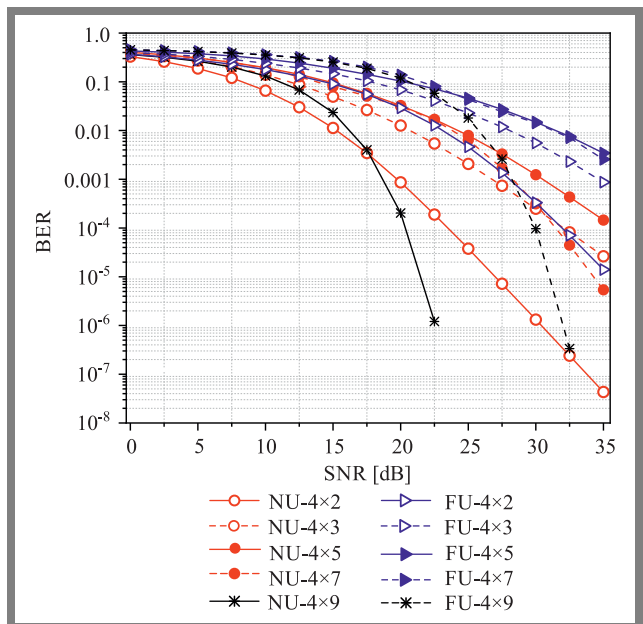


Fig. 7. BER performance of NU and FU for a varying number of transmit antennas N .

cause of the trade-off between diversity gain and multiplexing gain. More specifically, in the 4×2 and 4×3 scenarios, due to an increased number of parallel channels in the structure of GSVD, a decrease in performance is observed. Additionally, with an increase in the number of BS antennas in the $2M > N > M$ scenario, the number of parallel channels is constant (M), meaning that BER performance remains almost unchanged for NU and FU. Considering $N \geq 2M$, GSVD-MIMO channels decomposed into a number M of parallel Gaussian channels. It is shown in Eqs. (36) and (43), that the number of antennas does not affect the system's performance. In this case, BER outperforms other scenarios in the high SNR regime. Therefore, allowing a fixed number of receive antennas and an adjustable number of transmit antennas, in the low SNR regime, the transmitter should be equipped with a small number of aerials. However, assuming that the transmitted power at BS can be allocated at high levels permissively, BER performance is better when the number of transmit antennas satisfies the $N = 2M + 1$ condition.

Figure 8 shows the average BER performance of NU and FU when the number of receive antennas M increases. Here, the number of transmit antennas is modeled as $N = 5$. With an increase in the number of users' antennas M , it delivers better results in terms of BER performance. However, in the special case of $N \geq 2M (2 \times 5)$, the performance achieved is superior to all other solutions. So, if the number of transmit antennas is fixed, the number of receive antennas should satisfy the $N \geq 2M$ condition.

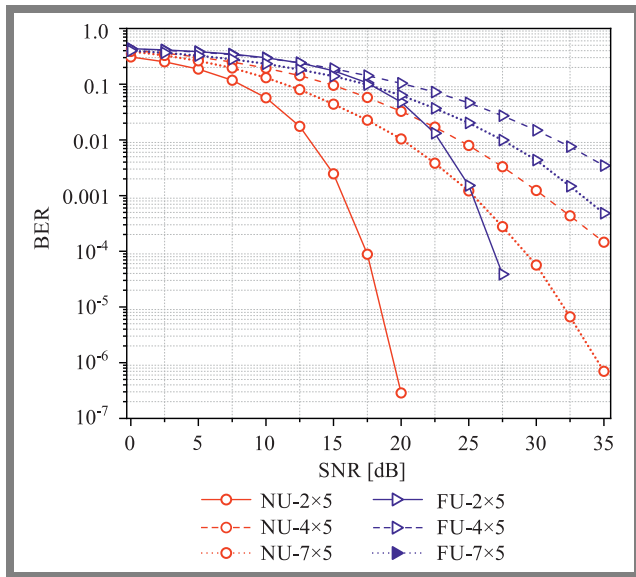


Fig. 8. BER performance of NU and FU for a varying number of receive antennas M .

6. Conclusions

In this paper, we consider the downlink MIMO-NOMA system with one base station and two users: near user and far user. The generalized singular value decomposition (GSVD) is applied to linear precoding and detection schemes at the BS and

the end users respectively. Through mathematical analyses, we obtain the approximate expression BER performance for NU and FU with the joint-modulation at BS and the joint maximum-likelihood detector at each user. It was shown that the exact results, approximate expressions, and simulation results are completely consistent with each other. Furthermore, the joint maximum-likelihood detector is almost similar to the ideal SIC and significantly outperforms symbol-level SIC in terms of average BER performance. In comparison with other precoding schemes, GSVD offers the same performance as zero-forcing precoding, outperforming block diagonalization and simultaneous triangularization in terms of BER performance when the antenna configuration satisfies the $N \geq 2M$ condition. In $M \leq N < 2M$ scenarios, simultaneous triangularization precoding should be considered due to its superior performance not only in terms of BER, but also in terms of the ergodic achievable rate region [12]. However, fact that it may be applied to all antenna configurations in a significant advantage of GSVD. We also investigated average BER performance of GSVD with a varying number of antennas. It has been observed that the system's performance is superior when the number of antennas satisfies the $N \geq 2M$ condition. The analysis performed may be extended to downlink and uplink MIMO-NOMA with more than two users, where each user uses higher-order modulation level (M -ary modulation), to conduct further studies. Moreover, it is better to consider multiple performance metrics to arise a trade-off.

Acknowledgements

This research has been funded by University of Science, Vietnam National University, Ho Chi Minh City (VNU-HCM) under grant number DT-VT 2022-05.

Appendix A: proof of Lemma 1

The joint probability density function of unordered $W_i = \frac{\alpha_i^2}{\beta_i^2}$ in [14] is represented as:

$$f_{W_1, \dots, W_q}(w_1, \dots, w_q) = c_{J2} \prod_{i=1}^q \frac{w_i^{M-q}}{(1+w_i)^{N+q}} \times \prod_{i < j}^q (w_i - w_j)^2, \quad (61)$$

where c_{J2} is defined in Lemma 1. We can rewrite

$W_i = \frac{X_i}{1-X_i}$, and the joint probability distribution of X_i will be expressed as:

$$\begin{aligned} F_{X_1, \dots, X_q}(x_1, \dots, x_q) &= \Pr(X_1 \leq x_1, \dots, X_q \leq x_q) \\ &= \Pr\left(W_1 \leq \frac{x_1}{1-x_1}, \dots, W_q \leq \frac{x_q}{1-x_q}\right) \\ &= \int_0^{\frac{x_1}{1-x_1}} \dots \int_0^{\frac{x_q}{1-x_q}} f_{W_1, \dots, W_q}(w_1, \dots, w_q) dw_1 \dots dw_q. \end{aligned} \quad (62)$$

Using the Leibniz integral rule, we obtain the joint probability density function of unordered X_i :

$$\begin{aligned} f_{X_1, \dots, X_q}(x_1, \dots, x_q) &= \frac{\partial F_{X_1, \dots, X_q}(x_1, \dots, x_q)}{\partial x_1 \dots \partial x_q} \quad (63) \\ &= c_{J2} \prod_{i < j}^q \left(\frac{x_i}{1-x_i} - \frac{x_j}{1-x_j} \right)^2 \prod_{i=1}^q x_i^{M-q} (1-x_i)^{N-M+2q-2} \\ &= c_{J2} \prod_{i < j}^q (x_i - x_j)^2 \prod_{i=1}^q x_i^r (1-x_i)^r. \end{aligned}$$

The proof is completed.

Appendix B: proof of Theorem 1

Using exponential bound for Q -function in [31]:

$$Q(x) \simeq \frac{1}{12} e^{-\frac{1}{2}x^2} + \frac{1}{4} e^{-\frac{2}{3}x^2}. \quad (64)$$

Then, the error probability on the i -th parallel channel in Eq. (28) is:

$$\begin{aligned} P_i^{n1} &\simeq \frac{1}{4} \left[\frac{1}{3} e^{-\frac{a_1 \rho \alpha_i^2}{2t_1^2}} + \frac{1}{6} e^{-\frac{a_2 \rho \alpha_i^2}{2t_1^2}} - \frac{1}{6} e^{-\frac{a_3 \rho \alpha_i^2}{2t_1^2}} \right. \\ &\quad \left. + e^{-\frac{2a_1 \rho \alpha_i^2}{3t_1^2}} + \frac{1}{2} e^{-\frac{2a_2 \rho \alpha_i^2}{3t_1^2}} - \frac{1}{2} e^{-\frac{2a_3 \rho \alpha_i^2}{3t_1^2}} \right] = \frac{1}{4} g(\alpha_i^2). \end{aligned} \quad (65)$$

From joint-PDF function in Eq. (7), conducting the average BER in fading channel:

$$\begin{aligned} \bar{P}_{n1} &\simeq \frac{c_{J1}}{4N} \int_0^1 \dots \int_0^1 \left[\sum_{i=1}^N g(x_i) \right] \left[\prod_{i=1}^N x_i^{M-N} (1-x_i)^{M-N} \right. \\ &\quad \left. \times \prod_{i < j}^N (x_i - x_j)^2 \right] dx_1 \dots dx_N. \end{aligned} \quad (66)$$

Consider term:

$$\begin{aligned} \prod_{i < j}^N (x_j - x_i) &= \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \\ x_1^2 & x_2^2 & \dots & x_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{N-1} & x_2^{N-1} & \dots & x_N^{N-1} \end{pmatrix} \\ &= \det V, \end{aligned} \quad (67)$$

where V is the Vandermonde matrix. Applying the Leibniz formula for the determinant of V :

$$\prod_{i < j}^N (x_j - x_i) = \sum_{\sigma \in S_N} \text{sgn}(\sigma) \prod_{i=1}^N x_i^{\sigma(i)-1}, \quad (68)$$

S_N is the set of the permutations of $\{1, 2, \dots, N\}$. Carrying out algebraic manipulation, we get:

$$\begin{aligned} \prod_{i < j}^N (x_j - x_i)^2 &= \sum_{\sigma \in S_N} \prod_{i=1}^N x_i^{2\sigma(i)-2} + 2 \sum_{\sigma_1, \sigma_2 \in S_N} \text{sgn}(\sigma_1) \\ &\quad \times \text{sgn}(\sigma_2) \prod_{i=1}^N x_i^{\sigma_1(i)+\sigma_2(i)-2}. \end{aligned} \quad (69)$$

Substituting Eq. (69) into Eq. (66):

$$\begin{aligned} \bar{P}_{n1} &\simeq \frac{c_{J1}}{4N} \int_{[0,1]^N} \left[\sum_{i=1}^N g(x_i) \right] \left[\prod_{i=1}^N x_i^{M-N} (1-x_i)^{M-N} \right] \\ &\quad \times \left[\sum_{\sigma \in S_N} \prod_{i=1}^N x_i^{2\sigma(i)-2} \right] dx_1 \dots dx_N \quad (70) \\ &\quad + \frac{c_{J1}}{2N} \int_{[0,1]^N} \left[\sum_{i=1}^N g(x_i) \right] \left[\prod_{i=1}^N x_i^{M-N} (1-x_i)^{M-N} \right] \\ &\quad \times \left[\sum_{\sigma_1, \sigma_2 \in S_N} \text{sgn}(\sigma_1) \text{sgn}(\sigma_2) \times \prod_{i=1}^N x_i^{\sigma_1(i)+\sigma_2(i)-2} \right] dx_1 \dots dx_N, \end{aligned}$$

where $\int_{[0,1]^N} = \int_0^1 \dots \int_0^1$ for brevity. By performing algebraic operations, we obtain the approximate expression of \bar{P}_{n1} as:

$$\begin{aligned} \bar{P}_{n1} &\simeq \frac{c_{J1}}{4N} \sum_{\sigma \in S_N} \sum_{j=1}^N \int_0^1 g(x_j) x_j^{p_{j1}} (1-x_j)^{q_1} dx_j \\ &\quad \times \prod_{\substack{i=1 \\ i \neq j}}^N \int_0^1 x_i^{p_{i1}} (1-x_i)^{q_1} dx_i + \frac{c_{J1}}{2N} \sum_{\sigma_1, \sigma_2 \in S_N} \text{sgn}(\sigma_1) \\ &\quad \times \text{sgn}(\sigma_2) \sum_{j=1}^N \int_0^1 g(x_j) x_j^{p_{j2}} (1-x_j)^{q_1} dx_j \\ &\quad \times \prod_{\substack{i=1 \\ i \neq j}}^N \int_0^1 x_i^{p_{i2}} (1-x_i)^{q_1} dx_i. \end{aligned} \quad (71)$$

Considering the following integral, $I_1 = \int_0^1 t^p (1-t)^q dt$ and $I_2 = \int_0^1 e^{-at} t^p (1-t)^q dt$ and applying two equations Eqs. (8.380) and (8.384) from [29], we can rewrite I_1 as:

$$I_1 = B(p+1, q+1), \quad (72)$$

$B(x, y)$ is the beta function. Using Eq. (3.383) from [29] I_2 is:

$$I_2 = B(p+1, q+1) {}_1F_1(p+1; p+q+2; -a), \quad (73)$$

where ${}_1F_1(a; b; z)$ is the generalized hypergeometric function. Applying Eqs. (72) and (73) to (71), the proof is completed.

References

- [1] L. Dai, *et al.*, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends", *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015 (DOI: 10.1109/MCOM.2015.7263349).
- [2] Y. Liu, Z. Qin, M. El-kashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond", *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017 (DOI: 10.1109/JPROC.2017.2768666).
- [3] Y. Saito, *et al.*, "Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access", in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2013 (DOI: 10.1109/VTC-Spring.2013.6692652).
- [4] K. Higuchi and A. Benjebbour, "Non-orthogonal Multiple Access (NOMA) with Successive Interference Cancellation for Future Radio Access", *IEICE Transactions on Communications*, vol. E98.B, pp. 403–414, 2015 (DOI: 10.1587/transcom.E98.B.403).

- [5] Q. Sun, S. Han, I. C-L, and Z. Pan, "On the Ergodic Capacity of MIMO NOMA Systems", *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, 2015 (DOI: 10.1109/LWC.2015.2426709).
- [6] S. Ali, E. Hossain, and D.I. Kim, "Non-Orthogonal Multiple Access (NOMA) for Downlink Multiuser MIMO Systems: User Clustering, Beamforming, and Power Allocation", *IEEE Access*, vol. 5, pp. 565–577, 2017 (DOI: 10.1109/ACCESS.2016.2646183).
- [7] Z. Ding, R. Schober, and H.V. Poor, "A General MIMO Framework for NOMA Downlink and Uplink Transmission Based on Signal Alignment", *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4438–4454, 2016 (DOI: 10.1109/TWC.2016.2542066).
- [8] H. Weingarten, Y. Steinberg, and S.S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel", *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006 (DOI: 10.1109/TIT.2006.880064).
- [9] Z. Chen and X. Dai, "MED Precoding for Multiuser MIMO-NOMA Downlink Transmission", *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5501–5505, 2017 (DOI: 10.1109/TVT.2016.2627218).
- [10] A. Krishnamoorthy, Z. Ding, and R. Schober, "Precoder Design and Statistical Power Allocation for MIMO-NOMA via User-Assisted Simultaneous Diagonalization", *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 929–945, 2021 (DOI: 10.1109/TCOMM.2020.3036453).
- [11] D. Senaratne and C. Tellambura, "GSVD Beamforming for Two-User MIMO Downlink Channel", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 6, pp. 2596–2606, 2013 (DOI: 10.1109/TVT.2013.2241091).
- [12] A. Krishnamoorthy, M. Huang, and R. Schober, "Precoder Design and Power Allocation for Downlink MIMO-NOMA via Simultaneous Triangularization", in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2021 (DOI: 10.1109/WCNC49053.2021.9417424).
- [13] Z. Chen, Z. Ding, X. Dai, and R. Schober, "Asymptotic Performance Analysis of GSVD-NOMA Systems with a Large-Scale Antenna Array", *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 575–590, 2019 (DOI: 10.1109/TWC.2018.2883102).
- [14] Z. Chen, Z. Ding, and X. Dai, "On the Distribution of the Squared Generalized Singular Values and Its Applications", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 1030–1034, 2019 (DOI: 10.1109/TVT.2018.2885122).
- [15] M.F. Hanif and Z. Ding, "Robust Power Allocation in MIMO-NOMA Systems", *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1541–1545, 2019 (DOI: 10.1109/LWC.2019.2926277).
- [16] C. Rao, Z. Ding, and X. Dai, "The Distribution Characteristics of Ordered GSVD Singular Values and Its Applications in MIMO-NOMA", *IEEE Communications Letters*, vol. 24, no. 12, pp. 2719–2722, 2020 (DOI: 10.1109/LCOMM.2020.3017796).
- [17] C. Rao, Z. Ding, and X. Dai, "GSVD-Based MIMO-NOMA Security Transmission", *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1484–1487, 2021 (DOI: 10.1109/LWC.2021.3071365).
- [18] Y. Qi and M. Vaezi, "Secure Transmission in MIMO-NOMA Networks", *IEEE Communications Letters*, vol. 24, no. 12, pp. 2696–2700, 2020 (DOI: 10.1109/LCOMM.2020.3016999).
- [19] X. Wang, F. Labeau, and L. Mei, "Closed-Form BER Expressions of QPSK Constellation for Uplink Non-Orthogonal Multiple Access", *IEEE Communications Letters*, vol. 21, no. 10, pp. 2242–2245, 2017 (DOI: 10.1109/LCOMM.2017.2720583).
- [20] F. Kara and H. Kaya, "BER Performances of Downlink and Uplink NOMA in the Presence of SIC Errors over Fading Channels", *IET Communications*, vol. 12, no. 15, pp. 1834–1844, 2018 (DOI: 10.1049/iet-com.2018.5278).
- [21] T. Assaf, A. Al-Dweik, M.E. Moursi, and H. Zeineldin, "Exact BER Performance Analysis for Downlink NOMA Systems Over Nakagami-Fading Channels", *IEEE Access*, vol. 7, pp. 134539–134555, 2019 (DOI: 10.1109/ACCESS.2019.2942113).
- [22] J.S. Yeom, H.S. Jang, K.S. Ko, and B.C. Jung, "BER Performance of Uplink NOMA With Joint Maximum-Likelihood Detector", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10295–10300, 2019 (DOI: 10.1109/TVT.2019.2933253).
- [23] C. Yan, *et al.*, "Receiver Design for Downlink Non-Orthogonal Multiple Access (NOMA)", in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–6, 2015 (DOI: 10.1109/VTC-Spring.2015.7146043).
- [24] —, "Wireless Technology Evolution Towards 5G: 3GPP release 13 to release 15 and beyond", 2017. (<https://www.5gamericas.org/wireless-technology-evolution-towards-5g-3gpp-release-13-to-release-15-and-beyond/>).
- [25] C.F. Van Loan, "A General Matrix Eigenvalue Algorithm", *SIAM Journal on Numerical Analysis*, vol. 12, no. 6, pp. 819–834, 1975 (<https://www.jstor.org/stable/2156413>).
- [26] A. Edelman and B.D. Sutton, "The Beta-Jacobi Matrix Model, the CS Decomposition, and Generalized Singular Value Problems", *Foundations of Computational Mathematics*, vol. 8, no. 2, pp. 259–285, 2008 (DOI: 10.1007/s10208-006-0215-9).
- [27] J. Tiefeng, "Limit theorems for beta-Jacobi ensembles", *Bernoulli*, vol. 19, no. 3, pp. 1028–1046, 2013 (DOI: 10.3150/12-BEJ495, <https://projecteuclid.org/journalArticle/Download?urlId=10.3150%2F12-BEJ495>).
- [28] A. Goldsmith, "Wireless Communications", *Cambridge University Press*, 2005 (DOI: 10.1017/CBO9780511841224).
- [29] D. Zwillinger and A. Jeffrey, *Table of integrals, series, and products*, 7th ed. Elsevier, 2007 (ISBN 978-0-12-373637-6).
- [30] F.W.L. Oliver, D.W. Lozier, R.F. Boisvert, and C.W. Clark, *NIST Handbook of Mathematical Functions*, 2010 (https://assets.cambridge.org/97805211/92255/copyright/9780521192255_copyright_info.pdf).
- [31] M. Chiani, D. Dardari, and M.K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels", *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 840–845, 2003 (DOI: 10.1109/TWC.2003.814350).



Ngo Thanh Hai received his B.Sc. from the University of Science, Vietnam National University, Ho Chi Minh City (VNU-HCM) and M.Sc. in Electronics and Telecommunications from Posts and Telecommunications Institute of Technology. Since 2019, he has been working as a researcher at the Faculty of Electronics and Telecommunications, University

of Science, VNU-HCM. His research activities focus on non-orthogonal multiple access (NOMA), multiple-input multiple-output (MIMO) and covert wireless communications.

E-mail: ngohai@hcmus.edu.vn

Department of Telecommunications and Networks, University of Science, VNU-HCM, District 5, Ho Chi Minh City, Vietnam



Dang Le Khoa received his B.E. and Ph.D. degrees in Radio Physics and Electronics from the University of Science, Vietnam National University, Ho Chi Minh City (VNU-HCM). He is the head of the Telecommunications and Networks Department, University of Science, VNU-HCM.

His current research interests are in the areas of wireless communications and digital signal processing for telecommunication.

E-mail: dlkhoa@hcmus.edu.vn

Department of Telecommunications and Networks, University of Science, VNU-HCM, District 5, Ho Chi Minh City, Vietnam

Performance Comparison of Optimization Methods for Flat-Top Sector Beamforming in a Cellular Network

Pampa Nandi and Jibendu Sekhar Roy

School of Electronics Engineering, Kalinga Institute of Industrial Technology University Bhubaneswar, Odisha, India

<https://doi.org/10.26636/jtit.2022.162122>

Abstract — The flat-top radiation pattern is necessary to form an appropriate beam in a sectored cellular network and to provide users with best quality services. The flat-top pattern offers sufficient power and allows to minimize spillover of signal to adjacent sectors. The flat-top sector beam pattern is relied upon in sectored cellular networks, in multiple-input multiple-output (MIMO) systems and ensures a nearly constant gain in the desired cellular sector. This paper presents a comparison of such optimization techniques as real-coded genetic algorithm (RGA) and particle swarm optimization (PSO), used in cellular networks in order to achieve optimum flat-top sector patterns. The individual parameters of flat-top sector beams, such as cellular coverage, ripples in the flat-top beam, spillover of radiation to the adjacent sectors and side lobe level (SLL) are investigated through optimization performed for 40° and 60° sectors. These parameters are used to compare the performance of the optimized RGA and PSO algorithms. Overall, PSO outperforms the RGA algorithm.

Keywords — flat-top sector beam, particle swarm optimization, real-coded genetic algorithm

1. Introduction

In cellular communication, a radio communication link is established between two users via a base station. The high-speed multimedia communication network consists of a number of cells covering the serviced area. Channel capacity, interference level, data rate and numerous security features are some of the issues that are of great significance for any cellular network. In order to enhance channel capacity and security, each cell is divided into a number of sectors. Each sector is serviced by a dedicated directional antenna. Unfortunately, these directional antennas radiate non-uniformly over the terrain. Therefore, in a cellular communication system, it is necessary to install antenna systems that are capable of generating the desired radiation patterns with almost constant power levels available to all users.

The flat-top radiation pattern is necessary to form a derived beam in a sectored cellular network and, hence, to provide the best quality of service to the users [1]–[3] and to assure lower spillover of signal to the neighboring sectors. It is found that the classical methods are easy to implement, but suffer from several drawbacks [4]–[5]. All the known methods have limitations regarding their coverage area, the

formation of ripples on the flat-top or the spillover of the beam to adjacent cells. These drawbacks can be minimized by deploying specific optimization methods.

2. Related Work

In [6], differential evolution (DE) is used to generate an optimally shaped beam pattern with multiple constraints. In cellular networks, in order to enhance network capacity and to ensure better spectral efficiency, cells are divided into a number of angular sectors. For the purpose of forming a sector beam with reduced SLL level, the zero forcing algorithm is used in [7]. For the design of the reflect-array aerial type with a flat-top beam for a remote sensing satellite system, the genetic algorithm (GA) is used in [8]. The formation of a flat-top pattern using a dipole array is reported in [9], using GA, and its performance is compared with results simulated with the use of high frequency structure simulator (HFSS) software.

Despite the large number of reports available in the literature concerned with the generation of flat-top beams using array synthesis and optimization methods, comparisons of performance of specific optimization methods are not available. Therefore, the aim of this work is to use specific optimization techniques, such as RGA and PSO, to generate a flat-top sector beam of a phased array antenna which covers the desired sector and is characterized by low ripples and reduced SLL. In addition, in both RGA and binary genetic algorithms (BGA), the phase-only optimization and amplitude-only optimization are considered for flat-top beam generation, and their performance is compared. Performance related to the maximum SLL, half power beam width (HPBW), maximum ripple formation in the beam pattern, cellular area coverage and spillover of the beam to the neighboring sectors is compared with that of the generated flat-top beams using RGA and PSO. RGA and PSO optimization methods provide good results in terms of flat-top beam generation, but the computation times encountered are very long compared to the classical approach. PSO is characterized by better performance in terms of flat-top sector beam generation. In MIMO, different types of beam patterns of antenna arrays are required [10], with an antenna array for flat-top cellular sector beam being one of them.

3. Description of Optimization Methods

3.1. Real Coded Genetic Algorithm (RGA)

GA is an evolutionary optimization technique and a stochastic method, searching for the global minimum by following the principles of genetics and natural selection. GA deals simultaneously with a large number of variables for global optimization. It is a stochastic method and any variable, whether discrete or continuous, may be used directly in the optimization process. The parameters considered include genes, chromosomes, population sizes, crossovers, selection, and mutations of the biological world. BGA changes the variables into an encoded binary string, while RGA works with continuous valued variables to optimize the cost function, although both algorithms follow the same principles of genetic recombination and natural selection [11]. RGA requires an adjustment of different operator or parameter values. Once all applicable conditions are satisfied, RGA needs fewer iterations to reach the optimum value and provide the best result.

The real coded genetic algorithm operates based on a real value parameter. Chromosomes are formed by groups of random (0 to 1) valued genes and a set of such chromosomes constitutes the initial population [12]. After evaluating the cost of each chromosome from this population, 50% best valued chromosomes are kept for the natural selection process and the rest is discarded. These selected parent chromosomes create offspring by combing the weighted portions of both parents. Weight b is calculated using a random number r and cross over operator μ , as [11], [12]:

$$b = \begin{cases} (2r)^{\frac{1}{1+\mu}} & \text{if } r > 0.5 \\ \left(\frac{1-r}{2}\right)^{\frac{1}{1+\mu}} & \text{otherwise} \end{cases} \quad (1)$$

The newly generated offspring are:

$$\begin{aligned} \text{Offspring1} &= \frac{(1+b)\text{parent}_1 + (1-b)\text{parent}_2}{2} \\ \text{Offspring2} &= \frac{(1-b)\text{parent}_1 + (1+b)\text{parent}_2}{2} \end{aligned} \quad (2)$$

Mutation is performed on some randomly selected chromosomes to continue the search in a diversified direction, according to the probability of mutation. If η is the mutation operator, then mutation weight p is:

$$p = \begin{cases} (2r)^{\frac{1}{1+\eta}} - 1 & \text{if } r \leq 0.5 \\ p = 1 - (2-2r)^{\frac{1}{1+\eta}} & \text{otherwise} \end{cases} \quad (3)$$

3.2. Particle Swarm Optimization (PSO)

PSO is another evolutionary algorithm which has evolved from the behavior of animals which do not have a leader in their population or swarm (e.g. bird flocks). The PSO algorithm (search technique) is used to identify the best settings or parameters required to achieve a desired objective [13]. In the PSO approach, each single solution in the search space of an objective function is commonly known as a bird or a particle, and the set of random particles is the initial swarm. Each particle resides at a position within the search space, the

fitness of each particle represents the quality of its position. Particles may evaluate their actual positions using the function to be optimized. The randomly generated solutions or swarms propagate in the design space, over a number of iterations, towards the optimal solution. The velocity of each particle is updated by its own best position solution found so far, with the best particle ($pbest$), the best solution that has been found so far by its neighbors ($lbest$) and another best value that is tracked by the particle swarm optimizer, obtained until now by any particle – global best ($gbest$). The swarm converges towards the optimal position by updating its information at every iteration and by checking the termination conditions.

Each particle is characterized by position vector $x_i(t)$ and velocity vector $v_i(t)$. Individual knowledge of the $pbest$ particle, its own best-so-far position and social knowledge $gbest$ is the $pbest$ value in the swarm. In PSO, the velocity update equation is [13]:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot \text{rand} \cdot [pbest - x_i(t)] + c_2 \cdot \text{rand} \cdot [gbest - x_i(t)], \quad (4)$$

and the position update equation is:

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (5)$$

where i is the number of iterations, v_i is particle velocity at i -th iteration, x_i is the current particle position or solution, w is the inertia weight factor, a random number between (0, 1). c_1 , c_2 are the learning factors or constriction factors, such as: c_1 (known as a cognitive parameter) and c_2 (known as a social parameter) are acceleration factors that guarantee the convergence and improve its velocity. Usually, $c_1 + c_2 = 4$. A particle updates its velocity and position using the procedures described above, at every iteration toward the best solution.

4. Simulation Study

For simulation purposes, MATLAB has been used to generate an optimized flat-top cellular sector pattern, while RGA and PSO have been used as optimization tools. The ordinary sector beam, the desired flat-top sector beam and a typical optimized flat-top sector beam in a cellular network are shown in Fig. 1.

The flat-top pattern is characterized by three parameters: SLL, ripple and transition width. If one of them decreased, the others will increase. The techniques relied upon for shaping these patterns focus mostly on controlling the amplitude or phase of the current feeding each antenna element. In this paper, RGA and PSO are used for the generation of optimized flat-top patterns.

A linear antenna array with an element spacing value of d is shown in Fig. 2.

In the case of amplitude-only optimization, common amplitude distribution with a fixed phase of 0° or 180° for the desired sector beam pattern is considered. For the phase-only synthesis, the excitation phase distribution of the antenna array varies within the range of $0^\circ \leq \varphi \leq 180^\circ$ to achieve a flat-top beam through optimization. For an array of isotropic N an-

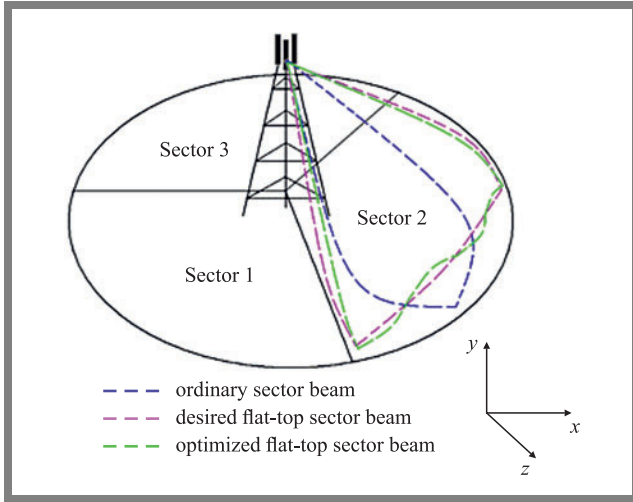


Fig. 1. Flat-top sector beam in a cellular network.

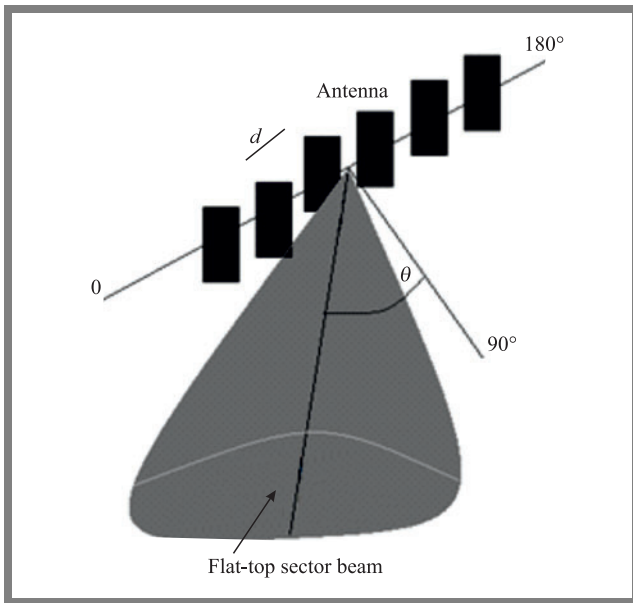


Fig. 2. Linear antenna array with a flat-top sector beam.

tennas with the inter-element distance of d along the y -axis, the array factor $AF(\theta)$ in the principal y - z plane is [14], [15]:

$$AF(\theta) = \sum_{n=1}^N a_n e^{j\varphi_n} e^{j(n-1)kd \sin \theta}, \quad (6)$$

where $k = 2\pi/\lambda$, n is the number of elements, λ is the wavelength, a_n and φ_n are the current amplitude and phase of the n -th antenna element, θ is the polar angle measured from the broadside direction (as shown in Fig. 2). The cost function needed to achieve the desired flat-top array pattern is expressed as:

$$\text{Cost} = [AF_d(\theta_{\text{sec}}) - AF(\theta_{\text{sec}})]^2 + [SLL_d - SLL_{\text{max}}]^2, \quad (7)$$

where θ_{sec} is the angular sector region of the beam pattern, $AF_d(\theta_{\text{sec}})$ is the desired pattern, $AF(\theta_{\text{sec}})$ is the observed pattern, SLL_d is the desired SLL and SLL_{max} is the maximum SLL of the observed pattern. $AF(\theta_{\text{sec}})$ is the normalized array factor of the observed pattern within the beam range θ_{sec} and for the sector beam range of θ_{sec} , $AF_d(\theta_{\text{sec}}) = 1$. Here, the

goal is to minimize the cost function and to generate a flat-top beam which covers the desired sector area with a minimum SLL value.

4.1. Assumptions for Simulation Tests

In this simulation, the antennas in the array are assumed to be of the isotropic variety and are distributed linearly and uniformly. The various parameters considered in the simulation for RGA- and PSO-based optimization are as follows. For amplitude-only optimization with RGA, the crossover rate = 0.7, the crossover operator $\mu = 20$, the mutation rate = 0.15, the mutation operator $\eta = 20$, the population size = 600 and the number of iterations = 1000. For phase-only optimization with RGA, the crossover rate = 0.6, the crossover operator $\mu = 20$, the mutation rate = 0.2, the mutation operator $\eta = 20$, the population size = 600 and the number of iterations = 1000. For amplitude-only and phase-only optimizations using PSO, cognitive $c_1 = 1$, the social parameter $c_2 = 3$, the constriction factor $c = 1$, the inertia weight $W = 0$ to 1, the swarm size = 600 and the number of iterations = 1000. In RGA, single point crossover and roulette wheel selection are used.

4.2. Flat-top Beamforming by Amplitude-only Optimization Using RGA

Non-linear amplitude optimization is used to identify the distribution of the excitation current amplitude of the individual elements in order to achieve the desired flat-top pattern. In this section, RGA is used as the optimization tool, with Eq. (7) being the cost function used to determine the optimum excitation amplitude distribution of the array. A uniform linear array (ULA) of $N = 11$ elements is considered, with the inter-element spacing of $d = \lambda/2$, with all elements having the same phase to radiate in the broadside direction (90°). The desired patterns are 40° and 60° sector flat-top beams

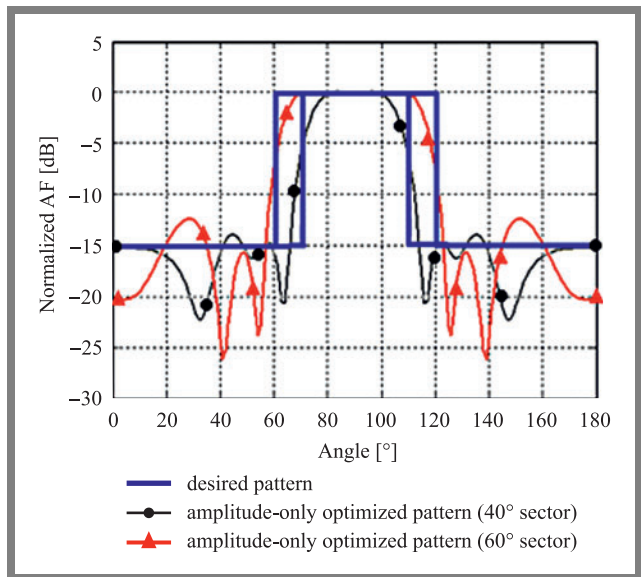


Fig. 3. 40° and 60° flat-top sector beams achieved by an amplitude-only RGA optimized array ($N = 11$).

having a maximum SLL of -15 dB. To generate such a desired pattern using the given array, Eqs. (6) and (7) are used to obtain the minimum value of the cost. The RGA-optimized simulated patterns generated using MATLAB are shown in Fig. 3 for a 40° sector and for a 60° sector, with broadside patterns for 11-element antenna arrays ($N = 11$).

For a 20-element array with $d = \lambda/2$, the excitation amplitude is optimized using RGA. The simulated radiation patterns and the desired patterns are plotted in Fig. 4.

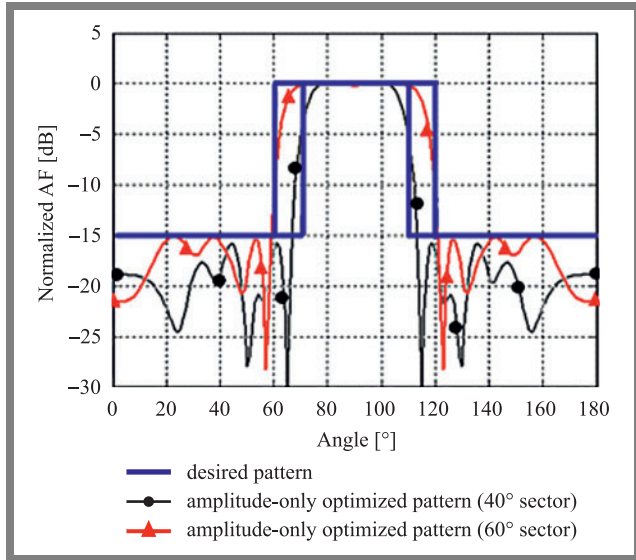


Fig. 4. 40° and 60° flat-top sector beams achieved by an amplitude-only RGA optimized array ($N = 20$).

The variations of RGA-optimized cost values with the number of iterations corresponding to Fig. 3 and Fig. 4 are plotted in Fig. 5.

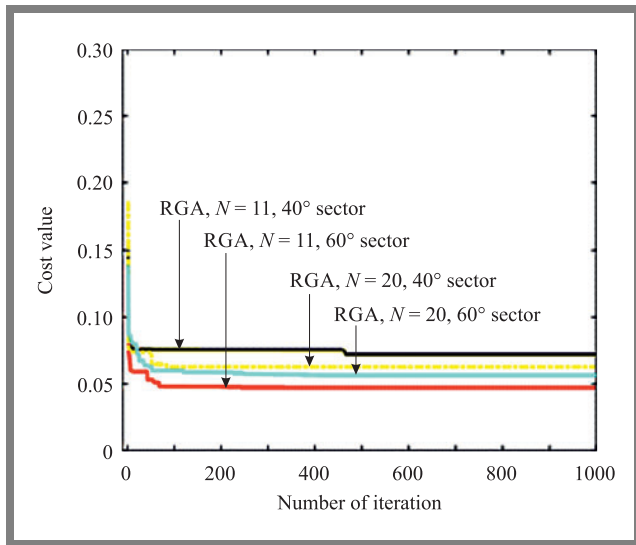


Fig. 5. Cost values for RGA amplitude-only optimized arrays.

In RGA optimized flat-top patterns in all the above cases (Fig. 3 and Fig. 4) the flat-top ripple disappears, but spillover to the neighbor sectors increases. In Fig. 3 optimized patterns do not satisfy the desired maximum SLL criteria, but

in Fig. 4 the optimized patterns satisfy the desired maximum SLL criteria.

4.3. Flat-top Beamforming by Amplitude-only Optimization Using PSO

In this section, PSO is used to obtain the optimum excitation current amplitude distribution of the individual elements to achieve the desired flat-top pattern. PSO is a continuous algorithm, beginning with a real random number, but it is simple to implement, as it does not possess the evolutionary operator. Equation (7) is the cost function used to determine the optimum excitation amplitude distribution of the array. A uniform linear array (ULA) of $N = 11$ elements is considered, having the inter-element spacing of $d = \lambda/2$. All elements have the same phase to radiate in the broadside direction (90°). The desired patterns are 40° and 60° sector flat-top

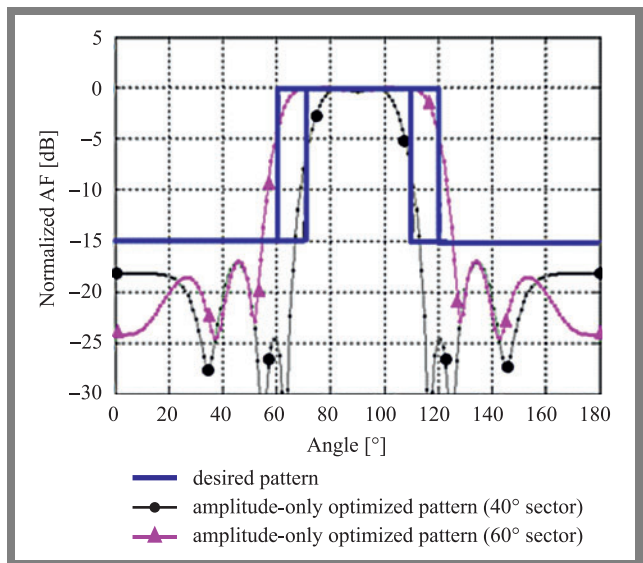


Fig. 6. 40° and 60° flat-top sector beams by amplitude-only PSO optimized array ($N = 11$).

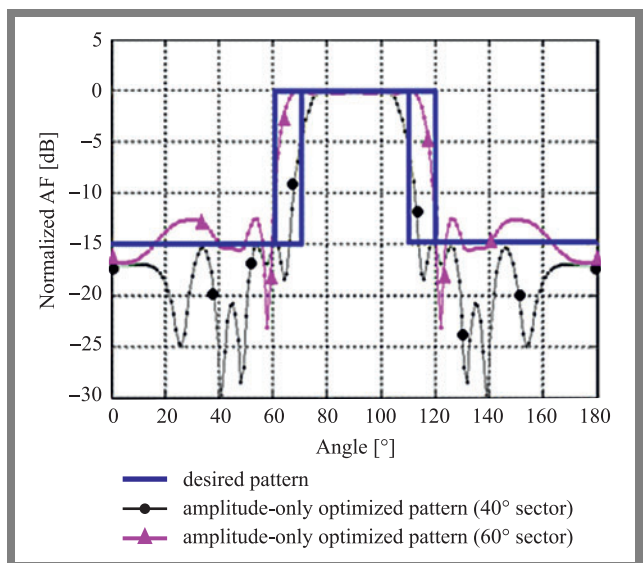


Fig. 7. 40° and 60° flat-top sector beams by amplitude-only PSO optimized array ($N = 20$).

beam generation having a maximum SLL of -15 dB. To generate such desired pattern using the given array, Eqs. (6) and (7) are used to obtain the minimum value of the cost function.

The simulated radiation pattern using PSO for the amplitude optimized array of 11 elements ($N = 11$) for 40° and 60° sector beam patterns are presented in Fig. 6.

The 40° and 60° flat-top sector beams by amplitude-only PSO optimized array for $N = 20$ are shown in Fig. 7.

The variation of PSO optimized cost values with number of iteration corresponding to Figs. 6 and 7 are plotted in Fig. 8.

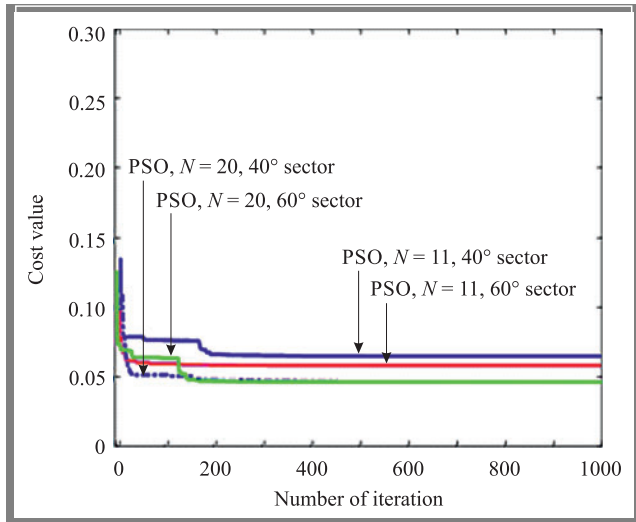


Fig. 8. Cost values for PSO amplitude-only optimized arrays.

4.4. Flat-top Beamforming by Phase-only Optimization Using RGA and PSO

In phase-only synthesis, design of an antenna array for generation of sector pattern is based on finding a common phase distribution of the array elements when the feeding current amplitude is fixed and equal value for all elements. Phase dis-

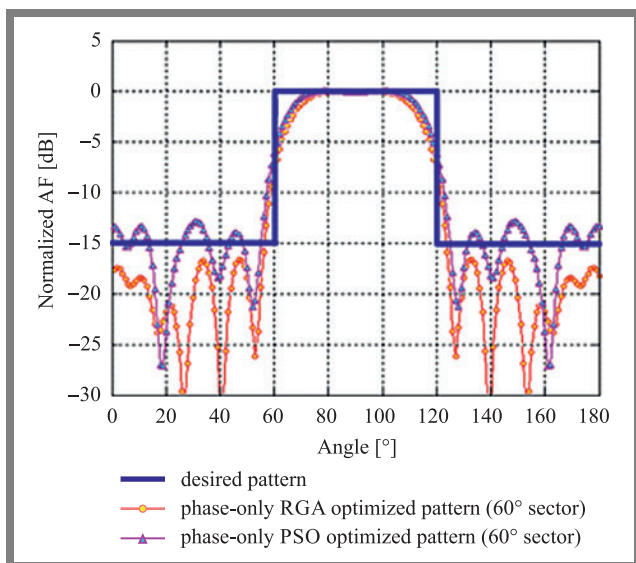


Fig. 9. 60° sector beam achieved by phase-only RGA and PSO optimized arrays ($N = 11$).

tribution can be determined by optimization of excitation current phase. RGA is applied to determine the phase distribution of each element in the array to generate a sector flat top beam pattern. Here, two ULAs of $N = 11$ and $N = 20$ are considered, which are placed along y -axis with half-wavelength array spacing. The desired pattern is a 60° sector flat top beam and SLL is -15 dB. Now a different approach is considered to generate such a desired radiation pattern using the given array by controlling the phase of excitation current while keeping a fixed excitation current amplitude. To determine the phase distribution of the array, phase-only optimization is performed using RGA on Eqs. (6) and (7) to identify the best pattern which will produce the desired pattern. Results for the phase-only RGA-optimized flat-top beam and the phase-only PSO-optimized flat-top beam are compared in Figs. 9 and 10.

The variations in RGA- and PSO-optimized cost values, observed with the increasing number of iterations corresponding to Figs. 9 and 10, are plotted in Fig. 11.

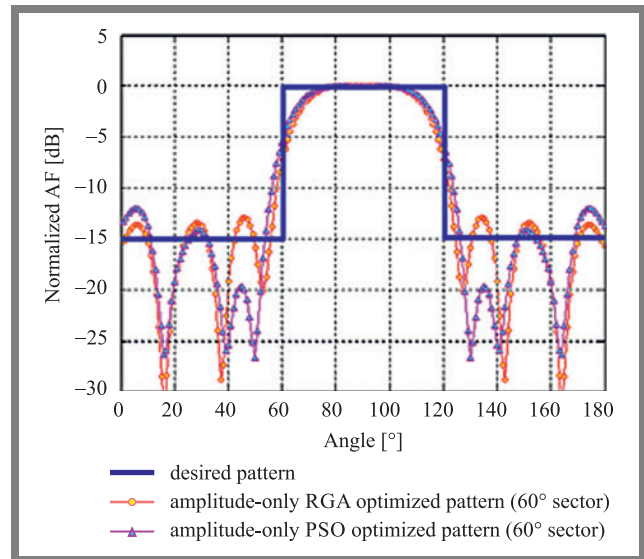


Fig. 10. 60° sector beam achieved by phase-only RGA and PSO optimized arrays ($N = 20$).

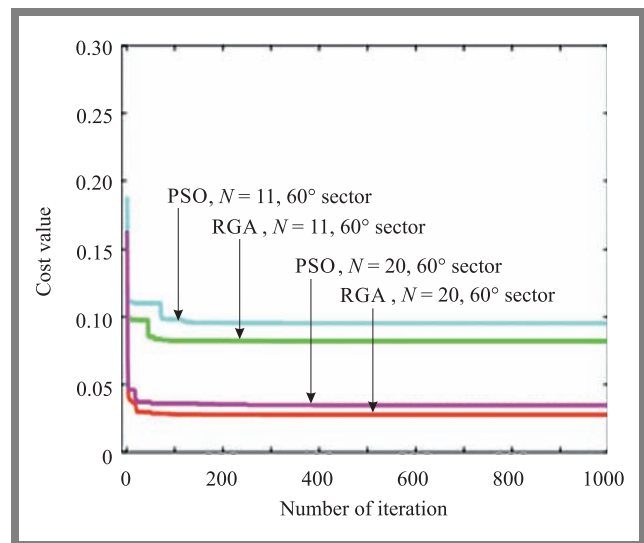


Fig. 11. Cost values for RGA and PSO phase-only optimized arrays.

Tab. 1. Phase-only optimization weights for flat-top beam formation using RGA and PSO.

Antenna elements	Sector area	Weights (RGA) phase-only	Weights (PSO) phase-only
N = 11	40°	1.88	1.50
		-1.75	-1.15
		1.53	2.50
		-1.76	-1.25
		1.14	2.05
		-1.24	-1.10
		1.38	2.40
		-1.07	-0.85
		1.65	1.31
		-1.36	0.53
		1.73	2.10
N = 20	60°	0.32	1.03
		2.98	-3.19
		-1.15	-0.64
		2.59	2.62
		-0.62	-0.55
		1.46	2.15
		0.94	0.14
		2.99	2.73
		0.96	-1.51
		-2.98	-2.76
		-0.61	-0.21
		2.99	-3.12
		0.24	-1.25
		-2.76	-2.97
		0.58	-0.36
		2.95	2.02
		-1.36	-1.87
-1.69	-1.05		
-0.75	1.03		
0.59	-1.95		

Tab. 2. Amplitude-only optimization weights for flat-top beam formation using RGA and PSO.

Antenna elements	Sector area	Weights (RGA) phase-only	Weights (PSO) phase-only
N = 11	40°	-0.15	-0.17
		-0.21	-0.19
		0.03	0.11
		0.35	0.37
		0.85	0.77
		0.91	0.90
		0.77	0.68
		0.45	0.28
		0.01	0.11
		0.30	-0.17
		0.20	0.04
	60°	0.60	0.38
		0.95	0.81
		0.70	0.72
		0.04	0.12
		-0.30	-0.30
		-0.08	-0.14
0.05	0.14		
0.20	0.12		
-0.15	-0.15		
0.00	0.02		
0.01	0.03		

N = 20	40°	0.01	0.02
		0.02	0.09
		-0.10	-0.11
		0.01	0.03
		0.03	0.02
		-0.15	-0.13
		-0.23	-0.22
		0.02	0.04
		0.35	0.37
		0.75	0.68
		0.95	0.92
	0.78	0.95	
	0.40	0.41	
	0.01	0.02	
	-0.30	-0.31	
	0.01	-0.05	
	0.04	0.06	
0.17	0.15		
0.08	0.09		
-0.08	-0.11		
60°	0.06	0.01	
	0.07	0.09	
	-0.06	-0.11	
	0.05	0.09	
	0.09	0.10	
	0.07	0.09	
	-0.29	-0.36	
	0.16	0.18	
	0.65	0.75	
	0.99	0.97	
0.79	0.85		
0.04	0.04		
-0.25	-0.26		
-0.12	-0.12		
0.08	0.09		
0.05	0.07		
-0.09	-0.10		
-0.08	-0.08		
0.03	0.04		
-0.03	0.02		

In the case of amplitude weight optimization, the excitation phase of each element is kept the same and the initial population is made up of random values whose maximum and minimum levels are restricted by the corresponding maximum and minimum values considered in the simulation. The other array synthesis method uses the phase-only optimization technique in which the excitation phase varies and amplitude is fixed at the unity level for all the elements.

The values of phase-only optimization weights using RGA and PSO are presented in Tab. 1.

The values of amplitude-only optimization weights using RGA and PSO are presented in Tab. 2.

5. Performance Comparison

The best cost values (the cost value at 1,000 iterations) for RGA and PSO optimizations are compared in Table 3.

Two ULAs with $N = 11$ and $N = 20$ are considered for synthesis purposes. Various outcomes of amplitude-only and phase-only optimized radiation patterns are presented in Tab. 4.

6. Conclusion

Tab. 3. Best cost values for RGA and PSO optimizations.

Optimization methods	Parameters	Best cost value	Computation time
RGA amplitude-only	$N = 11, 40^\circ$	0.074	937.7 s
	$N = 11, 60^\circ$	0.048	
	$N = 20, 40^\circ$	0.068	1846.3 s
	$N = 20, 60^\circ$	0.058	
PSO amplitude-only	$N = 11, 40^\circ$	0.068	734.9 s
	$N = 11, 60^\circ$	0.060	
	$N = 20, 40^\circ$	0.048	1491.1 s
	$N = 20, 60^\circ$	0.047	
RGA phase-only	$N = 11, 60^\circ$	0.080	899.2 s
	$N = 20, 60^\circ$	0.030	1763.6 s
PSO phase-only	$N = 11, 60^\circ$	0.092	719.4 s
	$N = 20, 60^\circ$	0.035	1483.3 s

While generating the desired flat-top sector radiation pattern by using the RGA optimization technique, a large population size and a large number of iterations are required in order to reach an acceptable value. Both RGA and PSO optimizations show better performance in flat-top sector beam formation in cellular networks, but the performance of PSO is better than that of RGA-based optimization. The deployment of PSO for flat-top beam generation is easier than in the case of RGA, and the computation time is also lower in PSO than in RGA. The performance of PSO is better than that of RGA if the flat-top sector beam is generated with the use of a large array. Application of these methods in multiple-beam arrays for flat-top sector beam formation in cellular networks may be the focal point of future studies.

Tab. 4. Performance comparison of amplitude-only and phase-only optimized array using RGA and PSO for a flat-top 60° sector.

Antenna elements	Parameters	Desired pattern	RGA amplitude-only optimized	PSO amplitude-only optimized	RGA phase-only optimized	PSO phase-only optimized
$N = 11$	Max. SLL [dB]	-15	-12.75	-17.18	-16.45	-13.83
	HPBW [$^\circ$]	60	52	54	51	52
	Max. ripple [dB]	0.0	0.0	0.03	0.0	0.03
$N = 20$	Max SLL [dB]	-15	-14	-13.76	-13.87	-12.52
	HPBW [$^\circ$]	60	53	56	50	52
	Max. ripple [dB]	0.0	0	0.04	0	0.04

Tab. 5. Performance of optimization methods used for flat-top beam formation in a cellular network.

Parameter	Amplitude-only optimization		Phase-only optimization	
	RGA	PSO	RGA	PSO
Implementation for flat-top beam generation	More difficult	Less effort	More difficult	Less effort
Ripple of flat-top beam	Almost zero	Very low but not almost zero like RGA	Almost zero	Very low but not almost zero like RGA
Coverage of desired sector by HPBW	Not good	Better than RGA	Not good	Better than RGA
Spillover of beam to the neighboring sectors (beyond coverage area and below -10 dB level)	Lesser than PSO	Low	Lesser than PSO	Low
Performance for a large array	Good	Better than RGA	Good	Better than RGA
Performance for a small array	Good	Good	Good	Good
Simulation time	High	Low	High	Low

References

- [1] J. Thronton, *et al.*, "Effect of antenna beam pattern and layout on cellular performance in high altitude platform communications", *Wireless Personal Communications*, vol. 35, no. 1–2, pp. 35–51, 2005 (DOI: 10.1007/s11277-005-8738-6).
- [2] X. Cai and W. Geyi, "An optimization method for the synthesis of flat-top radiation patterns in the near- and far-field regions", *IEEE Transactions Antennas & Propagations*, vol. 67, no. 2, pp. 980–987, 2018 (DOI: 10.1109/TAP.2018.2882653).
- [3] N. Vegesna, G. Yamuna, and S.K. Terlapu, "Design of linear array for shaped beams using enhanced flower pollination optimization algorithm", *Soft Computing*, vol. 38, 2022 (DOI: 10.1007/s00500-022-07146-0).

- [4] R. Wongsan, P. Krachodnok, and P. Kamphikul, "A sector antenna for mobile base station using MSA array with curved woodpile EBG", *Open Journal of Antennas and Propagation*, vol. 2, no. 1, pp. 1–8, 2014 (DOI: 10.4236/ojapr.2014.21001).
- [5] A. Sabharwal, D. Avidor, and L. Potter, "Sector beam synthesis for cellular systems using phased antenna arrays", *IEEE Transactions Vehicular Technology*, vol. 49, no. 5, pp. 1784–1792, 2000 (DOI: 10.1109/25.892583).
- [6] S.K. Goudos, "Shaped beam pattern synthesis of antenna arrays using composite differential evolution with eigenvector-based crossover operator", *International Journal of Antennas and Propagation*, pp. 1–10, 2015 (DOI: 10.1155/2015/295012).
- [7] S. Dai, M. Li, Q. Abbasi, and M. Imran, "A zero placement algorithm for synthesis of flat top beam pattern with low sidelobe level", *IEEE Access*, vol. 8, pp. 225935–225944, 2020 (DOI: 10.1109/ACCESS.2020.3045287).
- [8] S.A.M. Soliman, E.M. Eldesouki, and A.M. Attiya, "Analysis and design of an X-band reflectarray antenna for remote sensing satellite system", *Sensors*, vol. 22, no. 3, pp. 1166, 2022 (DOI: 10.3390/s22031166).
- [9] H-J. Zhou, Y-H. Huang, B-H. Sun, and Q-Z. Liu, "Design and realization of a flat-top shaped-beam antenna array", *Progress In Electromagnetics Research Letters*, vol. 5, pp. 159–166, 2008 (DOI: 10.2528/PIERL08111911).
- [10] R.S. Sohal, V. Grewal, and J. Kaur, "Analysis of different antenna array configurations in massive MIMO cellular system for line of sight", *Wireless Personal Communications*, vol. 120, no. 3, pp. 2029–2041, 2021 (DOI: 10.1007/s11277-021-08697-5).
- [11] K. Deb and A. Kumar, "Real-coded genetic algorithms with simulated binary crossover: studies on multimodal and multiobjective problems", *Complex Systems*, vol. 9, no. 6, pp. 431–454 1995 (<https://content.wolfram.com/uploads/sites/13/2018/02/09-6-1.pdf>).
- [12] R.L. Haupt and S.E. Haupt, "Practical Genetic Algorithms" Wiley, 2004 (DOI: DOI:10.1002/0471671746).
- [13] J. Kennedy and R. Eberhart, "Particle swarm optimization", *International Conference on Neural Networks, IEEE Xplore*, pp. 1942–1948, 1995 (DOI: 10.1109/ICNN.1995.488968).
- [14] C.A. Balanis, "Antenna Theory – Analysis and Design", Wiley-Interscience, 2005 (ISBN: 9780471667827).
- [15] M.M. Khodier and C.G. Christodoulou, "Side lobe level and null control using particle swarm optimization", *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 8, pp. 2674–2679, 2005 (DOI: 10.1109/TAP.2005.851762).



Pampa Nandi received her M.Tech. in Electronics and Telecommunication Engineering from KIIT University, Bhubaneswar, India, in 2010. She obtained her Ph.D. from the same University in 2017. She is an Assistant Professor at the KIIT Polytechnic, operating under the auspices of KIIT University. Her research interests focus on phased array antennas, thinned array antennas, and various optimization

problems related to antenna arrays.

E-mail: pampanandi@yahoo.com

School of Electronics Engineering, Kalinga Institute of Industrial Technology University Bhubaneswar, Odisha, India



Jibendu Sekhar Roy is a full professor and Ex-associate Dean, School of Electronics Engineering, KIIT University, Bhubaneswar, Odisha. Between 1998 and 2009 he was a professor at the ECE department, BIT, Mesra, Ranchi. He received his M.Sc. in Physics (Radio Physics & Electronics)

from the University of Burdwan, West Bengal, India. He received his Ph.D. degree from Jadavpur University, Calcutta and was a CSIR Research Associate. He was a Post-Doctoral Research Associate at CNRS, France. His areas of research cover microstrip antennas, antenna arrays, signal processing and wireless communication.

E-mail: drjsroy@rediffmail.com

School of Electronics Engineering, Kalinga Institute of Industrial Technology University Bhubaneswar, Odisha, India

An Extended Version of the Proportional Adaptive Algorithm Based on Kernel Methods for Channel Identification with Binary Measurements

Rachid Fateh, Anouar Darif, and Said Safi

Laboratory of Innovation in Mathematics, Applications and Information Technologies (LIMATI), Sultan Moulay Slimane University, Beni Mellal, Morocco

<https://doi.org/10.26636/jtit.2022.161122>

Abstract — In recent years, kernel methods have provided an important alternative solution, as they offer a simple way of expanding linear algorithms to cover the non-linear mode as well. In this paper, we propose a novel recursive kernel approach allowing to identify the finite impulse response (FIR) in non-linear systems, with binary value output observations. This approach employs a kernel function to perform implicit data mapping. The transformation is performed by changing the basis of the data in a high-dimensional feature space in which the relations between the different variables become linearized. To assess the performance of the proposed approach, we have compared it with two other algorithms, such as proportionate normalized least-mean-square (PNLMS) and improved PNLMS (IPNLMS). For this purpose, we used three measurable frequency-selective fading radio channels, known as the broadband radio access network (BRAN C, BRAN D, and BRAN E), which are standardized by the European Telecommunications Standards Institute (ETSI), and one theoretical frequency selective channel, known as the Macchi's channel. Simulation results show that the proposed algorithm offers better results, even in high noise environments, and generates a lower mean square error (MSE) compared with PNLMS and IPNLMS.

Keywords — binary measurement, BRAN channel identification, kernel methods, PNLMS, phase estimation

1. Introduction

Numerous measurement-related challenges faced in digital communication systems have been effectively resolved with the help of adaptive filtering algorithms dealing with signal enhancement, acoustic noise cancellation, echo cancellation, channel estimation, blind channel equalization, and system identification [1]–[4]. System identification is of crucial interest in the field of automatic control [5], used to determine the most adequate mathematical model based on the inputs, outputs, and perturbations of a simulated real system. The least mean squares (LMS) algorithm [6] and its variants – normalized LMS (NLMS) [7] and recursive least squares (RLS) – [8] are the most popular methods employed for identification of linear systems due to the statistical conceptual clarity of the mean square error cost function, simple mathematical

operations required, stability, and easy implementation [1]. Unfortunately, this algorithm is not valid for sparse system identification.

An attempt was made to overcome this limitation by proposing a novel adaptive technique for model system identification with sparse impulse response using an adaptive delay filter [9]. After that, Duttweiler introduced the concept of updating the proportional NLMS (PNLMS) [10] algorithm for network echo cancellation applications. Unfortunately, its convergence rate begins to slow down considerably after the fast initial period, finally becoming even slower than in the case of NLMS. To resolve this drawback, several versions of the PNLMS algorithm were developed. The examples include the well-known improved PNLMS (IPNLMS) algorithm [11], which uses a controlled mixture of PNLMS and NLMS algorithms, the μ -law PNLMS (MPNLMS) algorithm [12], an improved IPNLMS algorithm [13], an improved μ -law proportionate NLMS algorithm [14], l_0 -LMS [15], and the evaluation of block-sparse systems with an improved μ -law PNLMS algorithm (BS-MPNLMS) [16].

In addition, to exploit the sparsity characteristics of the estimated systems, certain subtypes of these techniques operating based on core idea presented above have also been introduced [17]–[19] with zero-attractors. In the field of system identification, the requirement for a high degree of precision in large complex systems has driven the need for good models that are capable of representing the non-linear structure of numerous real systems [20], [21]. For example, the Hammerstein model has been employed in diverse non-linear system applications and many related research works exist – see, for instance, [22]–[30].

Non-linear system identification continues to be a hot topic in the scientific community [31]. Volterra filters [32]–[34] and neural networks [35] are two of the most well-known techniques gaining a lot of attention. Each technique has its own set of benefits and drawbacks. For example, in the case of Volterra filters, the number of parameters to be estimated is determined by the filter order and its complexity. This allows to incorporate a high degree of complexity as far as the

range of parameters is concerned. The weak point of neural networks lies in the choice of the parametric form, as this can often only be performed in a more or less arbitrary way. A false decision may degrade performance.

The development of Kernel methods [36]–[38] has been speeding up in recent years, as they serve as an important tool for the advancement of new technologies, especially in terms of the reduction of computational time required to handle difficult tasks [39]. These techniques greatly increase the accuracy of processing thanks to their ability to detect any existing commonalities in the treated information. They depend on a key principle known as the kernel trick, which was first applied to the support vector machine (SVM) [40], [41], and was soon after used to recast many classical linear methods in high dimensional the reproducing kernel Hilbert space (RKHS) [42]–[44], and was then reformulated as an inner product to yield more powerful non-linear extensions [38]. The kernel trick makes it possible to attribute the non-linear nature to many previously classical linear techniques and, with no restraint, it can only be represented in the scalar products form of data measurement.

Up to date, a number of kernel adaptive-filtering (KAF) algorithms have been suggested in the research literature. The examples include, inter alia, kernel least mean square (KLMS) [45], kernel recursive least square (KRLS) [46] and kernel affine projection algorithm (KAPA) [47] used in the field of non-linear signal processing [48]. In order to achieve the highest level of performance of fundamental kernel algorithm varieties, different subtypes of these categories were identified, including quantized kernel least mean square (QKLMS) [49], quantized kernel recursive least square (QKRLS) [50], regularized kernel least mean square based on multiple time delay feedback (RKLMS-MDF) [51], kernel least mean square with adaptive kernel size (KLMS-AKS) [52], extended kernel recursive least square (Ex-KRLS) [53] and reduced kernel recursive least square (RKRLS) [54] that are used for channel identification [24]–[29], [55] and equalization in non-linear systems.

In this paper, we propose a novel recursive algorithm that is based on the positive definite kernel function, with our primary focus being on the improved proportionate normalized least mean square (IPNLMS) variety developed in [11]. As far as we are aware, the application of kernel-based adaptive nonlinear system identification methods with binary-valued output observations of the IPNLMS has not been studied yet. For validity and test purposes, the proposed algorithm is compared with the proportional normalized least mean square (PNLMS) algorithm and with its improved version (IPNLMS), where the goal is to identify impulse response parameters of Macchi and ETSI BRAN channels. The relation of the proposed algorithm with other algorithms described in the literature will be demonstrated based on two examples. First, an example using the Hammerstein system, in which the input sequence is randomly generated with a uniform distribution, will highlight how the proposed algorithm is capable of estimating, with good accuracy, the impulse response parameters of the practical frequency selective fading

channel (BRAN) and the theoretical frequency selective channel (Macchi). Second, it will be shown how the proposed algorithm is capable of converging faster and yielding a smaller estimation error than both PNLMS and IPNLMS.

This paper is arranged as follows. In Section 2 we introduce the architecture of a non-linear system (Hammerstein model) identification problem with binary-valued output observations and noise. In Section 3, we give an overview of some fundamental notations of the kernel methods, with that overview followed by a description of PNLMS, IPNLMS, and kernel extended IPNLMS algorithms. The effectiveness of the proposed recursive kernel algorithm is discussed based on some simulation results in Section 4. Finally, Section 5 concludes this paper.

2. Problem Statement and Assumptions

Let us consider the single-input single-output (SISO) Hammerstein model presented in Fig. 1. It is made up of a non-linear static function followed by a known-order finite impulse response (FIR).

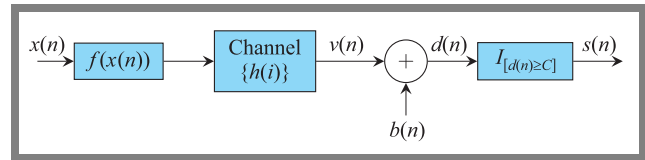


Fig. 1. Block diagram of a Hammerstein model with binary outputs and noises.

From Fig. 1, the output of the desired system is given by:

$$\begin{cases} v(n) = \sum_{i=0}^{L-1} h(i)f(x(n-i)), \\ d(n) = v(n) + b(n), \quad n = 0, 1, 2, \dots, N \end{cases}, \quad (1)$$

where $x(n)$ represents the input signal, $d(n)$ the output, $h(i)_{(i=0,1,\dots,L-1)}$, L the coefficients of the finite impulse response filter, $f(\cdot)$ is the nonlinearity, and $b(n)$ is the measurement noise.

A binary-valued sensor $I_{[\cdot]}$ with a fixed threshold $C \in \mathbb{R}$ can be used to measure the system's output $d(n)$. The output with a binary value $s(n)$ can be represented by the following formula:

$$s(k) = I_{|d(n)| \geq C} = \begin{cases} 1 & \text{if } d(n) \geq C \\ -1 & \text{otherwise} \end{cases}. \quad (2)$$

The following are the main assumptions that were made for the system model:

- input sequence $\{x(n)\}$, is i.i.d. (independent and identically distributed) bounded random process with zero mean,
- additive noise $\{b(n)\}$ is suggested, Gaussian and independent of $\{x(n)\}$ (bounded) and $\{d(n)\}$ (bounded),
- let $f(\cdot)$ be invertible and continuous for any finite x ,
- the system model does not include any delays, i.e. $h(0) \neq 0$,
- C value is available (known).

The above-mentioned assumptions are made in order to facilitate the analysis of the system and to obtain the best results in terms of the mean square error and the channel identification framework considered. The main purpose of this paper is to construct a recursive identification algorithm for finite impulse response (FIR) systems based on positive definite kernels and binary-valued observations $s(n)$, in order to recursively estimate the channel's parameters.

3. Proposed Adaptive Filtering Algorithm

In this section, we start by presenting the general idea behind kernel methods. We will define what a kernel is, specifying its properties and those of the kernel spaces. Next, we describe the adaptive algorithms used to identify channel impulse responses, i.e. the proportional normalized LMS algorithm (PNLMS) and the improved PNLMS algorithm (IPNLMS). This derivation order is equally the historical arrangement of the algorithms that were previously extended according to [10], [11]. Then, the kernel methods are incorporated into the IPNLMS algorithm strategy in order to produce a kernelized version of the improved proportionate normalized LMS algorithm based on binary measurements.

Kernel methods can be used to solve non-linear adaptive filtering problems in high dimensional spaces. The problem of dimensionality, referring to the number of parameters to be estimated, is then reduced to the amount of learning data available. A fundamental characteristic of kernel methods is that the resulting model is a linear combination of kernel functions whose order is identical to the size of the learning data, where all sources of input information $\{x(i)\}_{i=1}^N \in \mathcal{X}$ were mapped (implicit) into a high dimensional space \mathcal{H} (an inner product space) by taking advantage of the idea that the Mercer kernel function could be used to express an inner product in Hilbert spaces. Based on Mercer's theorem, the mapping $\Psi(\cdot)$ that was introduced by means of $\kappa(x(i), x(j))$ will be expressed by the following relationship [36], [63]:

$$\kappa(x(i), x(j)) = \langle \Psi(x(i)), \Psi(x(j)) \rangle_{\mathcal{H}}, \quad \forall x(i), x(j) \in \mathcal{X}, \quad (3)$$

where $\kappa(\cdot, \cdot)$ is a kernel function and $\Psi(\cdot)$ mapped \mathcal{X} to a space \mathcal{H} with an inner product $\langle \cdot, \cdot \rangle$. Commonly, dimension of (\mathcal{X}) is much smaller than the dimension of (\mathcal{H}) .

The block diagram shown in Fig. 2 illustrates an adaptive kernel-based channel estimation using the proposed algorithm, where $e(n)$ represents the estimation error and $y(n)$ is the estimated desired response. The learning procedure is conducted in two distinct phases at each time n (Fig. 2):

- 1) Initially, using the Hammerstein system (HS) with binary-valued output observations and noise, we obtain the binary output $s(n)$.
- 2) During the next phase, based on the transformation of the measured data into non-linear spaces (RKHSs) employing a Mercer kernel κ , channel coefficients $\theta(n)$ are adjusted according to the functional cost minimization principle.

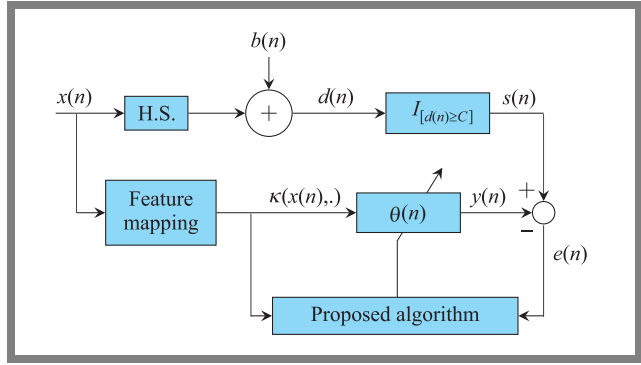


Fig. 2. Block schematic of kernel adaptive filter.

Let us start with some necessary preliminaries that need to be employed in the proposed algorithm in order to successfully determine the existing functional space \mathcal{H} .

Definition 1 positive definite kernel. A kernel is said to be positive definite, if it satisfies the following condition for each input data point $\{x(i)\}_{i=1}^N \in \mathcal{X}$:

$$\sum_{i,j=1}^N \alpha_i \alpha_j \kappa(x(i), x(j)) \geq 0, \quad (4)$$

for all $N \in \mathbb{N}$, $\{x(1), \dots, x(N)\} \subseteq \mathcal{X}$ and $\{\alpha_1, \dots, \alpha_N\} \subseteq \mathbb{R}$.

In particular, if $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is positive definite, then it can be expressed as an inner product in the feature space \mathcal{H} , where the data are projected. On the other hand, if we define a correspondence between the input data and a vector space, then the inner product in this vector space will be a positive definite kernel.

According to the Mercer theorem [36], [63], any kernel $\kappa(x(i), x(j))$ can be redefined as follows:

$$\kappa(x(i), x(j)) = \sum_{i=1}^{\infty} \zeta_i \Psi_i(x(i)) \Psi_i(x(j)), \quad (5)$$

where ζ_i and Ψ_i , $i = 1, 2, \dots$, denote the non-negative eigenvalues and the eigenfunctions, respectively.

Mapping Ψ_i in the reproducing kernel Hilbert space can be created as:

$$\Psi(x) = \left[\sqrt{\zeta_1} \Psi_1(x), \sqrt{\zeta_2} \Psi_2(x), \dots \right]^T. \quad (6)$$

Definition 2 reproducing kernel Hilbert spaces. Let \mathcal{H} denote a Hilbert space of real functions defined on an indexed set \mathcal{X} :

$$\mathcal{H} = \left\{ \sum_{j=1}^n \alpha_j \kappa(x(j), \cdot) : n \in \mathbb{N}, x(j) \in \mathcal{X}, \right. \\ \left. \alpha_j \in \mathbb{R}, j = 1, \dots, n \right\}. \quad (7)$$

\mathcal{H} is considered to be a reproducing kernel Hilbert space with an inner product noted $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and the norm $\|f\|_{\mathcal{H}} = \sqrt{\langle f, f \rangle_{\mathcal{H}}}$ if there exists a function $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ that has the following two properties:

- 1) for any element $x \in \mathcal{X}$, $\kappa(x, \cdot)$ belongs to \mathcal{H} ,

- 2) function κ is a reproducing kernel function, i.e. for any function $f \in \mathcal{H}$, we have: $\langle f, \kappa(x, \cdot) \rangle_{\mathcal{H}} = \sum_{j=1}^n \alpha_j \kappa(x(j), x) = f(x)$.

3.1. Derivation of the PNLMS Algorithm

The PNLMS algorithm was first proposed by assigning a step parameter to each coefficient using a diagonal step control matrix $G(n) \in \mathbb{R}^{(L) \times (L)}$ [10]. This algorithm is capable of exploiting low impulse response density to achieve a better adaptation than that observed in the case of the classical NLMS algorithm. The PNLMS algorithm needs more operations than the NLMS algorithm but has the benefit of converging faster than the latter. The PNLMS algorithm's practical update equations are given by:

$$e(n) = s(n) - \theta^\top(n-1)\mathbf{x}(n), \quad (8)$$

$$\mathbf{D}(n-1) = \text{diag}(d_0(n-1), d_1(n-1), \dots, d_{L-1}(n-1)), \quad (9)$$

$$\theta(n) = \theta(n-1) + \frac{\mu \mathbf{D}(n-1) \mathbf{x}(n) e(n)}{\delta_{\text{PNLMS}} + \mathbf{x}^\top(n) \mathbf{D}(n-1) \mathbf{x}(n)}, \quad (10)$$

where $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^\top$ represents the input signal, where the superscript $(\cdot)^\top$ is the transpose operator, $e(n)$ is the estimation error, $\mu \in \mathbb{R}_+^*$ is the fixed step-size, $d_l(n) \in \mathbb{R}_+^*$, and δ_{PNLMS} is a regularization parameter:

$$\delta_{\text{PNLMS}} = \frac{\delta_{\text{NLMS}}}{L}.$$

The original definition of the diagonal matrix element $\mathbf{D}(n)$ is:

$$d_l(n) = \frac{k_l(n)}{\frac{1}{L} \sum_{i=0}^{L-1} k_i(n)}, \quad l = 0, 1, \dots, L-1, \quad (11)$$

with

$$k_l(n) = \max\{|\theta_l(n)|, \rho \max\{\delta_p, |\theta_0(n)|, \dots, |\theta_{L-1}(n)|\}\}. \quad (12)$$

Parameters δ_p and ρ are used to protect $\theta_l(n)$ from stalling during the initialization step. The typical value of δ_p is equal to 0.01 and ρ ranges from $\frac{1}{L}$ to $\frac{5}{L}$.

3.2. Derivation of the IPNLMS Algorithm

Convergence speed of the PNLMS algorithm degrades significantly when dealing with non-sparse impulse responses. Improved PNLMS (IPNLMS) is proposed to avoid degradation in a scenario in which the impulse underlying the response is non-sparse. In the improved PNLMS algorithm, the diagonal element of $\mathbf{D}(n)$ is:

$$d_l(n) = \frac{1-\alpha}{2L} + \frac{|\theta_l(n)|(1+\alpha)}{2 \sum_{i=0}^{L-1} |\theta_i(n)| + \delta_{\text{IPNLMS}}}, \quad (13)$$

where $\alpha \in [-1, 1]$ and $\delta_{\text{IPNLMS}} = \frac{1-\alpha}{2L} \delta_{\text{NLMS}}$ is a one of the small positive numbers in order to prevent dividing by zero. We will utilize the kernel-based method to extend the improved PNLMS algorithm in the manner described in the next subsection.

3.3. Projection Over Kernel Methods

The proposed algorithm is presented in this section. The main idea is to operate the improved proportionate NLMS algorithm in the Gaussian kernel feature space that is linked to a reproducing kernel κ (continuous, normalized and symmetric), using the feature map $\Psi(\cdot)$ that enables us to transform the sample sequence as:

$$\Psi : \mathcal{X} \longrightarrow \mathcal{H} \\ x(i) \longrightarrow \kappa(x(i), \cdot), \quad 0 \leq i \leq N. \quad (14)$$

In order to generate the infinite-dimensional space model of the reproducing kernel, there exist various types of kernels such as sigmoid and radial Gaussian kernels defined, respectively, by:

$$\kappa_{a,c}^{\text{sig}}(x(i), x(j)) = \text{tanh}(a(x(i), x(j)) + c), \quad \forall a, c \in \mathbb{R}, \quad (15)$$

$$\kappa_\sigma^G(x(i), x(j)) = e^{-\frac{\|x(i) - x(j)\|^2}{2\sigma^2}}, \quad (16) \\ \forall (x(i), x(j)) \in \mathcal{X}^2,$$

where $\sigma > 0$ is the bandwidth of the kernel used to specify the form of the kernel function, a is the sigmoid scale, and c is the bias.

Throughout the remainder of the paper, we will use the Gaussian radial basis function (RBF) kernel, which the favorite option mostly thanks to its perfect approximation character and its numerical stability. In [57], one can find a more comprehensive list of Mercer kernels. $\Psi(\cdot)$ mapping generated by this type of kernel is a bit special. In fact, a data item will be mapped onto a Gaussian function that represents the data's similarity to all data in \mathcal{X} . Fig. 3 represents the application of the Gaussian radial-basis function kernel to the two pieces of data $x(i)$ and $x(j)$.

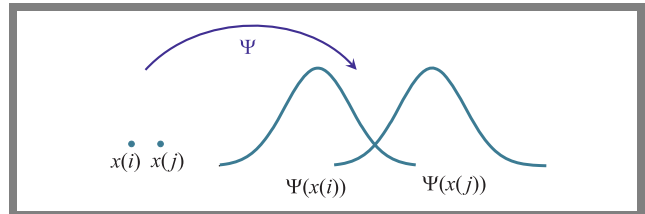


Fig. 3. Definition of characteristic map.

The four steps that make up the proposed identification algorithm are:

- 1) In the initial step, a transformation of the measured data inputs from the space \mathcal{X} into a high dimensional space (feature space \mathcal{H}) is realized to produce the following input data:

$$\{(\Psi(x(1)), s(1)), (\Psi(x(2)), s(2)), \dots, (\Psi(x(n)), s(n)), \dots\}. \quad (17)$$

- 2) In the second step, by applying the methodology of the improved proportionate NLMS algorithm to the input data sequence described in Eq. (17), we can minimize the cost function by:

$$E[|s(k) - \langle (\Psi(x(k))), \theta \rangle_{\mathcal{H}}|^2],$$

where θ denotes the weight vector in the feature space \mathcal{H} .

- 3) Next, we are proceeding directly in the feature space \mathcal{H} , under the assumption that our data has already been successfully modeled in the RKHS by means of the Ψ mapping function, i.e.:

$$\mathcal{X} \ni x \longrightarrow \Psi(x(n)) := \kappa(x, \cdot) \in \mathcal{H}. \quad (18)$$

- 4) The estimate of $\theta(n)$ is produced and is noted $\hat{\theta}(n)$:

$$\hat{\theta}(n) = \hat{\theta}(n-1) + \frac{\mu \mathbf{D}(n-1) \kappa(x(n), \cdot) e(n)}{\delta_{\text{KE-IPNLMS}} + \kappa(x(n), \cdot)^\top \mathbf{D}(n-1) \kappa(x(n), \cdot)}, \quad (19)$$

$$\mathbf{D}(n-1) = \text{diag}(d_0(n-1), d_1(n-1), \dots, d_{L-1}(n-1)), \quad (20)$$

where:

$$d_l(n) = \frac{1-\alpha}{2L} + \frac{|\hat{\theta}_l(n)|(1+\alpha)}{2 \sum_{i=0}^{L-1} |\hat{\theta}_i(n)| + \delta_{\text{KEIPNLMS}}}, \quad (21)$$

where α is the adjusting parameter, and $\delta_{\text{KE-IPNLMS}}$ is a small value used to avoid a denominator equaling zero.

The proposed identification algorithm update equations in kernel Hilbert space is summarized as Algorithm 1.

Algorithm 1. Kernel extended IPNLMS algorithm for channel identification.

Input: samples $\{x(n), s(n)\}$, $n = 1, 2, \dots, N$

Initialization: channel parameter $\theta(0)$ with zeros, adjusting parameter α , kernel bandwidth σ , and threshold C

Computation:

while $\{x(n), s(n)\}_{n=1}^N$ available **do**

1. compute $y(n)$ as: $y(n) = \kappa(x(n), \cdot)^\top \hat{\theta}(n-1)$
2. compute the prediction error as: $e(n) = s(n) - y(n)$
3. update weight vector using Eq. (19)
4. compute the diagonal matrix \mathbf{D} using Eq. (20)

end while

4. Results of Numerical Simulations

The main aim of this section was to investigate the effectiveness of the proposed kernel extended IPNLMS algorithm in terms of channel impulse response identification and to compare it with that of PNLMS and IPNLMS algorithms. Performance was measured using mean squares error (MSE) in decibels, expressed as:

$$\text{MSE} = 10 \log \left[\frac{1}{N} \sum_{n=1}^N (s(n) - y(n))^2 \right], \quad (22)$$

where N represents the data length, $s(n)$ is the binary output and $y(n)$ is the estimated desired response. The procedure involved 50 runs of Monte Carlo experiments to reduce measurement uncertainty. We used two models for the linear part: the Macchi channel and the ETSI BRAN channel to simulate the Hammerstein model. A typical non-linear part function is the hyperbolic function $\tanh(x)$, defined by:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (23)$$

The parameter settings selected for the simulations are: threshold is $C = 0.5$, step-size parameter for all the algorithms is $\mu = 0.05$, regularization parameter $\delta_{\text{NLMS}} = 0.01$, adjusting parameter $\alpha = -0.75$, kernel width $\sigma = 0.2$, data length $N = 2^{10}$, and $\text{SNR} = 16$ dB. Note that when we modify one of these simulation parameters, the others remain constant. The simulations are performed using Matlab software and are conducted for various SNRs defined as follows:

$$\text{SNR} = 10 \log \left[\frac{E(v^2(n))}{E(b^2(n))} \right], \quad (24)$$

where $E[\cdot]$ is the mathematical expectation.

4.1. Macchi Channel

To investigate the theoretical performance of the presented approach, we have relied on the Macchi channel. The impulse response of this channel is defined by vector $\mathbf{H} = [h(0), \dots, h(L-1)]^\top$ of coefficients $h(i)$ [64]:

$$\mathbf{H} = [0.8264, -0.1653, 0.8512, 0.1636, 0.81]^\top.$$

Figure 4 shows the characteristics of this channel. It has four zeros, two of them are outside of the unit circle, which implies that the channel is of the non-minimum phase. This channel's amplitude response is quite deep fading, and its phase response is far from linear.

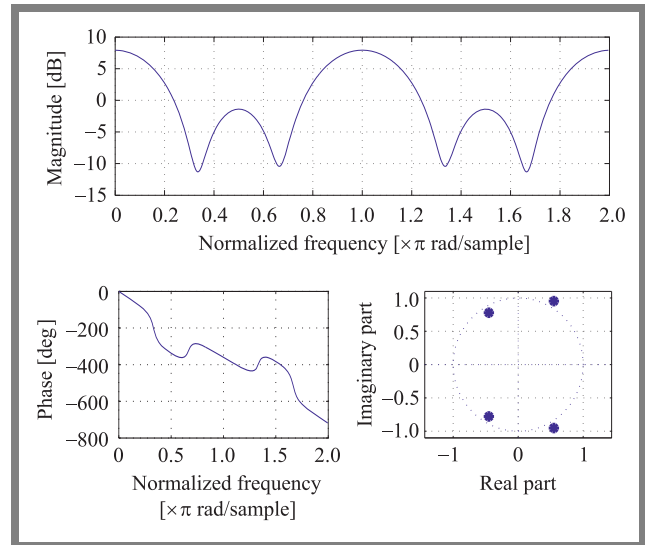


Fig. 4. Macchi channel.

Figure 5 shows the estimation of the magnitude and phase of the Macchi channel impulse response parameters, using the three algorithms for a data length of $N = 2^{10}$ and $\text{SNR} = 16$ dB.

We observe that the magnitude and phase estimated with the use of the proposed KE-IPNLMS algorithm follow the true model in perfect agreement with the measured data. But for both other algorithms (PNLMS and IPNLMS), we can see a significant difference between the measured and estimated parameters.

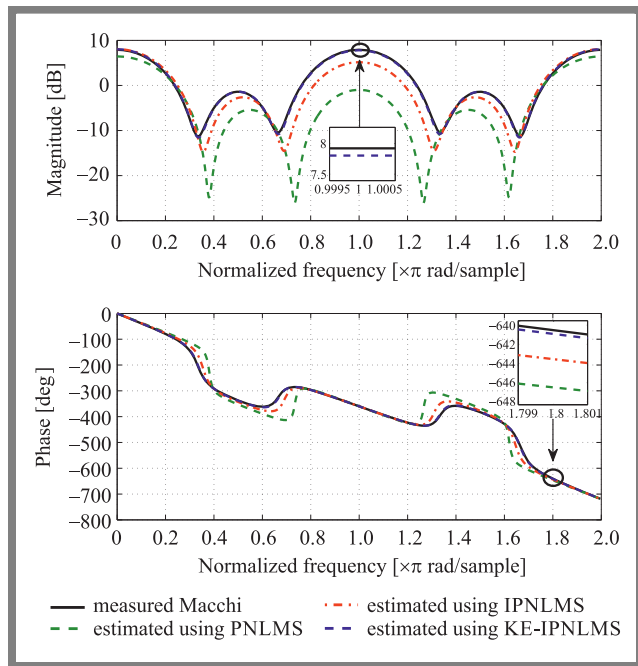


Fig. 5. Macchi channel magnitude and phase estimation for $N = 2^{10}$ and $SNR = 16$ dB.

4.2. ETSI BRAN Channel

The robustness of identification algorithms cannot be fully evaluated by simulating them on theoretical channels. As a result, we looked into mobile radio channel models. We focused on three models (ETSI BRAN C, BRAN D, and BRAN E) that represent fading radio channels. The associated model data are measured for 4G systems [65], [66]. The impulse response parameters of the ETSI BRAN radio channel are described by:

$$h(n) = \sum_{i=0}^{L-1} M_i \delta(n - \tau_i), \quad p = 18, \quad (25)$$

where L is the number of paths present, $M_i \in N(0, 1)$ is the magnitude of path i , τ_i is its delay time and $\delta(n)$ is the Dirac function. The magnitudes and time delays of 18 targets of the BRAN C, D and E channels are represented in Tables 1, 2 and 3, respectively.

Tab. 1. Delay and magnitudes of 18 targets of BRAN C radio channel.

Delay τ_i [ns]	Magnitude M_i [dB]	Delay τ_i [ns]	Magnitude M_i [dB]
0	-3.3	230	-3.0
10	-3.6	280	-4.4
20	-3.9	330	-5.9
30	-4.2	400	-5.3
50	0	490	-7.9
80	-0.9	600	-9.7
110	-1.7	730	-13.2
140	-2.6	880	-16.3
180	-1.5	1050	-21.2

Tab. 2. Delay and magnitudes of 18 targets of BRAN D radio channel.

Delay τ_i [ns]	Magnitude M_i [dB]	Delay τ_i [ns]	Magnitude M_i [dB]
0	0	230	-9.4
10	-10	280	-10.8
20	-10.3	330	-12.3
30	-10.6	400	-11.7
50	-6.4	490	-14.3
80	-7.2	600	-15.8
110	-8.1	730	-19.6
140	-9.0	880	-22.7
180	-7.9	1050	-27.6

Tab. 3. Delay and magnitudes of 18 targets of BRAN E radio channel.

Delay τ_i [ns]	Magnitude M_i [dB]	Delay τ_i [ns]	Magnitude M_i [dB]
0	-4.9	320	0
10	-5.1	430	-1.9
20	-5.2	560	-2.8
40	-0.8	710	-5.4
70	-1.3	880	-7.3
100	-1.9	1070	-10.6
140	-0.3	1280	-13.4
190	-1.2	1510	-17.4
240	-2.1	1760	-20.9

Figures 6, 7 and 8 illustrate BRAN C, D and E channel zeros, respectively.

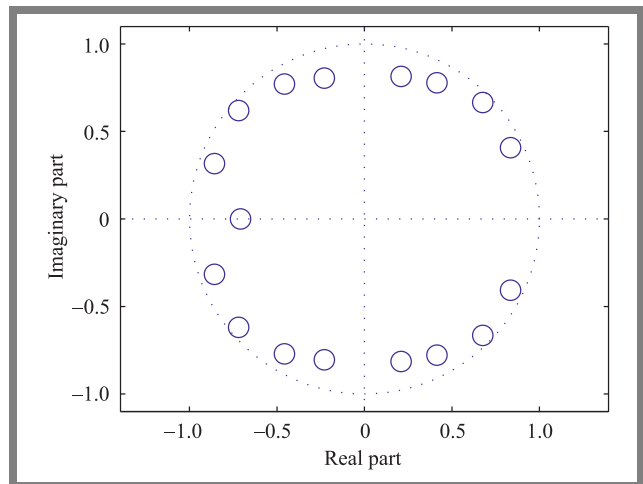


Fig. 6. Zeros of the BRAN C model.

To assess the accuracy of the proposed KE-IPNLMS algorithm, we looked at four different BRAN models with defined properties (i.e. known parameters), then we tried to recuperate these parameters under an additive noise.

Gaussian for an $SNR = 16$ dB and data length $N = 2^{10}$, and we compared them with the two other algorithms proposed in the literature during 50 Monte Carlo runs. Figure 9 illus-

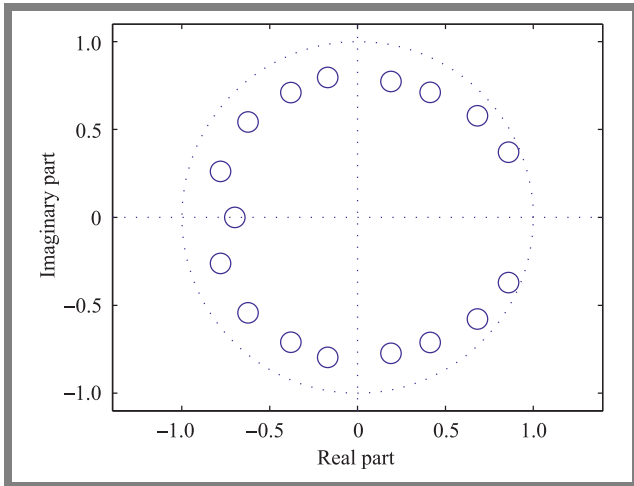


Fig. 7. Zeros of the BRAN D model.

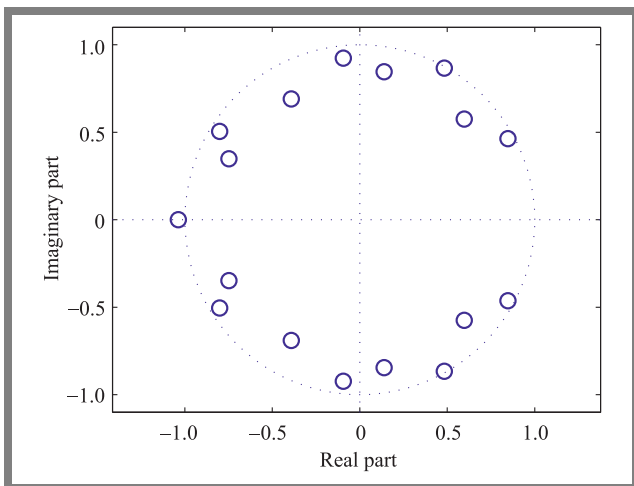


Fig. 8. Zeros of the BRAN E model.

trates the estimated magnitude and phase parameters of the BRAN C radio channel impulse response, using the algorithms presented previously, for a data length of $N = 2^{10}$ and an $SNR = 16$ dB. When the proposed KE-IPNLMS algorithm is used, the magnitude and phase response is estimated with reasonable precision, but several fluctuations are observed when PNLMS and IPNLMS algorithms are used.

Figure 10 demonstrates the magnitude and phase estimation of the BRAN D radio channel impulse response obtained using the proposed KE-IPNLMS algorithm, compared with PNLMS and IPNLMS algorithms for a data length of $N = 2^{10}$ and for $SNR = 16$ dB. The estimated magnitude and phase curves, obtained using the proposed algorithm (KE-IPNLMS), follow the real model with only a slight deviation. When BRAN D radio channel impulse response Parameters are estimated using the IPNLMS algorithm, some minor differences exist between the estimated magnitude and the real model (measured values), and an obvious difference is evident if the PNLMS algorithm is employed. In practical channels when multipath fading is severe for the learning sequence duration, the estimates could yield poor quality results.

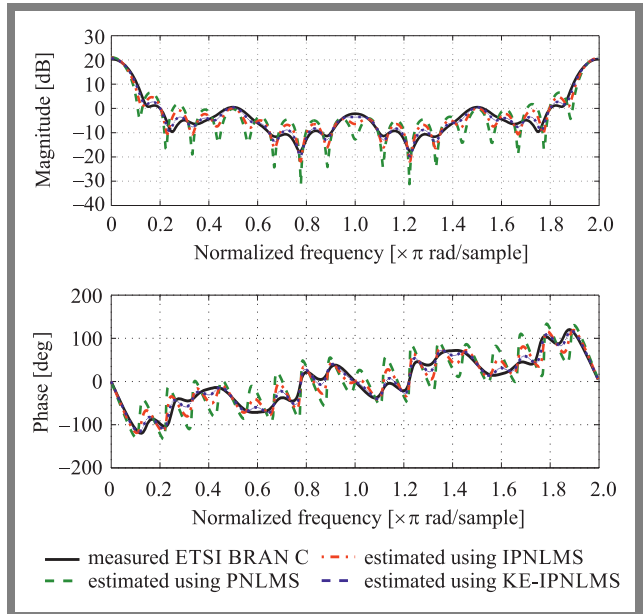


Fig. 9. BRAN C channel magnitude and phase estimation for $N = 2^{10}$ and $SNR = 16$ dB.

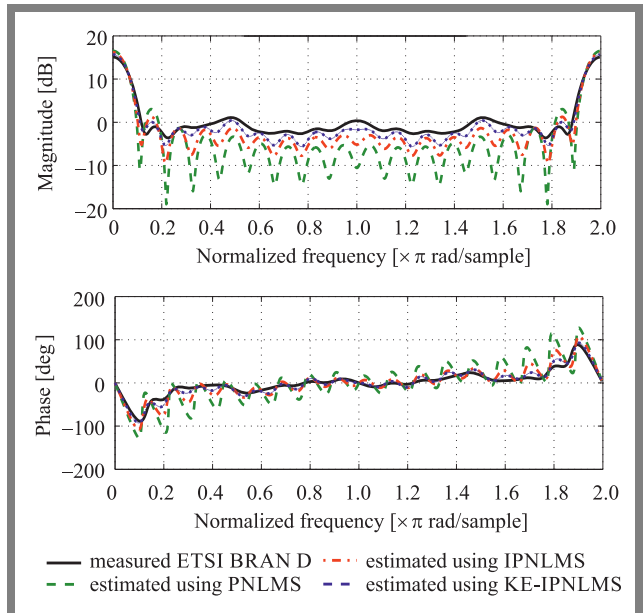


Fig. 10. BRAN D channel magnitude and phase estimation for $N = 2^{10}$ and $SNR = 16$ dB.

Figure 11 shows the estimated magnitude and phase of the BRAN E radio channel impulse response parameters, for a data length of $N = 2^{10}$ and for $SNR = 16$ dB. It should be observed that with the proposed KE-IPNLMS algorithm, the estimated magnitude and phase have the same forms as those measured. When compared with the PNLMS and IPNLMS algorithm, we note that the estimated magnitude follows the variations of the real model's parameters. Performance of the PNLMS algorithm degrades during phase estimation and a large difference between the estimated BRAN E radio channel impulse response and the measured phase is observed. To summarize, Gaussian noise exerts a significant impact on the phase, but only a minor impact on the amplitude estimates.

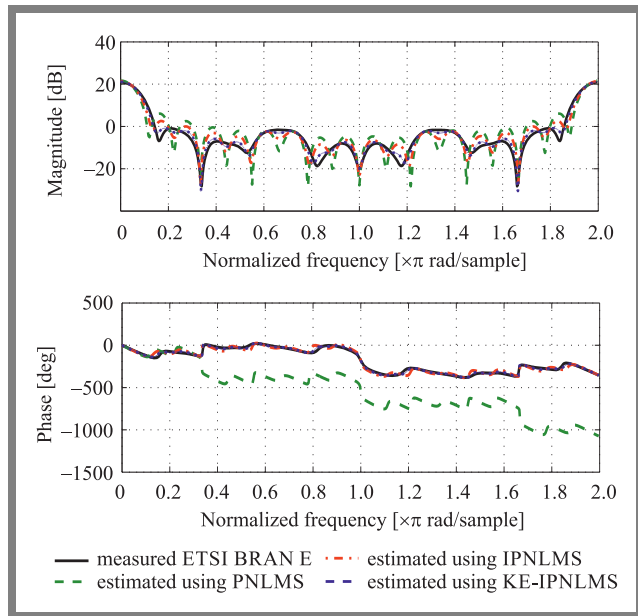


Fig. 11. BRAN E channel magnitude and phase estimation for $N = 2^{10}$ and $SNR = 16$ dB.

4.3. Performance in Noisy Environment

Here, we test the performance of the algorithms in a Gaussian noise environment, where SNR varies from 0 to 30 dB and for a fixed data length of $N = 2^{10}$. The results are summarized in Tables 4–7 for 50 Monte Carlo runs. With all these different results taken into consideration, we have several more important points to make.

The proposed KE-IPNLMS algorithm has offers excellent convergence performance in comparison to its PNLMS and IPNLMS counterparts, for all signal-to-noise ratio values, even in a high noise environment ($SNR = 0$ dB), since the MSE values of the proposed KE-IPNLMS algorithm are very low, contrary to those obtained by means of PNLMS and IPNLMS algorithms.

When SNR is adjusted from 0 to 30, even if the MSE criterion decreases for the three algorithms, the influence of the Gaussian noise disappears and the proposed KE-IPNLMS algorithm demonstrates its superiority over the remaining varieties.

As shown in Tables 4–7, performance of the proposed KE-IPNLMS solution is substantially better than that of other algorithms. For example, in the case of Macchi channel, with $SNR = 30$ dB, MSE values obtained using the proposed KE-IPNLMS algorithm are seven and four times lower than MSE values obtained by means of PNLMS and IPNLMS algorithms, respectively.

Based on Tables 6 and 7, we have observed that when $SNR = 10$ dB, the MSE value achieved by the proposed KE-IPNLMS algorithm equals only 21% and 37% of the MSE value obtained using PNLMS and IPNLMS algorithms, respectively, in the case of the BRAN D impulse response channel, as well as 38% and 56% of the MSE value using PNLMS and IPNLMS algorithms, respectively, in the cases of the BRAN E impulse response channel. These results give

Tab. 4. MSE values of all algorithms for different SNR and a data length $N = 2^{10}$ in the case of the Macchi channel.

SNR [dB]	Algorithm	MSE [dB]
0	PNLMS	-01.39
	IPNLMS	-04.25
	Proposed	-09.88
10	PNLMS	-02.53
	IPNLMS	-04.75
	Proposed	-18.41
20	PNLMS	-02.59
	IPNLMS	-04.87
	Proposed	-22.12
30	PNLMS	-02.81
	IPNLMS	-05.16
	Proposed	-22.47

Tab. 5. MSE values of all algorithms for different SNR and a data length $N = 2^{10}$ in the case of the BRAN C channel.

SNR [dB]	Algorithm	MSE [dB]
0	PNLMS	-03.80
	IPNLMS	-04.90
	Proposed	-05.17
10	PNLMS	-07.64
	IPNLMS	-09.44
	Proposed	-12.06
20	PNLMS	-07.79
	IPNLMS	-10.15
	Proposed	-13.87
30	PNLMS	-08.19
	IPNLMS	-10.26
	Proposed	-14.42

Tab. 6. MSE values of all algorithms for different SNR and a data length $N = 2^{10}$ in the case of the BRAN D channel.

SNR [dB]	Algorithm	MSE [dB]
0	PNLMS	-03.49
	IPNLMS	-05.26
	Proposed	-07.54
10	PNLMS	-04.32
	IPNLMS	-06.73
	Proposed	-11.04
20	PNLMS	-04.25
	IPNLMS	-06.95
	Proposed	-11.18
30	PNLMS	-04.64
	IPNLMS	-07.09
	Proposed	-11.49

a clear indication regarding the high accuracy of the proposed KE-IPNLMS algorithm.

Tab. 7. MSE values of all algorithms for different SNR and a data length $N = 2^{10}$ in the case of the BRAN E channel.

SNR [dB]	Algorithm	MSE [dB]
0	PNLMS	-02.00
	IPNLMS	-04.14
	Proposed	-04.79
10	PNLMS	-07.52
	IPNLMS	-09.16
	Proposed	-11.61
20	PNLMS	-8.00
	IPNLMS	-10.51
	Proposed	-14.28
30	PNLMS	-08.21
	IPNLMS	-10.68
	Proposed	-14.79

4.4. Performance Using Different Data Lengths

Now, we shall focus on the impact that parameter factor N exerts on the performance of the proposed KE-IPNLMS algorithm. Note that N is a data length that impacts the estimated channel parameters and the level of the mean square error. The results are averaged by means of 50 Monte Carlo trials.

The MSE evolution curves of these three algorithms are plotted in Figs. 12–15 for different channel impulse responses. From these results, we may observe that the impact of N is evident, which is linked to the regularity of the evaluated mean square error. It is clearly seen that the proposed algorithm offers the best performance and is also statistically important. For example, in Fig. 12, if N is 8000, MSE is lower than -30 dB in the case of the proposed KE-IPNLMS algorithm. However, we get an MSE that is close to -15 dB and just be-

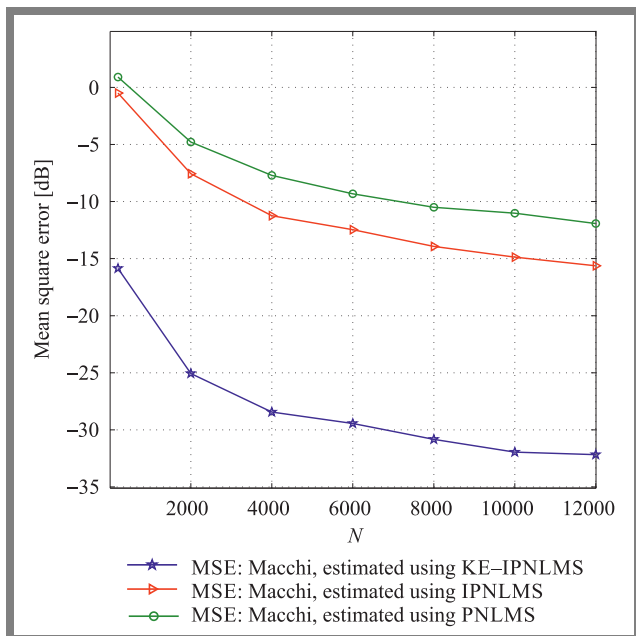


Fig. 12. Comparison of algorithms in terms of MSE for various data lengths N and for a fixed SNR = 16 dB, Macchi channel.

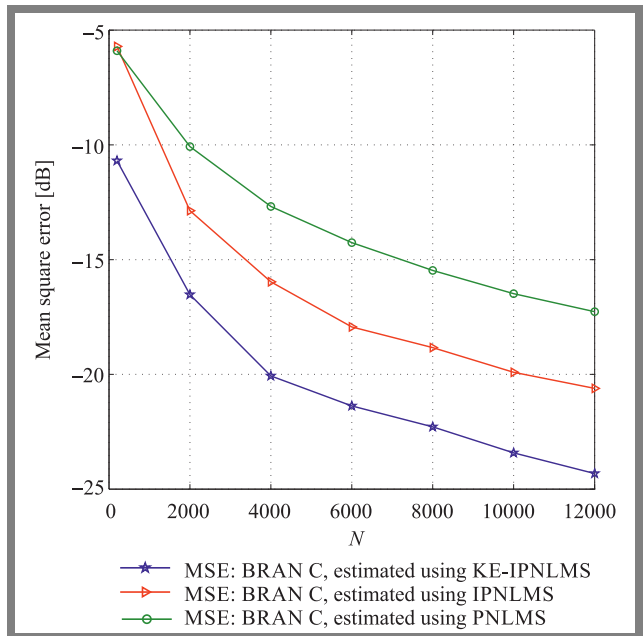


Fig. 13. Comparison of algorithms in terms of MSE for various data lengths N and for a fixed SNR = 16 dB, BRAN C channel.

low -10 dB when we use IPNLMS and PNLMS algorithms, respectively.

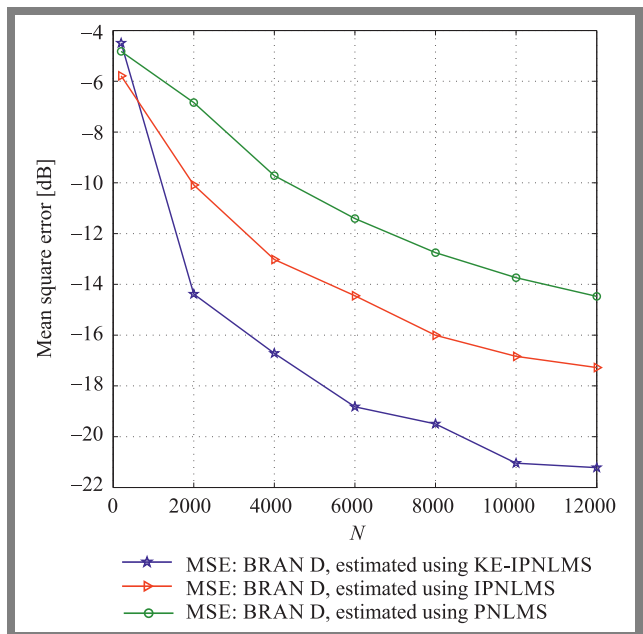


Fig. 14. Comparison of algorithms in terms of MSE for various data lengths N and for a fixed SNR = 16 dB, BRAN D channel.

From Figs. 13–15 it is evident that the data length is low ($N \leq 2000$), a very slow convergence is observed. Each time we increase the data length, we notice an improvement in the convergence speed. This shows that the speed of convergence of the three algorithms is proportional to data length. We can evidently see that the proposed KE-IPNLMS algorithm converges most quickly and has the lowest mean square error. During this time, the mean square error values of the IPNLMS algorithm are inferior to those of the PNLMS algorithm, but

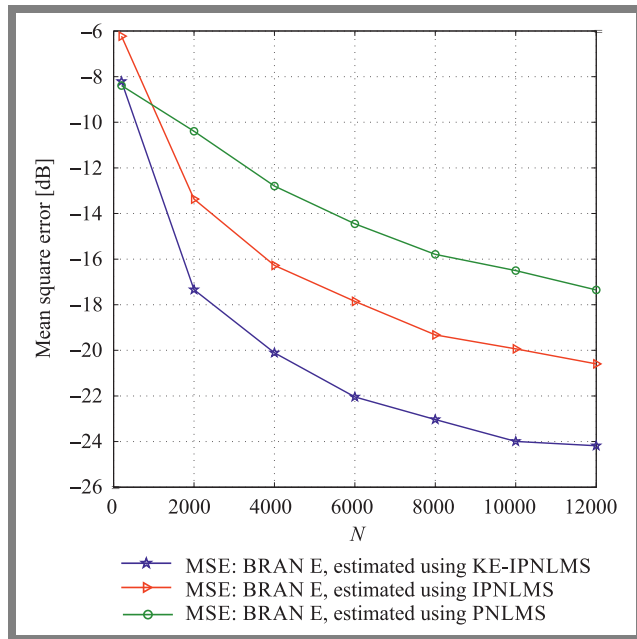


Fig. 15. Comparison of algorithms in terms of MSE for various data lengths N and for a fixed $SNR = 16$ dB, BRAN E channel.

it converges at a slow rate, which implies that the parameters estimated using the proposed KE-IPNLMS algorithm are very close to the exact values when compared to those given by the PNLMS and IPNLMS algorithms. It is very important to select an appropriate value of N and SNR in order to achieve a successful result.

Based on our study of the performance and convergence speed of these algorithms, we remarked that this proposed algorithmic version yields good experimental results in terms of channel identification from output binary measurements.

5. Conclusion and Future Scope

Numerical simulations for the Hammerstein system identification problem with binary measurements on the output have confirmed that the proposed KE-IPNLMS algorithm outperforms PNLMS and IPNLMS in terms of identification of magnitude and phase of channel impulse response parameters (BRAN (C, D and E) and Macchi channels), while only requiring linear computational complexity. In all simulations, we obtained good results in terms of channel identification even more in highest noise power (i.e low SNR) by using the proposed KE-IPNLMS algorithm.

The future work will focus on the development of an extension of this algorithm to MIMO systems and on comparing it with the existing methods, including quantized kernel recursive least squares (QKRLS), quantized kernel Least Incosh (QKLL) and cumulant-based methods.

References

[1] S. Haykin, "Adaptive Filter Theory", fourth ed., Prentice Hall, Delhi, 2002 (ISBN: 9780130901262).

- [2] M.M. Sondhi, "The history of echo cancellation", *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 95–102, 2006 (DOI: 10.1109/MSP.2006.1708416).
- [3] A.H. Sayed, "Fundamentals of Adaptive Filtering", *John Wiley & Sons*, 2003 (ISBN: 9780471461265).
- [4] P.S.R. Diniz, "Adaptive Filtering: Algorithms and Practical Implementations", *Springer*, 2008 (DOI: 10.1007/978-0-387-68606-6).
- [5] L. Ljung, "System identification: theory for the user", *Upper Saddle River (NJ): Prentice Hall PTR*, 1999 (DOI: 10.1002/047134608x.w1046).
- [6] E. Ferrara, "Fast implementations of LMS adaptive filters", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 474–475, 1980 (DOI: 10.1109/tassp.1980.1163432).
- [7] B. Widrow and S.D. Stearns, "Adaptive Signal Processing", *Upper Saddle River: Prentice – Hall*, 1985 (ISBN: 9780130040299).
- [8] R. Martinek, J. Rziky, R. Jaros, P. Bilik, and M. Ladrova, "Least Mean Squares and Recursive Least Squares Algorithms for Total Harmonic Distortion Reduction Using Shunt Active Power Filter Control", *Energies*, vol. 12, no. 8, pp. 1545, 2019 (DOI: 10.3390/en12081545).
- [9] D. Etter, "Identification of sparse impulse response systems using an adaptive delay filter", *In 1985 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, IEEE, vol. 10, pp. 1169–1172, 1985 (DOI: 10.1109/icassp.1985.1168275).
- [10] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers", *IEEE Transactions on speech and audio processing*, vol. 8, no. 5, pp. 508–518, 2000 (DOI: 10.1109/89.861368).
- [11] J. Benesty and S.L. Gay, "An improved PNLMS algorithm", *In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. II-1881–II-1884, 2002 (DOI: 10.1109/icassp.2002.5744994).
- [12] H. Deng and M. Doroslovacki, "Improving convergence of the PNLMS algorithm for sparse impulse response identification", *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181–184, 2005 (DOI: 10.1109/lsp.2004.842262).
- [13] P.A. Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation", *Signal Processing*, vol. 86, no. 6, pp. 1182–1192, 2006 (DOI: 10.1016/j.sigpro.2005.09.015).
- [14] L. Liu, M. Fukumoto, and S. Saiki, "An improved μ -law proportionate NLMS algorithm", *In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2008, pp. 3797–3800. (DOI: 10.1109/icassp.2008.4518480).
- [15] Y. Gu, J. Jin, and S. Mei, " l_0 norm constraint LMS algorithm for sparse system identification", *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009 (DOI: 10.1109/lsp.2009.2024736).
- [16] Z. Jin, X. Ding, Z. Jiang, and Y. Li, "An Improved μ -law Proportionate NLMS Algorithm for Estimating Block-Sparse Systems", *In 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT)*, IEEE, pp. 205–209, 2019 (DOI: 10.1109/iceict.2019.8846290).
- [17] Y. Chen, Y. Gu, and A.O. Hero, "Sparse LMS for system identification", *In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 3125–3128, 2009 (DOI: 10.1109/icassp.2009.4960286).
- [18] Y. Li and M. Hamamura, "Zero-attracting variable-step-size least mean square algorithms for adaptive sparse channel estimation", *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 9, pp. 1189–1206, 2015 (DOI: 10.1002/acs.2536).
- [19] Y. Li, Y. Wang, F. Albu, and J. Jiang, "A general zero attraction proportionate normalized maximum correntropy criterion algorithm for sparse system identification", *Symmetry*, vol. 9, no. 10, pp. 229, 2017 (DOI: 10.3390/sym9100229).
- [20] O. Nelles, "Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes", *Springer Nature*, 2020 (DOI: 10.1007/978-3-030-47439-3).
- [21] S.A. Billings and S. Chen, "Identification of non-linear rational systems using a prediction-error estimation algorithm", *International Journal of Systems Science*, vol. 20, no. 3, pp. 467–494, 1989 (DOI: 10.1080/00207728908910143).
- [22] F. Ding, X.P. Liu, and G. Liu, "Identification methods for Hammerstein nonlinear systems", *Digital Signal Processing*, vol. 21, no. 2, pp. 215–238, 2011 (DOI: 10.1016/j.dsp.2010.06.006).
- [23] R. Fateh, A. Darif, and S. Safi, "Performance Evaluation of MC-CDMA Systems with Single User Detection Technique using Ker-

- nel and Linear Adaptive Method”, *Journal of Telecommunications and Information Technology*, no. 4, pp. 1–11 2021 (DOI: 10.26636/jtit.2021.151621).
- [24] R. Fateh and A. Darif, “Mean Square Convergence of Reproducing Kernel for Channel Identification: Application to Bran D Channel Impulse Response”, *International Conference on Business Intelligence*, pp. 284–293, 2021 (DOI: 10.1007/978-3-030-76508-8_20).
- [25] R. Fateh, A. Darif, and S. Safi, “Channel Identification of Non-linear Systems with Binary-Valued Output Observations Based on Positive Definite Kernels”, *E3S Web of Conferences, EDP Sciences*, vol. 297, 2021 (DOI: 10.1051/e3sconf/202129701020).
- [26] M. Zidane, S. Safi, and M. Sabri, “Compensation of fading channels using partial combining equalizer in MC-CDMA systems”, *J. of Telecommun. and Informat. Technol.*, no. 1, pp. 5–11, 2017 (http://dlibra.itl.waw.pl/dlibra-webapp/Content/1962/ISSN_1509-4553_1_2017_5.pdf).
- [27] S. Safi, M. Frikel, A. Zeroual, and M. M’Saad, “Higher order cumulants for identification and equalization of multicarrier spreading spectrum systems”, *J. of Telecommun. and Informat. Technol.*, vol. 2, no. 1, pp. 74–84, 2011 (<https://www.itl.waw.pl/czasopisma/JTIT/2011/2/74.pdf>).
- [28] R. Fateh, A. Darif, and S. Safi, “Kernel and Linear Adaptive Methods for the BRAN Channels Identification”, *In International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 579–591, 2020 (DOI: 10.1007/978-3-030-90639-9_47).
- [29] R. Fateh, A. Darif, and S. Safi, “Identification of the Linear Dynamic Parts of Wiener Model Using Kernel and Linear Adaptive”, *In International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 387–400, 2020 (DOI: 10.1007/978-3-030-90639-9_31).
- [30] E. Eskinat, S.H. Johnson, and W.L. Luyben, “Use of Hammerstein models in identification of nonlinear systems”, *AICHE Journal*, vol. 37, no. 2, pp. 255–268, 1991 (DOI: 10.1002/aic.690370211).
- [31] S.A. Billings, “Identification of nonlinear systems—a survey”, *IEE Proceedings D-Control Theory and Applications*, IET, vol. 127, no. 6, pp. 272–285, 1980 (DOI: 10.1049/ip-d:19800047).
- [32] A. Stenger and R. Rabenstein, “Adaptive Volterra filters for nonlinear acoustic echo cancellation”, *IEEE Workshop Nonlinear Signal Image Process (NSIP)*, vol. 2, pp. 679–683, 1999 (<http://www.eurasiac.org/Proceedings/Ext/NSIP99/Nsip99/papers/146.pdf>).
- [33] L. Janjanam, S.K. Saha, R. Kar, and D. Mandal, “Volterra filter modelling of non-linear system using artificial electric field algorithm assisted Kalman filter and its experimental evaluation”, *ISA transactions*, 2020 (DOI: 10.1016/j.isatra.2020.09.010).
- [34] T. Ogunfunmi, “Adaptive Nonlinear System Identification: The Volterra and Wiener Model Approaches”, *Springer Science*, 2007 (ISBN 9780387263281).
- [35] S. Haykin, “Neural Networks and Learning Machines, Englewood Cliffs”, *Prentice-Hall*, 2009 (ISBN 9780387263281).
- [36] N. Aronszajn, “Theory of reproducing kernels”, *Transactions of the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.
- [37] J. Shawe-Taylor and N. Cristianini, “Kernel methods for pattern analysis”, *Cambridge university: press*, 2004 (DOI: 10.1017/CBO9780511809682).
- [38] W. Liu, J.C. Principe, and S. Haykin, “Kernel adaptive filtering: a comprehensive introduction”, *John Wiley & Sons*, vol. 57, 2011 (ISBN: 9780470447536).
- [39] G. Camps-Valls and L. Bruzzone, “Kernel methods for remote sensing data analysis”, *John Wiley & Sons*, 2009 (DOI: 10.1002/9780470748992).
- [40] C. Cortes and V. Vapnik, “Support-vector networks”, *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995 (DOI: 10.1007/bf00994018).
- [41] B. Scholkopf and A.J. Smola, “Learning with kernels: support vector machines, regularization, optimization, and beyond, Adaptive Computation and Machine Learning series”, *MIT Press*, 2001 (<https://alex.smola.org/papers/2001/SchHerSmo01.pdf>).
- [42] Z. Momani, M. Al-Shridah, O.A. Arqub, M. Al-Momani, and S. Momani, “Modeling and analyzing neural networks using reproducing kernel Hilbert space algorithm”, *Appl Math Inf Sci*. vol. 12, pp. 89–99, 2018 (DOI: 10.18576/amis/120108).
- [43] O.A. Arqub and H. Rashaideh, “The RKHS method for numerical treatment for integrodifferential algebraic systems of temporal two-point BVPs”, *Neural Computing and Applications*, vol. 30, no. 8, pp. 2595–2606, 2018 (DOI: 10.1007/s00521-017-2845-7).
- [44] H. Sun, “Mercer theorem for RKHS on noncompact sets”, *Journal of Complexity*, vol. 21, no. 3, pp. 337–349, 2005 (DOI: 10.1016/j.jco.2004.09.002).
- [45] W. Liu, P.P. Pokharel, and J.C. Principe, “The kernel least-mean-square algorithm”, *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 543–554, 2008 (DOI: 10.1109/tsp.2007.907881).
- [46] Y. Engel, S. Mannor, and R. Meir, “The kernel recursive least-squares algorithm”, *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004 (DOI: 10.1109/tsp.2004.830985).
- [47] W. Liu and J.C. Principe, “Kernel affine projection algorithms”, *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–12, 2008 (DOI: 10.1155/2008/784292).
- [48] R. Castro Garcia, “Structured nonlinear system identification using kernel-based methods, Ph.D. dissertation”, *Faculty of Engineering Science*, 2017 (<https://lirias.kuleuven.be/retrieve/473580>).
- [49] B. Chen, S. Zhao, P. Zhu, and J.C. Principe, “Quantized kernel least mean square algorithm”, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, 2011 (DOI: 10.1109/tnnls.2011.2178446).
- [50] B. Chen, S. Zhao, P. Zhu, and J.C. Principe, “Quantized kernel recursive least squares algorithm”, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1484–1491, 2013 (DOI: 10.1109/tnnls.2013.2258936).
- [51] S. Wang, Y. Zheng, and C. Ling “Regularized kernel least mean square algorithm with multiple-delay feedback”, *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 98–101, 2015 (DOI: 10.1109/lsp.2015.2503000).
- [52] B. Chen, J. Liang, N. Zheng, and J.C. Principe, “Kernel least mean square with adaptive kernel size”, *Neurocomputing*, vol. 191, pp. 95–106, 2016 (DOI: 10.1016/j.neucom.2016.01.004).
- [53] W. Liu, I. Park, Y. Wang, and J.C. Principe, “Extended kernel recursive least squares algorithm”, *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3801–3814, 2009 (DOI: 10.1109/tsp.2009.2022007).
- [54] H. Zhou, J. Huang, and F. Lu, “Reduced kernel recursive least squares algorithm for aero-engine degradation prediction”, *Mechanical Systems and Signal Processing*, vol. 95, pp. 446–467, 2017 (DOI: 10.1016/j.ymssp.2017.03.046).
- [55] M. Zidane and R. Dinis, “A new combination of adaptive channel estimation methods and TORC equalizer in MC-CDMA systems”, *International Journal of Communication Systems*, vol. 33, no. 11, pp. e4429, 2020 (DOI: 10.1002/dac.4429).
- [56] C.J. Burges, “A tutorial on support vector machines for pattern recognition”, *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998 (DOI: 10.1023/a:1009715923555).
- [57] A. Sánchez and V. David, “Advanced support vector machines and kernel methods”, *Neurocomputing*, vol. 55, no. 1–2, pp. 5–20, 2003 (DOI: 10.1016/s0925-2312(03)00373-4).
- [58] U. Soverini and T. Söderström, “Frequency domain identification of FIR models in the presence of additive input-output noise”, *Automatica*, vol. 115, pp. 108879, 2020 (DOI: 10.1016/j.automatica.2020.108879).
- [59] H. Zhang, T. Wang, and Y. Zhao, “Asymptotically Efficient Recursive Identification of FIR Systems With Binary-Valued Observations”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 2687–2700, 2019 (DOI: 10.1109/tsmc.2019.2916022).
- [60] J. Guo, Y. Zhao, C.Y. Sun, and Y. Yu, “Recursive identification of FIR systems with binary-valued outputs and communication channels”, *Automatica*, vol. 60, pp. 165–172, 2015 (DOI: 10.1016/j.automatica.2015.06.030).
- [61] T. Yuan, Q. Liu, and J. Guo, “Identification of FIR Systems with Quantized Input and Binary-Valued Observations Under A Priori Parameter Constraint”, *39th Chinese Control Conference (CCC)*, IEEE, pp. 1099–1104, 2020 (DOI: 10.23919/ecc50068.2020.9188983).
- [62] M. Poulliquen, T. Menard, E. Pigeon, O. Gehan, and A. Goudjil, “Recursive system identification algorithm using binary measurements”, *2016 European Control Conference (ECC)*, IEEE, pp. 1353–1358, 2016 (DOI: 10.1109/ecc.2016.7810477).
- [63] C.J. Burges, “A tutorial on support vector machines for pattern

recognition”, *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998 (DOI: 10.1023/a:1009715923555).

- [64] O. Macchi and C.A. Faria da Roccha, “Travassos-Romano JM. Égalisation adaptative autodidacte par rétroprédiction et prédiction”, XIV colloque GRETSI, pp. 491–493, 1993 (http://www.gretsi.fr/data/colloque/pdf/1993_017-0005_11877.pdf).
- [65] ETSI. “Broadband radio access network (BRAN); hyperlan type 2; physical layer”, Technical report, 2001 (<https://www.etsi.org/deliver/etsits/101400101499/101475/01.02.0260/ts101475v010202p.pdf>).
- [66] ETSI. “Broadband Radio Access Network (BRAN); hyperlan type 2; Requirements and architectures for wireless broadband access”, 1999 (<https://www.etsi.org/deliver/etsitr/101000101099/101031/02.02.0160/tr101031v020201p.pdf>).



Rachid Fateh received his B.Sc. in Mathematics, Computer Science and Applications from Ibn Zohr University, Ouarzazate, Morocco in 2017 and M.Sc. in Telecommunications Systems and Computer Networks from Sultan Moulay Slimane University, Beni Mellal, Morocco in 2019. He is currently pursuing his Ph.D. in Mathematics and Computer Science from

Sultan Moulay Slimane University, Beni Mellal, Morocco. His research interests include system identification, equalization of channel communications with a reproducing kernel of Hilbert space, signal processing and estimation.

 <https://orcid.org/0000-0002-0574-2105>

E-mail: fateh.smi@gmail.com

Laboratory of Innovation in Mathematics, Applications and Information Technologies (LIMATI), Sultan Moulay Slimane University, Beni Mellal, Morocco



Anouar Darif received his B.Sc. in IEEA (Informatique Electrothec- nique, Electronique and Automatique) from Dhar El Mahraz Faculty of Sciences at Mohamed Ben Abdel- lah University Fez, Morocco in 2005. He received the Diplôme d’Etudes Supérieures Approfondies in Comput- er Sciences and Telecommunications from the Faculty of Sciences, Rabat

in 2007. He received his Ph.D. degree in Computer Sciences and Telecommunications from the Faculty of Sciences, Rabat in 2015. He is currently a Research and Teaching Associate at the Multidisciplinary Faculty, University of Sultan Moulay Slimane Beni Mellal, Morocco. His research interests include wireless sensor networks, mobile edge computing, Internet of Things and telecommunications (channels identification, IR-UWB, WCDMA, LTE, LTE-A).

 <https://orcid.org/0000-0001-8026-9189>


E-mail: anouar.darif@gmail.com

Laboratory of Innovation in Mathematics, Applications and Information Technologies (LIMATI), Sultan Moulay Slimane University, Beni Mellal, Morocco



Said Safi received his B.Sc. in Physics (option: Electronics) from Cadi Ayyad University, Marrakech, Morocco in 1995, M.Sc. and Ph.D. degree from Chouaib Doukkali Uni- versity and Cadi Ayyad Univerrsty, in 1997 and 2002, respectively. He was a Professor of information the- ory and telecommunication systems at the National School for Applied Sci- ences, Tangier, Morocco, from 2003 to 2005. Since 2006,

he has been a Professor of applied mathematics and programming at the Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco. In 2008, he received a Ph.D. in Telecommunication and Informatics from Cadi Ayyad University. In 2015, he received the degree of Professor in Sciences at Sultan Moulay Slimane University. His general interests span the areas of communications and signal processing, estimation, time-series analysis, and system identification subjects. His current research topics focus on transmitter and receiver diversity techniques for single- and multi-user fading communication channels, and wide-band wireless communication systems.

 <https://orcid.org/0000-0003-3390-9037>

E-mail: safi.said@gmail.com

Laboratory of Innovation in Mathematics, Applications and Information Technologies (LIMATI), Sultan Moulay Slimane University, Beni Mellal, Morocco

Modeling of Microwave Cavities Based on SIBC-FDTD Method for EM Wave Focalization by TR Technique

Zhigang Li¹, Younes Aimer², and Tayeb H. C. Bouazza³

¹ESECA, Department of ENSEEIHT, National Polytechnic, Institute of Toulouse, Toulouse, France,

²Mines Saint-Etienne, Centre of Microelectronics in Provence, Department of Flexible Electronics, Gardanne, France,

³XLIM Laboratory UMR-CNRS 7252, Institute of Technology of Angouleme, University of Poitiers, Poitiers, France

<https://doi.org/10.26636/jtit.2022.153021>

Abstract — The time reversal (TR) techniques used in electromagnetics have been limited for a variety of reasons, including extensive computations, complex modeling and simulation, processes as well as, large-scale numerical analysis. In this paper, the SIBC-FDTD method is applied to address these issues and to efficiently model TR systems. An original curvilinear modeling method is also proposed for constructing various obstacles in a 2D microwave cavity and for processing the corners of the cavity. The EM waves' spatio-temporal focalization has been realized, and results of the simulations further prove the accuracy and effectiveness of this modeling method. Furthermore, they demonstrate that the microwave cavity processes may significantly improve the focalization quality in terms of SLL enhancement.

Keywords — *curvilinear modeling, EM waves focalization, SIBC-FDTD method, TR technique.*

1. Introduction

The time-reversal technique was first proposed by Bogert from Bell Labs in 1957. In 2005, Lerosey *et al.* introduced the TR's spatio-temporal focalization in electromagnetics [1]. In their experiments, electromagnetic wave signals emitted by the antenna were recorded by a receiving antenna known as the time-reversal mirror (TRM). The recorded signal was time-reversely retransmitted by the TRM, thus resulting in focalization of EM waves at the initial position of the transmitting antenna. In recent years, more researches have focused on the applications of this phenomenon, e.g. in the fields of acoustics [2]–[3], medicine [4]–[5], and communications [6]–[9].

To simulate the propagation of an EM wave throughout the entire TR process, the finite-difference time-domain (FDTD) method, one of the most effective numerical calculation methods, has been widely used [10]–[13]. Its main idea is to transform the propagation of EM waves into spatial and temporal propulsion by discretizing Maxwell equations. It needs to be borne in mind that boundary conditions are also critical for simulating propagation of EM waves in a confined space. However, since good conductors have a low skin depth and sharp field variations, a tiny grid is required to calculate the distribution of the EM field on their surfaces,

which significantly increases computational complexity. To solve this problem, researchers have introduced the surface impedance boundary condition (SIBC) to the FDTD method. This method can directly obtain the field distribution at the interface without considering the interior of the conductor, thus drastically improving computation efficiency.

In 1992, Maloney *et al.* proposed an efficient implementation of SIBC in the FDTD method based on exponential approximation, and various experiments have demonstrated its applicability [14]. However, the degree of computational complexity failed to decrease significantly.

In recent years, Mao *et al.* have constructed a new absorbing boundary condition by setting surface impedance to free space in order to terminate the outer boundary of the FDTD computational domain [15]–[16]. This approach dramatically reduced computational complexity without affecting the level of accuracy.

Based on these facts, this study aims to contribute to this increasingly popular area by exploring more effective methods for modeling TR systems.

The paper is organized as follows. Section 2 presents the derived theoretical formulas of the SIBC-FDTD method, which can be used to model the EM wave propagation process, microwave cavities, and obstacles. It continues by explaining the principle of TR modeling and shows how this approach naturally leads to focalization. In Section 3, several simulations are conducted to verify effectiveness of the modelling method and to demonstrate how the presented cavity improves the quality of focalization. Furthermore, a series of parallel simulations is performed to re-validate this conclusion by employing two customized metrics, namely side-lobe level (SLL) and spatial side-lobe level (SSLL).

2. TR System Modeling Approach

2.1. EM Wave Propagation Modeling

With Maxwell's equations serving as a point of departure, time-domain propulsion formulas for the FDTD method may be obtained by performing the central difference and discrete

approximation to its differential form. Thus, the continuous electromagnetic wave propagation problem is converted into a discrete numerical problem, which is well suited to deal with electromagnetic field calculations in large and complex structures.

Assuming that the study space is passive and that electrical and magnetic losses are not considered, the Maxwell time-domain differential equations used for constructing the FDTD algorithm can be expressed as:

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t}, \quad (1)$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (2)$$

where \vec{E} is the electric field intensity in V/m, \vec{H} is the magnetic field intensity in A/m, \vec{B} is the magnetic flux density in Wb/m², and \vec{D} is the electric displacement in C/m².

Besides, constitutive relations are essential to complement Maxwell's equations and describe the medium's properties. The constitutive relations for linear, isotropic and non-dispersive media can be written as:

$$\vec{D} = \varepsilon \cdot \vec{E}, \quad (3)$$

$$\vec{B} = \mu \cdot \vec{H}, \quad (4)$$

where ε is the permittivity of the medium in F/m, μ is the permeability in H/m.

In solving the one-dimensional (1D) problem, it is assumed that the electromagnetic wave propagates along the x -axis, so both E_x and H_x are equal to zero, and Maxwell's equations can be simplified as:

$$\frac{\partial E_z}{\partial x} = \mu \frac{\partial H_y}{\partial t}, \quad (5)$$

$$\frac{\partial H_y}{\partial x} = \epsilon \frac{\partial E_z}{\partial t}. \quad (6)$$

It can be noticed that the left and right-hand sides of the equation are space and time partial derivative term, respectively. The distribution of E_z and H_y in the 1D Yee cell is given in Fig. 1. In space and time dimensions, the \vec{E} and \vec{H} components are staggered at an interval of half the grid length, an essential basis for the FDTD method's iteration.

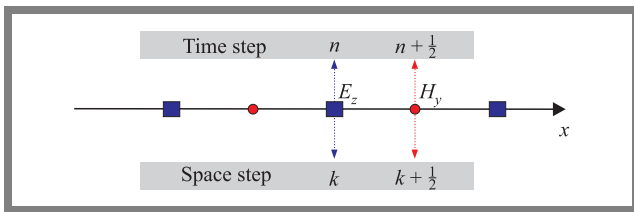


Fig. 1. Distribution of E_z and H_y in the 1D Yee cell.

By using Taylor formulas, the discrete form of Eqs. (5) and (6) can be achieved as:

$$H_y^{n+\frac{1}{2}} \left(k + \frac{1}{2} \right) = H_y^{n-\frac{1}{2}} \left(k + \frac{1}{2} \right) + \frac{\Delta t}{\mu \Delta x} [E_z^n(k+1) - E_z^n(k)], \quad (7)$$

$$E_z^{n+1}(k) = E_z^n(k) + \frac{\Delta t}{\epsilon \Delta x} \left[H_y^{n+\frac{1}{2}} \left(k + \frac{1}{2} \right) - H_y^{n+\frac{1}{2}} \left(k - \frac{1}{2} \right) \right]. \quad (8)$$

Moreover, because FDTD is an infinite approximation method, stability conditions must be set to ensure the convergence of the solution. The size of the grid must also be limited to reduce errors. The relationship is given by:

$$\delta \leq \frac{\lambda_{\min}}{10}, \quad (9)$$

$$c \cdot \Delta t \leq \frac{1}{\sqrt{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2 + \left(\frac{1}{\Delta z}\right)^2}}, \quad (10)$$

where $\delta = \min[\Delta x, \Delta y, \Delta z]$ and λ_{\min} is the minimum wavelength in the computational frequency range, Δx , Δy and Δz are the space steps, Δt is the time step, and c is the speed of light.

Assuming $\Delta x = \Delta y = \Delta z$, the relationship can be replaced like $c \cdot \Delta t \leq \delta$, $c \cdot \Delta t \leq \frac{\delta}{\sqrt{2}}$ and $c \cdot \Delta t \leq \frac{\delta}{\sqrt{3}}$ for 1D, 2D and 3D, respectively. This means that the time interval must not be greater than the time it takes for the wave to pass through one Yee cell at the speed of light.

2.2. 1D SIBC-FDTD Iterative Formulations for Microwave Cavities Modeling

SIBC describes the relationship between the tangential electric field and the tangential magnetic field at two different media interfaces in the frequency domain. The first-order SIBC in the frequency domain is given by:

$$\vec{E}(\omega) = Z_s(\omega)[\vec{n} \times \vec{H}(\omega)], \quad (11)$$

where $s = j\omega$, ω is the angular frequency, $Z_s(\omega)$ is the surface impedance of the conductor, \vec{n} is a unit vector normal to the surface of the lossy metal. Since the surface impedance of a good conductor is

$$Z_s(\omega) = (1 + j)\sqrt{\frac{\omega\mu}{2\sigma}} = \sqrt{\frac{j\omega\mu}{\sigma}}, \quad (12)$$

$Z_s(\omega)$ can be separated into the sum of surface resistance $R_s(\omega)$ and surface inductance $L_s(\omega)$ as:

$$Z_s(\omega) = R_s(\omega) + j\omega L_s(\omega). \quad (13)$$

At a specific frequency, $R_s(\omega)$, $L_s(\omega)$ can be treated as approximate constants, as $R_s(\omega) = \sqrt{\frac{\omega\mu}{2\sigma}}$, $L_s(\omega) = \sqrt{\frac{\mu}{2\sigma\omega}}$. Hence, Eq. (11) can be rewritten as:

$$\vec{E}(\omega) = [R_s + j\omega L_s][\vec{n} \times \vec{H}(\omega)]. \quad (14)$$

By using the inverse Fourier transform, the relationship between \vec{E} and \vec{H} can be expressed in the time domain as:

$$\vec{E}(t) = \left[R_s + L_s \frac{\partial}{\partial t} \right] [\vec{n} \times \vec{H}(t)]. \quad (15)$$

Hence, the schematic diagram of the SIBC-FDTD method in the 1D case may be shown in Fig. 2.

In the FDTD iteration formulas, by replacing the differential operations of $E_z(1 + \frac{1}{2})$ and $E_z(n\Delta x + \frac{1}{2})$ in the space with

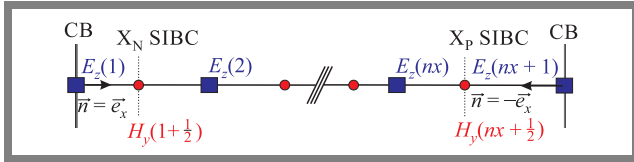


Fig. 2. Schematic diagram of the 1D SIBC-FDTD method.

a half-cell step, the Faraday-Maxwell law can be derived in the following form:

$$H_y^{n+\frac{1}{2}}\left(1 + \frac{1}{2}\right) = \frac{\mu\Delta x - \Delta tR_s + 2L_s}{\mu\Delta x + \Delta tR_s + 2L_s} H_y^{n-\frac{1}{2}}\left(1 + \frac{1}{2}\right) + \frac{2\Delta t}{\mu\Delta x + \Delta tR_s + 2L_s} E_z^n(2), \quad (16)$$

$$H_y^{n+\frac{1}{2}}\left(n\Delta x + \frac{1}{2}\right) = \frac{\mu\Delta x - \Delta tR_s + 2L_s}{\mu\Delta x + \Delta tR_s + 2L_s} H_y^{n-\frac{1}{2}}\left(n\Delta x + \frac{1}{2}\right) + \frac{2\Delta t}{\mu\Delta x + \Delta tR_s + 2L_s} E_z^n(n\Delta x). \quad (17)$$

2.3. 2D SIBC-FDTD Iterative Formulations for Microwave Cavities Modeling

In the case of 2D (TM wave) scenario, the fundamental components comprise H_x , H_y , and E_z . The schematic diagram of the 2D SIBC-FDTD method is similar to [16], as shown in Fig. 3.

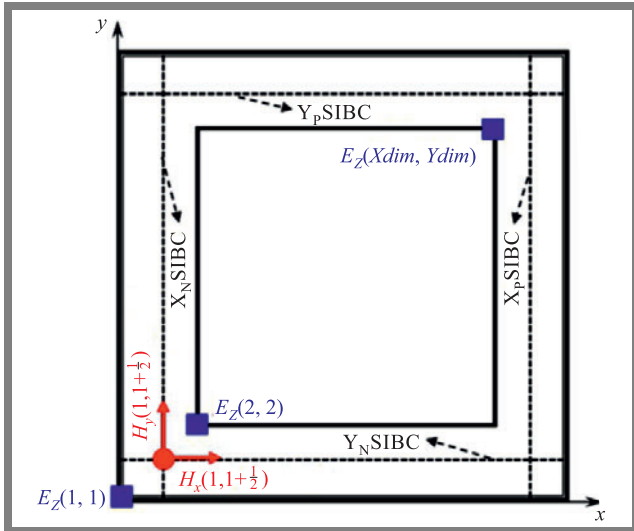


Fig. 3. Schematic diagram of the 2D SIBC-FDTD method.

For any Yee cell on the x -axis negative boundary (X_N SIBC), only H_y existed. The geometry is shown in Fig. 4.

By applying the SIBC equation:

$$E_z\left(1 + \frac{1}{2}, j\right) = R_s H_y\left(1 + \frac{1}{2}, j\right) + L_s \frac{\partial H_y\left(1 + \frac{1}{2}, j\right)}{\partial t}$$

in FDTD, the derived iteration formula on X_N is shown in Eq. (18). Similarly, the iterative recipes on X_P , Y_N and Y_P can be denoted in Eqs. (19), (20), (21), respectively.

$$X = \frac{\mu\Delta x - \Delta tR_s + 2L_s}{\mu\Delta x + \Delta tR_s + 2L_s} H_y^{n-\frac{1}{2}},$$

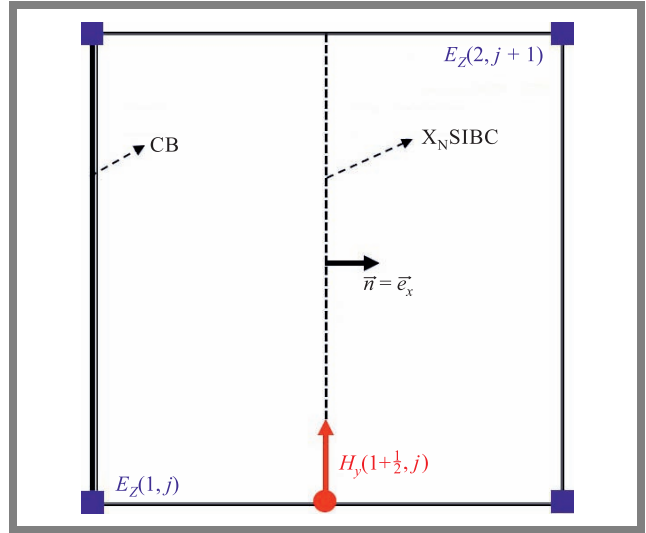


Fig. 4. 2D Yee cell.

$$Y = \frac{\mu\Delta y - \Delta tR_s + 2L_s}{\mu\Delta y + \Delta tR_s + 2L_s} H_x^{n-\frac{1}{2}},$$

$$H_y^{n+\frac{1}{2}}\left(1 + \frac{1}{2}, j\right) = X\left(1 + \frac{1}{2}, j\right) + \frac{2\Delta t}{\mu\Delta x + \Delta tR_s + 2L_s} E_z^n(2, j), \quad (18)$$

$$H_y^{n+\frac{1}{2}}\left(x_{\text{dim}} + \frac{1}{2}, j\right) = X\left(x_{\text{dim}} + \frac{1}{2}, j\right) - \frac{2\Delta t}{\mu\Delta x + \Delta tR_s + 2L_s} E_z^n(x_{\text{dim}}, j), \quad (19)$$

$$H_x^{n+\frac{1}{2}}\left(i, 1 + \frac{1}{2}\right) = Y\left(i, 1 + \frac{1}{2}\right) - \frac{2\Delta t}{\mu\Delta y + \Delta tR_s + 2L_s} E_z^n(i, 1), \quad (20)$$

$$H_x^{n+\frac{1}{2}}\left(i, y_{\text{dim}} + \frac{1}{2}\right) = Y\left(i, y_{\text{dim}} + \frac{1}{2}\right) + \frac{2\Delta t}{\mu\Delta y + \Delta tR_s + 2L_s} E_z^n(i, y_{\text{dim}}). \quad (21)$$

2.4. Curvilinear Modeling

Unlike linear modeling of the microwave cavity, curvilinear modeling is more complicated due to the lack of apparent boundaries in processing corners and constructing obstacles. The solution is to create a matrix of the same dimension corresponding to E_z (blue dots) and to assign E_z values to the matrix as the basis for determining boundaries, as shown in Fig. 5:

$$E_z[i, j] = \begin{cases} 0 & \text{if } E_z \text{ is inside the obstacle} \\ 1 & \text{if } E_z \text{ is outside the obstacle} \end{cases}. \quad (22)$$

For example, if the blue dot is inside the obstacles, the matrix's corresponding point will be allocated a 1. Otherwise, it will be a 0.

Determination of the boundary location requires judging the values of two adjacent points in the matrix as shown in

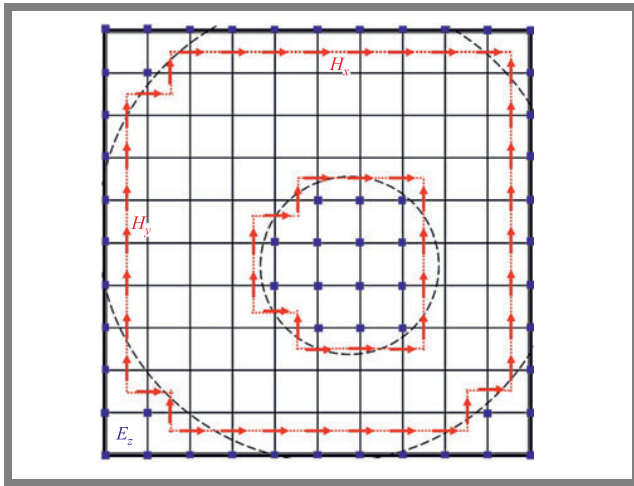


Fig. 5. Schematic diagram of the curvilinear modeling method.

Eq. (23). If the values of two adjacent points on the same row are different, the H_y iteration formula should be utilized as the boundary condition. Similarly, if the values of two adjacent points on the same column are different, the H_x iteration formula should be adopted:

$$\begin{cases} H_x [i, j + \frac{1}{2}] \text{ is the boundary, if } E_z[i, j] \neq E_z[i, j + 1] \\ H_y [i + \frac{1}{2}, j] \text{ is the boundary, if } E_z[i, j] \neq E_z[i + 1, j] \end{cases} \quad (23)$$

It follows from Fig. 5 that the area enclosed by the red arrow (H_x, H_y) inside the cavity is the modeling circular obstacle, and the area surrounded by the red arrow on the cavity's boundary is the cavity after the corner has been processed.

2.5. Time Reversal Modeling

The complete TR experiment is usually divided into two stages [17]–[18], the forward stage and the backward stage,

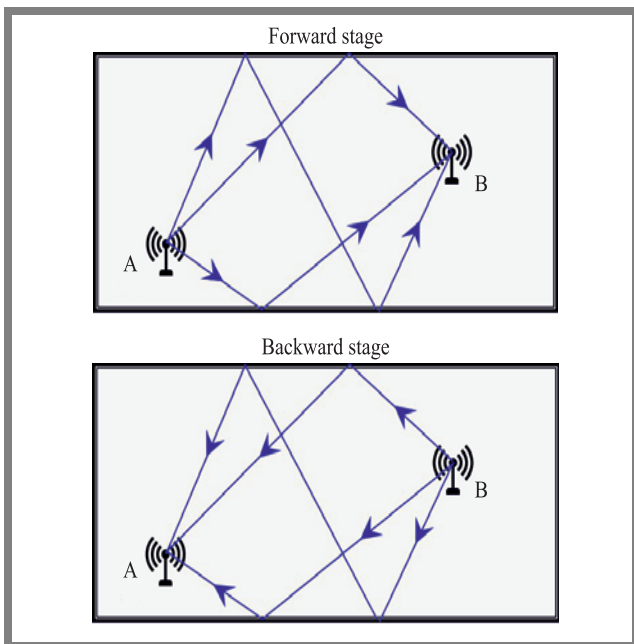


Fig. 6. TR system.

as shown in Fig. 6. During the forward stage, the EM waves signal emitted by source A is received and recorded by the antenna placed at B (TRM) after complex propagation and reflection in the cavity. During the backward stage, the first action is to reverse the signal received at B on the time axis and then the reversed signal is retransmitted, meaning that the signal received at B first will be emitted last.

Then the reversed signal will be focused at source A after a while (approximately the length of the receiving time). This process is known as spatio-temporal focalization characteristic of the TR technique.

3. Simulation Results and Discussion

3.1. Simulation Settings

In this TR simulation, the volume of the metallic microwave cavity is $1.165 \times 1.165 \text{ m}^2$, and the total cavity space is divided into 104×104 grids according to the FDTD stability conditions. Hence, the pixel size used for SIBC-FDTD iteration is approximately 0.0113 m. Moreover, a Gaussian pulse signal is excited at the initial source point A, as shown in Fig. 7. The pulse duration is 8 ns, the carrier frequency is 2.4 GHz, and the amplitude pulse is 10^4 V/m .

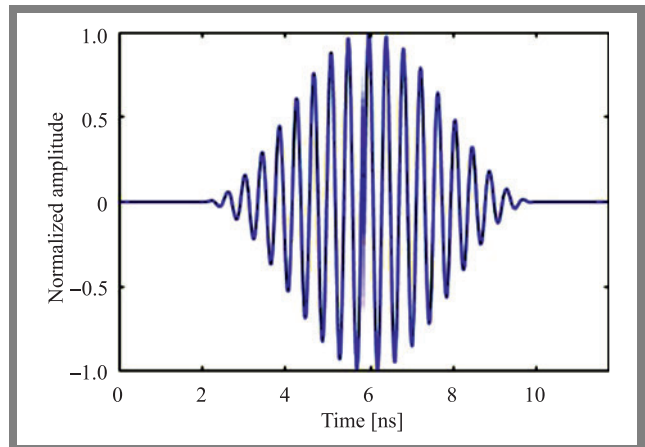


Fig. 7. Initial Gaussian pulse.

Moreover, the signal received by the TRM at point B during the forward stage is shown in Fig. 8. Due to the loss of the cavity's metallic boundaries, the TRM receives a series of signals whose amplitude decays gradually.

To evaluate the focalization quality in the time domain and frequency domain, we have defined such parameters as side-lobe level (SLL) and spatial side-lobe level (SSLL), respectively. SLL refers to the ratio between the first maximum amplitude and the second maximum amplitude of the signal recorded at focalization point A. SSLL, in turn, refers to the ratio between the first spatial maximum amplitude and the second spatial maximum amplitude inside the cavity.

Furthermore, to verify the effectiveness of the curvilinear modeling method and to assess the relationship between cavity complexity and focalization quality, we have set up two in-

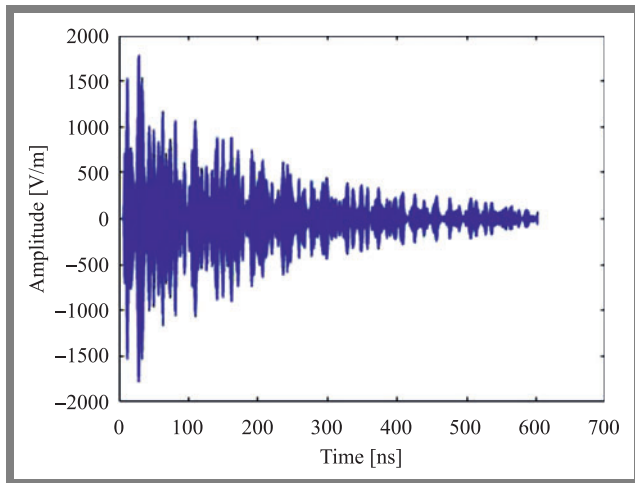


Fig. 8. Recorded signals by TRM at point B.

dependent groups: group 1 with no processing corners and no obstacles, and group 2 with processing corners and obstacles.

3.2. Analysis of Simulation Results

Inside the modeled microwave cavity, the processed corners and the circular obstacle generated using the SIBC-FDTD method can be observed directly in Fig. 9b. Simultaneously, there is no EM wave propagation outside the SIBC boundary, which is in agreement with our modeling theory. Also, at 604.03 ns, the temporal EM wave focalization is more clearly visualized in the group 2 (Fig. 9b) than in group 1 (Fig. 9a).

Figures 10 and 11 quantitatively show the entire focalization process from 0 ns to 604.03 ns. With the reflection of the EM wave emitted from TRM (point B) inside the cavity, the energy gradually accumulates at point A and finally focuses here. Furthermore, by comparing the two figures on the (a) side (group 1), one may notice that additional focalization points exist inside the cavity, for example, at the time points of 350 ns, 480 ns, 530 ns, 560 ns and 590 ns, which means that energy distribution is more dispersed. The comparison of the two figures on the (b) side (group 2) demonstrates that no other apparent focalization points are created and most of the energy is refocused at point A.

Table 1 shows the values of SLL and SSLL obtained from Figs. 10 and 11. Both are boosted more in group 2 than in group 1, which means that the focalization quality is optimized. This is caused by the fact that the first maximum amplitude in group 2 significantly increases to 2434 V/m due to the complex cavity environment reducing the propagation losses.

Tab. 1. Group 1 vs. group 2.

Parameters		Group 1	Group 2
Fig. 10	E_z first peak	713.4 V/m	2434 V/m
	E_z second peak	289.7 V/m	919.2 V/m
	SLL	2.46	2.64
Fig. 11	E_z first peak	988.8 V/m	2434 V/m
	E_z second peak	932.1 V/m	1260 V/m
	SSLL	1.06	1.93

Another possible explanation is that the processed corners and the obstacle have increased propagation complexity, relatively decreasing the EM waves' resonance losses in the rectangular cavity. Group 2 also offers better spatial focalization quality with an SSLL value of approx. 1.93. Most of the emitted energy is refocused at the position of the initial source, except for the boundary loss.

To further validate the modeling approach and the results obtained, two other parallel simulations of group 3 and group 4 are performed by randomly changing the position of the inserted obstacles. The results are shown in Table 2.

It should be highlighted that, compared with group 1, the first maximum amplitude of both group 3 and group 4 increases significantly, which fits in with our earlier observations, proving that the amplitude of the focalization point is positively correlated with cavity complexity. Also, the SLL values are equal to 2 or higher for all groups, meaning that the TR technique's temporal focalization quality is not easily influenced by the external environment, which is an inherent advantage of the TR technique itself. One may also notice that SSLL values are greatly improved for both group 3 and group 4, resulting in better spatial focalization quality.

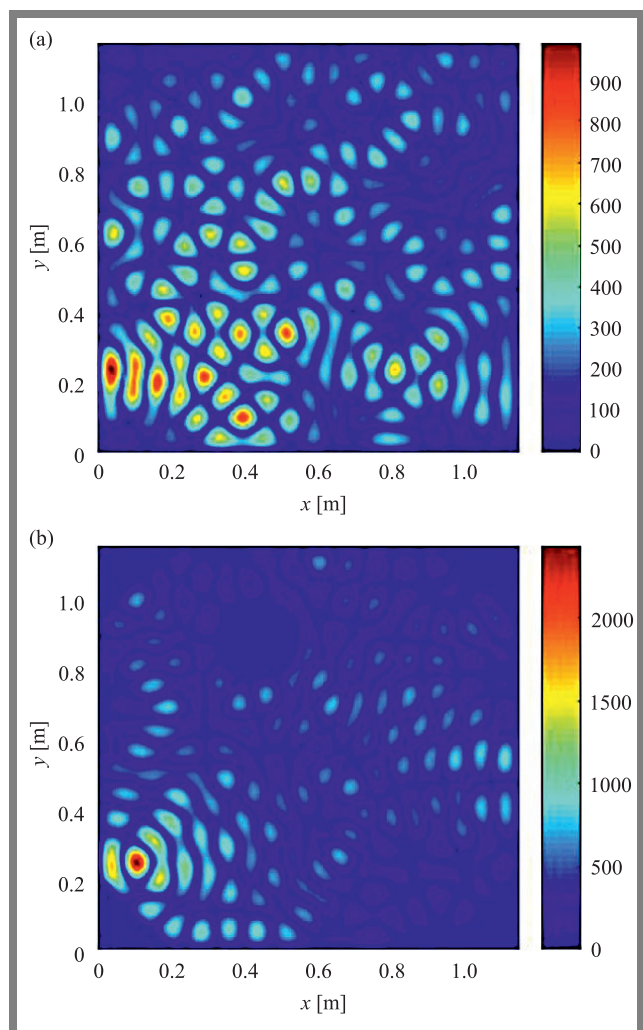


Fig. 9. Focalization point A for E_z at 604.03 ns: a) group 1 and b) group 2.

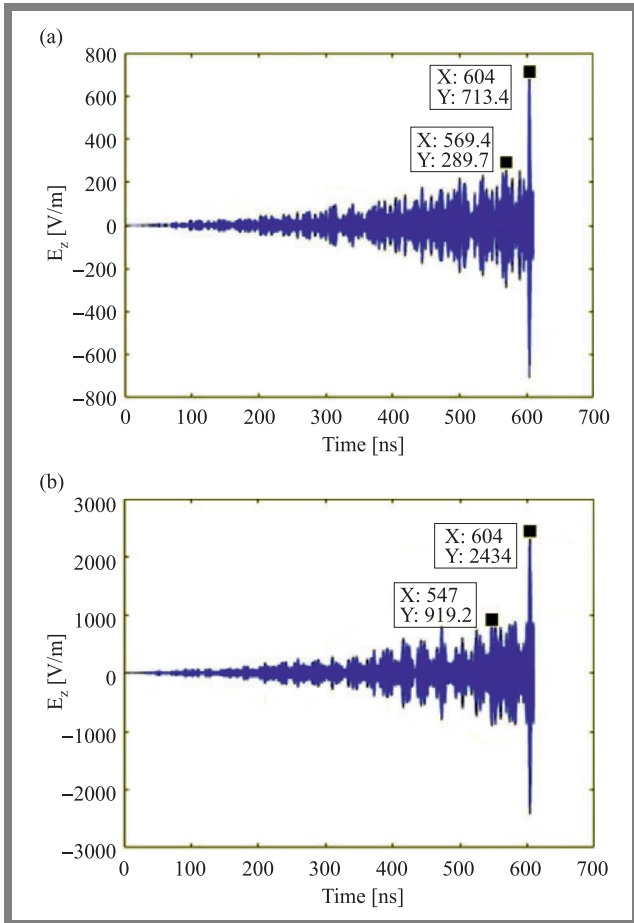


Fig. 10. Signal recorded at focalization point A: a) group 1 and b) group 2.

Three additional sets of simulations were completed by randomly changing the TRM position ten times in order to assess the effect of the modeled cavity on the focalization quality. For each set of simulations, the position of the initial source A remains fixed. In addition, the inserted obstacle and the processed corner positions are the same for all simulation groups, and the cavity is left untreated for all reference groups. SLL and SLL results are shown in Fig. 12.

The yellow/blue dots and the purple/green dots represent the SLL and SLL values for each independent simulation, respectively. The yellow/purple dots and blue/green dots correspond to the results of the reference and simulation groups. The black and red lines refer to the average values of SLL and

Tab. 2. Results of repeated simulations with randomly changed obstacle location.

Parameters	Group 1	Group 3	Group 4
E_z first peak	713.4 V/m	1990 V/m	1946 V/m
E_z second peak	289.7 V/m	935.6 V/m	993.8 V/m
SLL	2.46	2.13	1.96
E_z first peak	988.8 V/m	1990 V/m	1990 V/m
E_z second peak	932.1 V/m	1322 V/m	1104 V/m
SLL	1.06	1.51	1.80

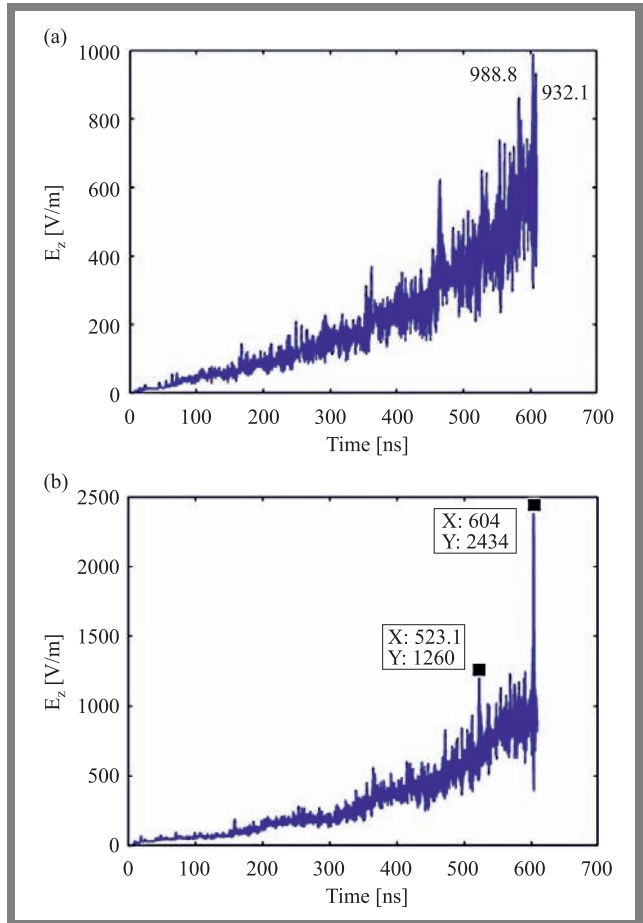


Fig. 11. Maximum amplitude recorded inside the cavity: a) group 1 and b) group 2.

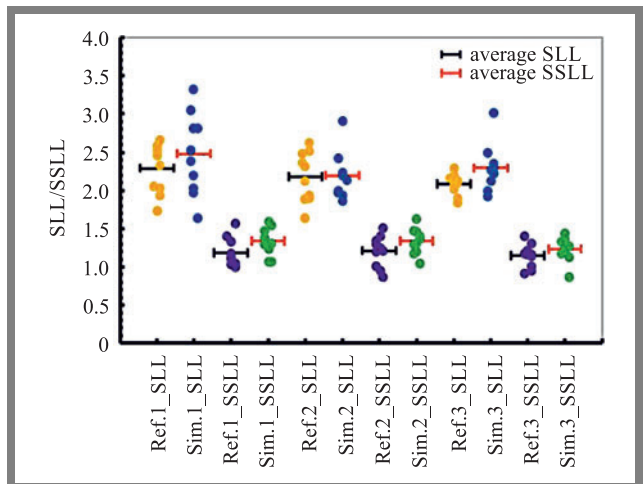


Fig. 12. Average SLL and SLL of the reference and simulation groups by randomly changing the location of TRM B.

SLL. What is striking about this chart is that the simulation group reported significantly more SLL and SLL average values than the reference group, which further verifies the positive correlation between the cavity's complexity and the focalization quality. However, for a single simulation, the difference observed in this study was not significant. This discrepancy could be attributed to the cavity's ergodicity, mean-

ing that our cavity is not complex enough. The round obstacle and arc-shaped corners cannot fully guarantee an improved focalization quality at all locations throughout the cavity.

4. Conclusion

This study was undertaken to model microwave cavities and obstacles and to realize the EM waves' spatio-temporal focalization in 2D. The most prominent finding that has emerged from the research concerned is that the derived SIBC-FDTD method's effectiveness and the original curvilinear modeling method. Obviously, when compared with other methods, such as conventional FDTD, PML-FDTD, etc., this approach will greatly reduce the computation complexity and CPU time due to the reasonable neglect of the outer solution space of the boundaries. Meanwhile, it has been one of the first attempts to thoroughly examine the SIBC-FDTD method in TR simulations. Multiple simulation results obtained in the course of this study also indicate that the processed corners and the inserted obstacle can significantly complicate the propagation environment, which will further improve focalization quality.

Despite these promising results, there is abundant room for further progress in determining the effect of different sizes, shapes of obstacles and even cavities on focalization quality. Meanwhile, further research should be undertaken to investigate the TR system's modeling using the SIBC-FDTD method in 3D.

References

- [1] G. Lerosey *et al.*, "Time reversal of electromagnetic waves and telecommunication", *Radio Science*, vol. 40, pp. 1–10, 2005 (DOI: 10.1029/2004RS003193).
- [2] Q. Li, C. He, Q. Zhang, and K. Cheng, "Passive time reversal based hybrid time-frequency domain equalizer for underwater acoustic communication", *2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–6, 2016 (DOI: 10.1109/ICSPCC.2016.7753713).
- [3] H. Karami, F. Rachidi, M. Azadifar, and M. Rubinstein, "An Acoustic Time Reversal Technique to Locate a Partial Discharge Source: Two-Dimensional Numerical Validation", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, pp. 2203–2205, 2020 (DOI: 10.1109/TDEI.2020.008837).
- [4] M. D. Hossain and A. S. Mohan, "A comparative study of coherent time reversal minimum variance beamformers for breast cancer detection", *2015 9th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, 2015 (<https://opus.lib.uts.edu.au/bitstream/10453/138599/4/Binder1.pdf>).
- [5] Y. Tao, T. Mu, and Y. Song, "Time reversal microwave imaging method based on SF-ESPRIT for breast cancer detection", *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 2094–2098, 2017 (DOI: 10.1109/CompComm.2017.8322906).
- [6] R. C. Qiu, C. Zhou, N. Guo, and J. Q. Zhang, "Time Reversal With MISO for Ultrawideband Communications: Experimental Results", *IEEE Antennas and Wireless Propagation Letters*, vol. 5, pp. 269–273, 2006 (DOI: 10.1109/LAWP.2006.875888).
- [7] H. Ma, B. Wang, Y. Chen, and K. J. Ray Liu, "Time-Reversal Tunneling Effects for Cloud Radio Access Network", *IEEE Transactions on Wireless Communications*, vol. 15, pp. 3030–3043, 2016 (DOI: 10.1109/TWC.2016.2515089).

- [8] P. Liao, B. Hu, Z. Lin, Q. Wen, and L. Zheng, "Effect of Signal Characteristics on Focusing Property of Time Reversal Electromagnetic Wave", *2019 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, pp. 1–3, 2019 (DOI: 10.1109/ICMMT45702.2019.8992282).
- [9] W. Lei and L. Yao, "Performance Analysis of Time Reversal Communication Systems", *IEEE Communications Letters*, vol. 23, pp. 680–683, 2019 (DOI: 10.1109/LCOMM.2019.2901484).
- [10] P. Kosmas and C. M. Rappaport, "FDTD-based time reversal for microwave breast cancer Detection-localization in three dimensions", *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 1921–1927, 2006 (DOI: 10.1109/TMTT.2006.871994).
- [11] H. Terchoune, *et al.* "Investigation of space-time focusing of time reversal using FDTD", *2009 IEEE MTT-S International Microwave Symposium Digest*, pp. 273–276, 2009 (DOI: 10.1109/MWSYM.2009.5165686).
- [12] X. Wei, W. Shao, S. Shi, Y. Cheng, and B. Wang, "An Optimized Higher Order PML in Domain Decomposition WLP-FDTD Method for Time Reversal Analysis", *IEEE Transactions on Antennas and Propagation*, vol. 64, pp. 4374–4383, 2016 (DOI: 10.1109/TAP.2016.2596899).
- [13] W. Fan, Z. Chen and W. J. R. Hoefer, "Source Reconstruction From Wideband and Band-Limited Responses by FDTD Time Reversal and Regularized Least Squares", *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, pp. 4785–4793, 2017 (DOI: 10.1109/TMTT.2017.2737991).
- [14] J. G. Maloney and G. S. Smith, "The use of surface impedance concepts in the finite-difference time-domain method", *IEEE Transactions on Antennas and Propagation*, vol. 40, pp. 38–48, 1992 (DOI: 10.1109/8.123351).
- [15] Y. Mao, A. Z. Elsherbeni, S. Li, and T. Jiang, "Surface impedance absorbing boundary for terminating FDTD simulations", *Applied Computational Electromagnetics Society Journal*, pp. 1035–1046, 2014 (<https://journals.riverpublishers.com/index.php/ACES/article/view/10807/9029>).
- [16] Y. Mao, A. Z. Elsherbeni, T. Jiang, and S. Li, "Mixed surface impedance boundary condition for FDTD simulations", *IET Microwaves, Antennas and Propagation*, vol. 11, pp. 1197–1202, 2017 (DOI: 10.1049/iet-map.2016.0649).
- [17] B. E. Anderson, M. Griffa, C. Larmat, T. J. Ulrich, and P. A. Johnson, "Time reversal", *Acoustics Today*, vol. 4, pp. 5–16, 2008 (<https://acousticstoday.org/time-reversal-brian-e-anderson/>).
- [18] H. Vallon, "Focusing High-Power Electromagnetic Waves Using Time-Reversal", *PhD thesis*, University of Paris Saclay, 2016 (<https://www.worldcat.org/title/focusing-high-power-electromagnetic-waves-using-time-reversal/oclc/948804731>).



Zhigang Li received his B.Sc. in Information Engineering from the Nanjing University of Aeronautics and Astronautics, China, in 2015, and his M.Sc. in Electronic Systems for Embedded and Communicating Applications from École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIH), France, in 2019. He is currently pursuing his Ph.D. at the University of Poitiers, France. His main research interests include electronics, electromagnetics, antennas, radar, and RF/mmW integrated circuit design.

E-mail: zhigang.li92@gmail.com

ESECA, Department of ENSEEIH, National Polytechnic, Institute of Toulouse, Toulouse, France



Younes Aimer received his B.Sc. and M.Sc. in Electronics and Telecommunications from the University of Saida – Algeria in 2011 and 2013, respectively. He joined the University of Saida and Poitiers, where he received his co-supervision Ph.D. degree in Signal Processing and Telecommunications – Electronics, Microelectronics, Nanoelectronics, and Microwaves in 2019.

He is presently a research engineer in Mines Saint-Etienne, Centre of Microelectronics in Provence at the Department of Flexible Electronics. His current research interests are digital communications, wireless and mobile communications, signal processing, electronic devices, IOT and wireless circuits.

E-mail: younes.aimer@emse.fr

Mines Saint-Etienne, Centre of Microelectronics in Provence, Department of Flexible Electronics, Gardanne, France



Tayeb H. C. Bouazza received his B.Sc. and M.Sc. in Electronics and Telecommunications from the University of Saida, Algeria in 2011 and 2013 respectively, and his Ph.D. in Electronics, Microelectronics, Nanoelectronics and Microwaves from XLIM Laboratory, University of Poitiers, France in 2021. His main research interests are in modelling nonlinear systems, signal processing and wireless communication.

E-mail: tayeb.habib.chawki.bouazza@univ-poitiers.fr

XLIM Laboratory UMR-CNRS 7252, Institute of Technology of Angouleme, University of Poitiers, Poitiers, France

Joint Optimization of Sum and Difference Patterns with a Common Weight Vector Using the Genetic Algorithm

Jafar Ramadhan Mohammed¹ and Duaa Alyas Aljaf²

¹Ninevah University, Mosul, Iraq,

²Mosul University, Mosul, Iraq

<https://doi.org/10.26636/jtit.2022.160722>

Abstract — A monopulse searching and tracking radar antenna array with a large number of radiating elements requires a simple and efficient design of the feeding network. In this paper, an effective and versatile method for jointly optimizing the sum and difference patterns using the genetic algorithm is proposed. Moreover, the array feeding network is simplified by attaching a single common weight to each of its elements. The optimal sum pattern with the desired constraints is first generated by independently optimizing amplitude weights of the array elements. The suboptimal difference pattern is then obtained by introducing a phase displacement π to half of the array elements under the condition of sharing some sided elements weights of the sum mode. The sharing percentage is controlled by the designer, such that the best performance can be met. The remaining uncommon weights of the difference mode represent the number of degrees of freedom which create a compromise difference pattern. Simulation results demonstrate the effectiveness of the proposed method in generating the optimal sum and suboptimal difference patterns characterized by independently, partially, and even fully common weight vectors.

Keywords — common weight vector, difference pattern, genetic algorithm, monopulse radar antenna, sum pattern

1. Introduction

Target searching and tracking in monopulse radar antennas requires simultaneous formation of both sum and difference patterns. The estimated angle of the target can be computed by dividing the difference pattern by the sum pattern [1]. To achieve high angular accuracy, the sum pattern must have a narrow main beam and low sidelobes. Primarily, these two factors are reversely related. Thus, there is always a tradeoff between the main beam width and the sidelobe level. Many numerical algorithms have been proposed in the literature for optimizing the excitation amplitudes and/or phases of the array elements to get the required sidelobe level under the desired beam width constraints. For example, see the approaches presented in [2]–[8].

On the other hand, the difference patterns should be also optimized to ensure low sidelobe levels in order to suppress the undesired interfering signals that could affect the angular estimate of the target's location. There are many techniques

in the literature that deal with such requirements, for example [9]–[11]. To cope with the requirements of optimal sum and difference patterns, ideally separate optimizations of two independent element weight vectors are needed [12]. However, these separate optimization methods were impractical, due to the use of two separate element weight vectors for one monopulse radar antenna. Moreover, their implementations are difficult and require a complex feeding network [13]–[14]. To tackle this problem, some researchers have been investigating the use of partially common weight vectors for optimizing sum and difference patterns. Ares *et al.* in [15] used simulated annealing to synthesize Taylor and Bayliss linear distributions with a common aperture tail. The common aperture tail technique has been successfully extended by the same authors [16] to the subarray configuration to obtain an optimum compromise between sum and difference patterns. Rocca *et al.* in [17] used convex optimization to optimally synthesize sum and difference patterns with arbitrary sidelobes and common excitation constraints. Then, the technique was further developed by the same authors to include the sparse theory [18] for the purpose of minimizing the number of array elements. In [19], Chun *et al.* used convex optimization again to synthesize asymmetric sum and difference patterns with a common complex weight vector. All the authors of the above-mentioned works assume that the problem is always convex and it can be solved by linear programming under some assumptions which cannot be sometimes satisfied especially when dealing with nonlinear problems such as phase-only synthesis problem and unequally spaced arrays. In fact, these synthesized problems are non-linear and non-convex and cannot be efficiently solved the use of with convex optimization methods. Therefore, global optimization approaches such as the genetic algorithm, particle swarm optimization, and evolutionary algorithms are more preferable than convex optimization methods due to the fact that they do not require any prior assumptions about the nature of the problems [20].

In [21], Keizer used iterative Fourier transform (IFT) to generate separately optimum sum and difference patterns, implicitly assuming that array elements are uniformly spaced at half the wavelength. In [22], Mohammed adopted the IFT technique to obtain the required sum and difference patterns

with a maximum allowable sharing percentage in the element excitations.

In light of the above discussions, there is a great need to find a general solution for optimizing the sum and difference patterns without any pre-specified assumptions or limitations. Also, both array patterns should be jointly optimized and the solution should be globally optimal. In this paper, a global genetic algorithm with a specific cost function that has the ability to jointly optimize both sum and difference patterns is considered to perform the required optimization process. The synthesis problem of the sum and difference patterns is first jointly formulated under the predefined constraint goals, such as beamwidth range, sidelobe envelope, and pattern nulling. Then, a single cost function is efficiently formulated to optimize the amplitudes of the common array weight vector. The proposed algorithm provides effective user-defined control over a wide range of sharing percentages ranging from 0% (i.e. independent weight vector) up to 100% (i.e. fully common weight vector).

A major novelty of our work is that the proposed optimization method can jointly synthesize arbitrary sum and difference patterns with a common weight vector, using a generalized genetic algorithm without any assumptions concerning convexity (as in [17]–[19]) and uniformity (as in [21]).

2. Problem Statement and Proposed Solution

Consider a linear array of an even number of isotropic elements $N = 2M$, which are equally spaced by $d = \lambda/2$ and symmetrically positioned with respect to the origin, along the x -axis. Also assume the indices of the elements on both sides of the array are: $-M, -M + 1, \dots, -1$ and $1, \dots, M - 1, M$ going from left-hand side to so right-hand side elements, as shown in Fig. 1. For such an antenna system, the array factor of the sum pattern can be written as [22]:

$$AF_{\text{Sum}}(u) = \sum_{n=-M}^{-1} a_n^s e^{j\frac{2n+1}{2}kdu} + \sum_{n=1}^M a_n^s e^{j\frac{2n-1}{2}kdu} \quad (1)$$

and in terms of the difference pattern, the array factor can be expressed by [22]:

$$AF_{\text{Dif}}(u) = \sum_{n=-M}^{-1} a_n^d e^{j\frac{2n+1}{2}kdu} - \sum_{n=1}^M a_n^d e^{j\frac{2n-1}{2}kdu}, \quad (2)$$

where $k = 2\pi/\lambda$, λ is the wavelength in free space, $u = \sin \theta$, and θ is the angle with respect to the normal direction of the array axis, and a_n^s and a_n^d are two separate weight vectors for the sum and difference patterns, respectively. First, the amplitude weights of a_n^s , $n = -M, \dots, M$ and $n \neq 0$ are independently optimized using the genetic algorithm to obtain the corresponding optimal sum pattern with the desired predefined constraints regarding main beam width, sidelobe level, and null control.

To proceed with the idea of the common weight vector, let us introduce a new parameter that specifies the common

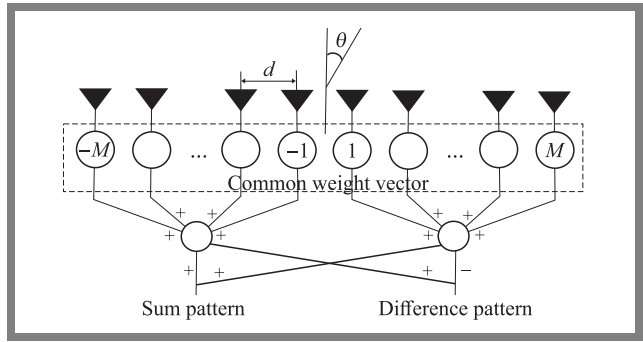


Fig. 1. Array configuration with a full common weight vector.

number of the array elements that share the same amplitude weights for the sum and difference modes, e.g. N_c . Thus, the remaining number of the uncommon array elements that will be available as degrees of freedom for optimizing the amplitudes of a_n^d to get the suboptimal difference pattern is $N - N_c$. To obtain the difference pattern, we also need to add a phase displacement of π to half of the elements of the array.

Now, depending on the chosen value of N_c , which is a user-defined parameter, three cases can be discussed. If $N_c = 0$, then two independent weight vectors for separately optimizing sum and difference patterns can be obtained. The second case involves a partially common weight vector with a certain sharing percentage that can be obtained for any value between $0 < N_c < N$. Finally, fully common weight vector between a_n^s and a_n^d can be obtained for $N_c = N$. Since the optimal amplitude weights for the sum and difference modes are usually very similar for the peripheral elements, one is inclined to start the value of N_c successively from the array ends. In this way, the searching spaces of the genetic optimizer are restricted, which helps significantly reduce the convergence speed of the optimizer and limit its run time by avoiding unnecessary random combinations of the array elements.

For symmetric amplitudes, the array factor of the difference pattern can be rewritten to include parameter N_c as follows:

$$AF_{\text{Dif}}(u) = \underbrace{\sum_{n=1}^{M-\frac{N_c}{2}} a_n^d \cos \left[\frac{2n-1}{2}kdu \right]}_{\text{Uncommon part } a_n^d \neq a_n^s} + \underbrace{\sum_{n=M-\frac{N_c}{2}+1}^M a_n^s \cos \left[\frac{2n-1}{2}kdu \right]}_{\text{Common part } a_n^d = a_n^s}. \quad (3)$$

Now, amplitudes of the two weight vectors a_n^s and a_n^d can be identified by minimizing the following cost function:

$$\text{cost} = \sum_{i=1}^I |AF_{\text{Sum}}(u_i) - UB_{\text{Sum}}(u_i)|^2 + \sum_{j=1}^J |AF_{\text{Dif}}(u_j) - UB_{\text{Dif}}(u_j)|^2, \quad (4)$$

Subject to:

$$|AF_{\text{Sum}}(u_o)|^2 = 1, \quad (5)$$

$$\text{if enforce symmetry then } a_n^s = a_{-n}^s, \quad a_n^d = -a_{-n}^d \quad (6)$$

for $n = 1, \dots, M$, and

$$a_n^d = a_n^s \text{ for } M - N_{c/2} + 1 \leq n \leq M. \quad (7)$$

u_i and u_j represent the sampling points in the sidelobe region (i.e. exempting the main beam region) of the sum and difference patterns, respectively, i and j represent the patterns points for the sum and difference modes, I and J are the total pattern points which are both set to be equal to 512 with evenly spaced in u space, UB_{Sum} and UB_{Dif} are the upper bounds of the constraint masks including the upper sidelobe envelope of each pattern, u_o indicates the target direction in the sum pattern. Note that the first null-to-null beam widths of the sum and difference patterns are included in the constraint masks UB_{Sum} and UB_{Dif} of Eq. (4).

To obtain a simultaneous nulling capability in both sum and difference patterns while also maintaining the same sidelobe structures, the constraint masks of UB_{Sum} and UB_{Dif} in the sidelobe regions are set at the same levels of -30 dB. Moreover, the first and second term of Eq. (4) can be separated, to allow each of them to act as a single independent cost function for the design constraints given in Eqs. (5), (6), and (7).

Clearly, the cost function from Eq. (4) is the sum of the squares of the excess pattern magnitudes exceeding the specified sidelobe mask. This minimizes the excess sidelobe power of both patterns outside specified upper-bound goals. Generally, a better solution may be obtained for lower cost values and the optimization process will be considered as converged when the cost value becomes lower than a specific threshold value which is chosen here to be -40 dB [22].

According to the cost function (4) and for given sum and difference patterns, each pattern points i and j that lie outside the specified sidelobe bounds contribute a certain value to the cost function equal to the power difference between the upper bound goal and the generated patterns.

3. Simulation Results

To validate the effectiveness of the described method, a number of numerical experiments have been performed. In the following examples, the synthesis of equally spaced linear arrays composed of $N = 20$ and $M = 100$ elements is considered. For the genetic optimizer, an initial population of 50 random array weights is generated and is evolved for 10,000 generations. Then, 25 pairs of parents are chosen by means of a tournament at each iteration to produce 2 children using 2 crossovers. Thus, the number of produced children becomes 50. From the total of 100 individuals, only best 50 survive to the next generation. This process repeats until a specified number of iterations is reached [22].

Next the sum and difference patterns are generated by jointly optimizing the amplitude weights of the sum and difference modes. Since the amplitude weights are assumed to be sym-

metric with respect to the center of the array, only half of the weights need to be optimized. The amplitudes are restricted to lie between 0 and 1 and phases between -180° and 180° (for the difference mode only).

In the first example, the synthesis of a linear array comprising $N = 2$ and $M = 20$ elements to independently generate optimum sum and difference patterns with two separate weight vectors a_n^s and a_n^d (i.e., the number of the common array elements $N_c = 0$) is considered. This case is considered as a benchmark for comparing other upcoming cases. The upper bounds of the sidelobe envelope of each pattern, i.e. UB_{Sum} and UB_{Dif} are set to -30 dB. The first null-to-null beam width (FNBW) of the sum pattern is constrained to be $u = \pm \frac{1}{N \cdot d}$, while that of the difference pattern is doubles. In addition, simultaneous two symmetric notches in the sum and difference sidelobe patterns centered at $u = \pm 0.57$ and ranging from $u = \pm 0.54$ to $u = \pm 0.6$ are placed.

The optimized sum and difference patterns along with their corresponding weights and cost functions are shown in Fig. 2. Clearly, implementation of such a feeding network system with these two separate weights (i.e. without a common weight vector) requires a large number of RF attenuators and phase shifters, equaling approx. From Fig. 2b, it can be observed that the optimum values of the excitation weights of the sum and difference modes are very similar to each other at both ends of the array. Accordingly, the amplitude weights of the side elements may be shared without any loss in system performance.

Table 1, shows the numerical results for the sharing percentage starting from 0 and reaching 100% (i.e. fully common weight vector), in incremental steps of 10% in each of the cases. For each case, performance measures related to taper efficiency, angle sensitivity K_r (1/rad), directivity, peak sidelobe level (i.e. peak sidelobe with respect to the maximum main beam), average sidelobes (i.e. area under the entire sidelobe region), and first null-to-null beam width (FNBW) are included. Further, the optimized weight vectors for the sum and difference patterns and for each considered case are presented as well. It can be observed that the greater the percentage of sharing weights, the poorer the difference side lobe pattern. Moreover, the remaining performance measures are slightly reduced as well when compared to the optimum values from the $N_c = 0\%$ scenario. The optimized sum and difference patterns for the two specific cases are highlighted in the following two examples.

In the second example, the results for the case of $N_c = 60\%$ and for a total number of array elements equal to $N = 20$ (i.e. 12 elements on both sides of the array are common for sum and difference modes) are shown in Fig. 3. In this case, the amplitude weights for the sum pattern are fixed at optimum levels (i.e. the same as in Fig. 1). The uncommon amplitude weights of the difference mode are optimized according to the cost function that was given in Eq. (4). Accordingly, little change in the sidelobe envelope of the difference pattern is noticed (Fig. 3a). However, the main beam shape and the null placement remain unchanged. Although the peak sidelobe level of the resulting difference pattern, -28 dB,

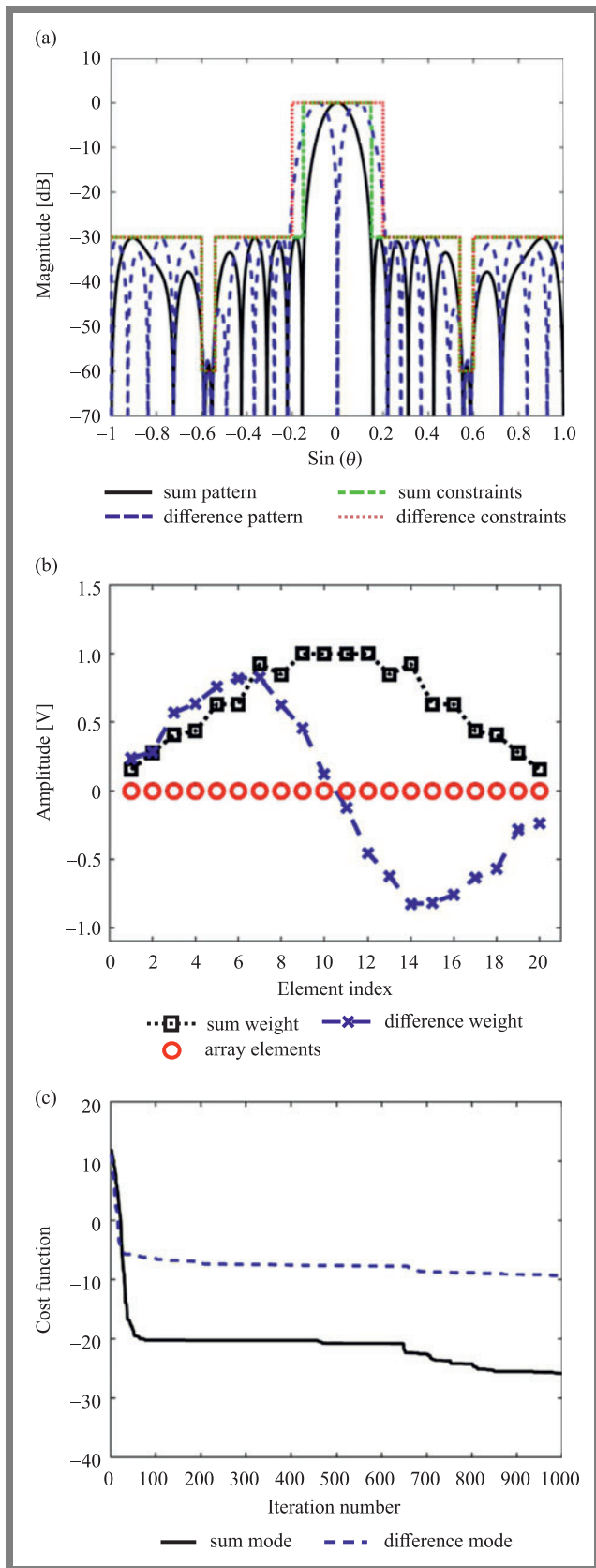


Fig. 2. Sum and difference patterns (a), their corresponding amplitude weights (b), and the cost function (c) for $N = 20$ and $N_c = 0\%$ (i.e. separate weights).

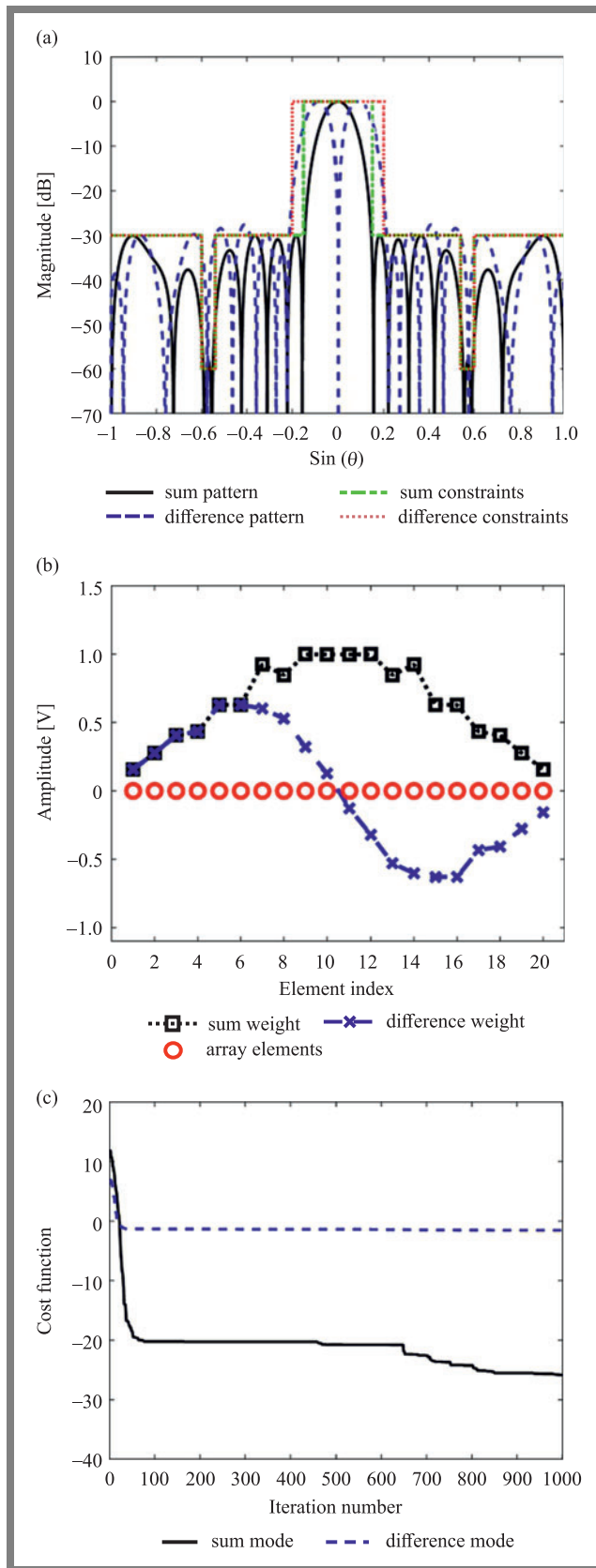


Fig. 3. Sum and difference patterns (a), its corresponding amplitude weights (b), and the cost function (c) for $N = 20$ and $N_c = 60\%$ (i.e. partially common weights).

Tab. 1. Performance of the optimized sum and difference patterns.

Performance	Optimized sum pattern	Optimized difference pattern											
		Common weight percentages N_c [%]											
		0	10	20	30	40	50	60	70	80	90	100	
Taper efficiency	0.82	0.53	0.51	0.51	0.52	0.53	0.53	0.52	0.47	0.45	0.41	0.35	
K_r [1/rad]		0.78	0.73	0.73	0.76	0.77	0.78	0.77	0.70	0.67	0.64	0.58	
Directivity [dB]	10.25	8.41	8.24	8.21	8.35	8.40	8.41	8.37	7.92	7.74	7.47	6.94	
Peak SLL [dB]	-30	-30	-30	-30	-30	-30	-29	-28	-25	-28	-22	-15	
Average SLL [dB]	-35.86	-33.57	-34.37	-35.31	-35.14	-32.06	-32.50	-32.74	-27.82	-27.25	-27.50	-18.46	
FNBW [deg]	0.29	0.41	0.43	0.44	0.42	0.41	0.41	0.42	0.71	0.74	0.84	0.84	
Weights	$a_n^S = a_{-n}^S$	$a_n^d = -a_{-n}^d$											
	0.15	0.23	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
	0.27	0.28	0.24	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
	0.40	0.56	0.49	0.46	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
	0.43	0.63	0.54	0.68	0.64	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	0.62	0.75	0.71	0.70	0.66	0.65	0.62	0.62	0.62	0.62	0.62	0.62	0.62
	0.62	0.81	0.70	0.86	0.87	0.57	0.56	0.62	0.62	0.62	0.62	0.62	0.62
	0.92	0.82	0.77	0.77	0.76	0.62	0.61	0.60	0.92	0.92	0.92	0.92	0.92
	0.84	0.62	0.59	0.70	0.66	0.48	0.46	0.52	0.70	0.84	0.84	0.84	0.84
	1.00	0.45	0.40	0.41	0.37	0.32	0.32	0.32	0.56	0.63	1.00	1.00	1.00
0.99	0.12	0.13	0.21	0.18	0.11	0.11	0.12	0.17	0.24	0.15	0.99	0.99	

Tab. 2. Comparison with other papers.

Method	Complexity reduction	Sharing percentage	Peak SLL for sum pattern	Peak SLL for difference pattern	Pre-assumption	Optimum solution
Independent weight vector method [12] and [13]	0%	0%	-30 dB	-30 dB	Two separate weight vectors	Yes
Alvarez method [15]	25%	50%	-30 dB	-23.8 dB	It uses pre-fixed Taylor and Bayliss distributions	Yes
Morabito and Rocca method [17]	30%	60%	-28 dB	-24 dB	The problem should be convex	Not always
Mohammed method [22]	30%	60%	-24 dB	-15.32 dB	Uses FFT and the inter-element spacing is uniform	No
Proposed	30%	60%	-30 dB	-28 dB	It doesn't need any assumption	Yes

is higher than, the prescribed mask limit of -30 dB, the average sidelobes of the resulting difference pattern amount to -32.74 dB, i.e. are lower than the mask limit. Moreover, complexity of the feeding network is reduced by more than half with respect to the first case (i.e. $N_c = 0\%$) with separate weights. In addition, the cost functions of this case were found to be satisfactory (Fig. 3c).

In the third example, the results for the case of $N_c = 100\%$ and for a total number of array elements equal to $N = 20$ (i.e. fully common weight vectors) are shown in Fig. 4. In this case, all the weights of the difference mode are enforced to be the same as those of the sum mode with a phase shift

equal to π . For such a specific case, there was a sudden change in the amplitude weights of the difference mode at the central elements of the array. This sudden change causes relatively high sidelobes in the difference pattern (Fig. 4a) and an unsatisfactory cost function (Fig. 4c).

Next, a larger array composed of $N = 100$ elements is considered. $N_c = 60\%$ and the levels of the UB_{Sum} and UB_{Dif} are set at the same levels as in the previous examples. The results are shown in Fig. 5 and match the observations from Fig. 3. This proves the generality of the proposed idea.

In the last example, the proposed method is compared with other published papers in terms of complexity reduction and

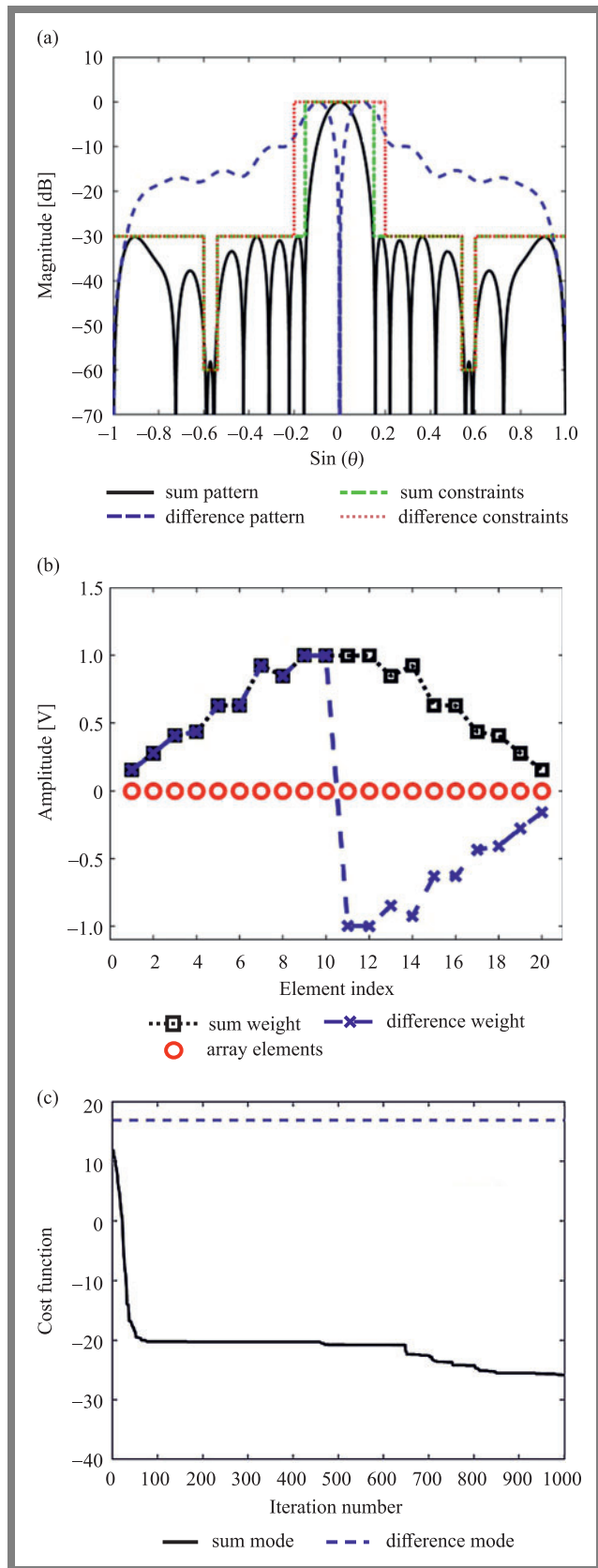


Fig. 4. Sum and difference patterns (a), its corresponding amplitude weights (b) and the cost function (c) for $N = 20$ and $N_c = 100\%$.

peak sidelobe level in the obtained difference pattern. The comparison is shown in Table 2.

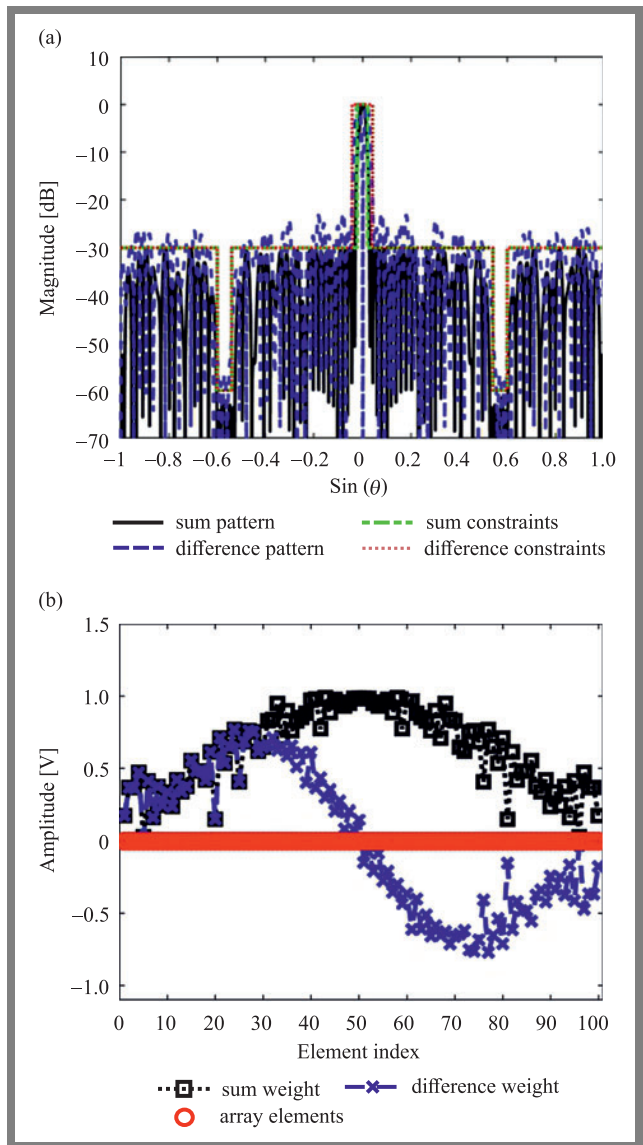


Fig. 5. Sum and difference patterns (a), and its corresponding amplitude weights (b) for $N = 100$ and $N_c = 60$.

4. Conclusions

In large tracking radar antenna arrays, complexity of the feeding networks is a major challenge. Thus, it is highly desired to simplify the feeding network as much as possible while generating the required sum and difference patterns. In a fully common weight vector case, where a single common attenuator and phase shifter is attached to each element for both sum and difference patterns, a significant reduction in the feeding network's complexity may be obtained by efficiently adjusting its amplitude and phase. However, this advantage comes at the cost of higher sidelobe levels in the difference pattern.

The problem of the high sidelobe level in the difference pattern was solved by using a partially common weight vector instead of its full counterpart and the complexity was found to be acceptable. It is found from the simulation that the sidelobe level of the difference pattern was reduced from -15 dB to more than -28 dB when switching from the fully

common weight vector to the partial one, with a sharing percentage of 60%. Also, it is found that the performance metrics of the difference pattern in terms of taper efficiency, angle sensitivity, directivity, average sidelobe, and beam width were reduced with an increase in the sharing percentage. The partially common weight vector of up to 60% was found to be an excellent choice for practical implementations.

References

- [1] M. I. Skolnik, "Radar Handbook", *McGraw-Hill*, 2008 (ISBN: 9780071485470).
- [2] J. R. Mohammed and K. H. Sayidmarie, "Sidelobe Cancellation for Uniformly Excited Planar Array Antennas by Controlling the Side Elements", *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 987–990, 2014 (DOI: 10.1109/LAWP.2014.2325025).
- [3] A. Safaai-Jazi and W. L. Stutzman, "A New Low-Sidelobe Pattern Synthesis Technique for Equally Spaced Linear Arrays", *IEEE Trans. On Antennas & Propagation*, vol. 64, no. 4, pp. 1317–1324, 2016 (DOI: 10.1109/TAP.2016.2526084).
- [4] J. R. Mohammed and K. H. Sayidmarie, "Synthesizing Asymmetric Sidelobe Pattern with Steered Nulling in Non-uniformly Excited Linear Arrays by Controlling Edge Elements", *International Journal of Antennas and Propagation*, vol. 2017, 2017 (DOI: 10.1155/2017/9293031).
- [5] J. R. Mohammed, "Obtaining Wide Steered Nulls in Linear Array Patterns by Controlling the Locations of Two Edge Elements", *AEU International Journal of Electronics and Communications*, vol. 101, pp. 145–151, 2019 (DOI: 10.1016/j.aeue.2019.02.004).
- [6] S. Koziel and A. Pietrenko-Dąbrowska, "Accelerated Gradient-Based Optimization of Antenna Structures Using Multifidelity Simulations and Convergence-Based Model Management Scheme", *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 12, pp. 8778–8789, 2021 (DOI: 10.1109/TAP.2021.3083742).
- [7] S. Koziel and A. Pietrenko-Dąbrowska, "Reliable EM-Driven Size Reduction of Antenna Structures by Means of Adaptive Penalty Factors", *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 2, pp. 1389–1401, 2022 (DOI: 10.1109/TAP.2021.3111285).
- [8] K. H. Sayidmarie and J. R. Mohammed, "Performance of a Wide Angle and Wideband Nulling Method for Phased Arrays", *Progress in Electromagnetics Research M*, vol. 33, pp. 239–249, 2013 (DOI: 10.2528/PIERM13100603).
- [9] J. R. Mohammed and K. H. Sayidmarie, "Performance Evaluation of the Adaptive Sidelobe Canceller with Various Auxiliary Configurations", *AEU International Journal of Electronics and Communications*, vol. 80, pp. 179–185, 2017 (DOI: 10.1016/j.aeue.2017.06.039).
- [10] S. Koziel and A. Pietrenko-Dąbrowska, "Expedited Acquisition of Database Designs for Reduced-Cost Performance-Driven Modeling and Rapid Dimension Scaling of Antenna Structures", *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 8, pp. 4975–4987, 2021 (DOI: 10.1109/TAP.2021.3074632).
- [11] S. Koziel and A. Pietrenko-Dąbrowska, "Robust Parameter Tuning of Antenna Structures by Means of Design Specification Adaptation", *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 12, pp. 8790–8798, 2021 (DOI: 10.1109/TAP.2021.3083792).
- [12] R. Haupt, "Simultaneous nulling in the sum and difference patterns of a monopulse antenna", *IEEE Transactions on Antennas and Propagation*, vol. 32, no. 5, pp. 486–493, 1984 (DOI: 10.1109/TAP.1984.1143352).
- [13] T. A. Milligan, "Bayliss line-source distribution", *Modern Antenna Design*, vol. 7, Section 4, pp. 158–161, 2005 (<http://www.radio-astronomy.org/library/Antenna-design.pdf>).
- [14] J. R. Mohammed, "Optimal Null Steering Method in Uniformly Excited Equally Spaced Linear Array by Optimizing Two Edge Elements", *Electronics Letters*, vol. 53, no. 13, pp. 835–837, 2017 (DOI: 10.1049/el.2017.1405).
- [15] M. Alvarez-Folgueiras, J. Rodriguez-Gonzales, and F. Ares-Pena, "Synthesizing Taylor and Bayliss linear distributions with common aperture tail", *Electron. Lett.*, vol. 45, no. 11, pp. 18–19, 2009 (DOI: 10.1049/el:20093322).
- [16] M. Alvarez-Folgueiras, J. Rodriguez-Gonzales, and F. Ares-Pena, "Optimal compromise among sum and difference patterns in monopulse antennas: use of subarrays and distributions with common aperture tail", *Journal of Electromagnetic Waves and Applications*, vol. 23, no. 17–18, pp. 2301–2311, 2009 (DOI: 10.1163/156939309790416206).
- [17] A. F. Morabito and P. Rocca, "Optimal synthesis of sum and difference patterns with arbitrary sidelobes subject to common excitations constraints", *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 623–626, 2010 (DOI: 10.1109/LAWP.2010.2053832).
- [18] A. F. Morabito and P. Rocca, "Reducing the number of elements in phase-only reconfigurable arrays generating sum and difference patterns", *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1338–1341, 2015 (DOI: 10.1109/LAWP.2015.2404939).
- [19] S. Kwak, J. Chun, D. Park, Y. K. Ko, and B. L. Cho, "Asymmetric Sum and Difference Beam Pattern Synthesis with a Common Weight Vector", *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1622–1625, 2016 (DOI: 10.1109/LAWP.2016.2519530).
- [20] D. W. Boeringer and D. H. Werner, "Particle Swarm Optimization versus Genetic Algorithms for Phased Array Synthesis", *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 3, pp. 771–779, 2004 (DOI: 10.1109/TAP.2004.825102).
- [21] W. P. M. N. Keizer, "Fast low sidelobe synthesis for large planar array antennas utilizing successive fast Fourier transforms of the array factor", *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 715–722, 2007 (DOI: 10.1109/TAP.2007.891511).
- [22] J. R. Mohammed, "Synthesizing Sum and Difference Patterns with Low Complexity Feeding Network By Sharing Element Excitations", *International Journal of Antennas and Propagation*, vol. 2017, Article ID 2563901, 2017 (<https://downloads.hindawi.com/journals/ijap/2017/2563901.pdf>).



Jafar Ramadhan Mohammed received his B.Sc. and M.Sc. degrees in Electronics and Communication Engineering from Mosul University, Iraq in 1998, and 2001, respectively, and a Ph.D. degree in Digital Communication Engineering from Panjab University, India in 2009. He is currently a Professor and Vice Chancellor for Scientific Affairs at Ninevah University, Mosul. His main research interests are in the area of digital signal processing and its applications, antennas and adaptive arrays.

E-mail: jafar.mohammed@uoninevah.edu.iq
Ninevah University, Mosul, Iraq



Duaa Alyas Aljaf received her B.Sc. degree in Electrical Engineering from the College of Engineering, Mosul University, in 2016 and M.Sc. in Communication Engineering from the College of Engineering, Al-Nahrain University in 2019. Currently, she works at the Autonomous Control Department at the Mosul Dam. She is currently pursuing her Ph.D. at the College of Engineering, Mosul University. Her research interests include antenna arrays, array pattern optimization for 5G systems and beyond.

E-mail: doaa.engineering@gmail.com
Mosul University, Mosul, Iraq

Semantic Knowledge Management and Blockchain-based Privacy for Internet of Things Applications

Manal Lamri¹ and Lyazid Sabri^{1,2}

¹Faculty of Mathematics and Informatics, University of Mohamed El Bachir EL Ibrahimi, Algeria,

²Laboratory of Images, Signals and Intelligent Systems, University of Paris-Est, France

<https://doi.org/10.26636/jtit.2022.161522>

Abstract — Design of distributed complex systems raises several important challenges, such as: confidentiality, data authentication and integrity, semantic contextual knowledge sharing, as well as common and intelligible understanding of the environment. Among the many challenges are semantic heterogeneity that occurs during dynamic knowledge extraction and authorization decisions which need to be taken when a resource is accessed in an open, dynamic environment. Blockchain offers the tools to protect sensitive personal data and solve reliability issues by providing a secure communication architecture. However, setting-up blockchain-based applications comes with many challenges, including processing and fusing heterogeneous information from various sources. The ontology model explored in this paper relies on a unified knowledge representation method and thus is the backbone of a distributed system aiming to tackle semantic heterogeneity and to model decentralized management of access control authorizations. We intertwine the blockchain technology with an ontological model to enhance knowledge management processes for distributed systems. Therefore, rather than relying on the mediation of a third party, the approach enhances autonomous decision-making. The proposed approach collects data generated by sensors into higher-level abstraction using n-ary hierarchical structures to describe entities and actions. Moreover, the proposed semantic architecture relies on hyperledger fabric to ensure the checking and authentication of knowledge integrity while preserving privacy.

Keywords — hyperledger fabric, ontology, security of distributed system, spatio-temporal knowledge representation

1. Introduction

The design of distributed complex systems capable of improving the perception of the user's context and making the interaction between humans and systems/robots more natural, requires data privacy protection, management and the sharing of a common and intelligible understanding of the environment [1]–[3]. Several technologies have emerged to ensure privacy and to identify when unauthorized users are trying to access the system [4]–[6]. As an emerging technology, blockchain offers tools that: protect sensitive personal data, solve reliability issues by supplying a secure communication architecture in a distributed application design (e.g. Industry 4.0), and eliminate the need for mediation of third-

authority organizations. Moreover, blockchain technologies offer the concept of smart contracts (SM), e.g. rules and actions based on predefined scenarios. SM is self-executing using data spread within the blockchain network, thus providing a higher level of autonomy required in IoT applications.

In contrast, traditional network architectures and numerous traditional privacy protection schemes store data in a centralized server to ensure that data is not disclosed. Moreover, decentralized management models rely on encryption techniques (e.g. public and private keys), guaranteeing access control authorizations between several domains. However, as demonstrated in [7], due to computing power constraints associated with IoT devices, such a cryptography technique cannot be suitably managed in all the layers. Besides, the authors highlighted another main challenge that IoT application designers face, namely heterogeneity of devices and services. Thus, adequate performance at the communication and storage level is not really achievable.

Since devices are energy-constrained, the use of permissioned blockchain hyperledger fabric (HF) [8] seems to be more suitable for dealing with IoT requirements. HF allows access and process data in real time, instantaneously taking adequate decisions to tackle emergencies, such as malicious actions or fall detection. However, in the process of designing these blockchain-based applications, many issues need to be taken into consideration, e.g. heterogeneous data fusion.

Many works point out that [9]–[12] ontologies are one of the best solutions available today to share knowledge, as they ensure secure communication that preserves privacy, offers authentication mechanisms and supports semantic interoperability across heterogeneous information sources. Consequently, intertwining ontology with blockchain provides a good level of autonomy by sharing the related IoT data, addresses data heterogeneity issues, and allows reasoning about SM semantics. To fulfill these goals, we essentially need the following:

- **Post-processing sensor information.** An ontological model must consider both the static (physical objects) and the dynamic (events) layer. Here, the aim is to link entities (e.g. humans, robots, sensors) with ontological-grounding concepts.

- **Architecture for reasoning and decision-making systems.** It offers high interoperability and abstraction of the entities, with mechanisms for secure event/knowledge delivery, simultaneously ensuring their privacy and confidentiality. More precisely, it will be applied in integrated smart surveillance systems for physical and digital assets and personal safety.
- **Assuring data integrity.** Thanks to the HF platform, each entity has its own Blockchain identity, which ensures data integrity and authentication while preserving privacy.

Defined in this manner, this model allows a narrative representation of events and establishes implicit semantic relationships (causality, purpose, etc.) between events observed in the environment. More precisely, it models inter-dependencies between contexts and expresses knowledge of: who is the initiator (i.e. agent) of the event/action? The objective is to enrich the interpretation of the context to ensure better adaptation. This approach relies on two types kinds of ontologies of the narrative knowledge representation language (NKRL) [13]: a binary ontology known as HClass and an *n*-ary structure known as hierarchy temporal (HTemp).

The latter uses semantic predicates/roles to represent dynamic Knowledge pertaining to: what are the context-related events that have been observed? Has an action been performed? Which entity (beneficiary) derives a benefit from the completion of the event/action?

Let us consider a scenario devoted to monitoring an elderly person wearing a fall sensor. Depending on knowledge analysis, the robot accomplishes tasks under different environmental conditions and recognizes situations/contexts (e.g. falls). Distributed IoT devices assist the robot in localizing the elderly person. A concept defined in an HClass ontology

becomes identifiable from the data provided by the sensors. Moreover, based on principles of cryptography, HF secures access to the robot’s embedded camera to allow the hospital staff to evaluate the patient’s health condition during the wait for the paramedics.

The remainder of the paper is organized as follows. Section 2 presents the background of the approach. The ontological knowledge representation approach is presented and detailed in Sections 3 and 4. A functional and architectural description of the proposed framework is given in Section 5, where an exemplified case study is presented as well. Finally, experimental evaluations and conclusions are given in Sections 6 and 7.

2. Hyperledger Fabric Blockchain Basics

Blockchain structures data into blocks. Each block contains a transaction (several transactions) and all blocks are organized into a cryptographic chain. Each transaction is secured, authenticated, and added to a secure, immutable data chain. A consensus-based algorithm allows to add new blocks to the blockchain network. The consensus algorithm is responsible for data integrity in the blockchain network and prevents service attacks for double-spending [14]–[15].

Hyperledger fabric has the form of a permissioned network. It is limited to a set of users and the consensus is achieved through a selective endorsement process. Thus, HF offers the following advantages: it saves time, removes cost (overhead and cost components), reduces risks (tampering, fraud, and cybercrime), and increases trust level [16]. The blockchain network has a single view of the dataset, and each node (e.g. participant) stores its code. Below, we describe each

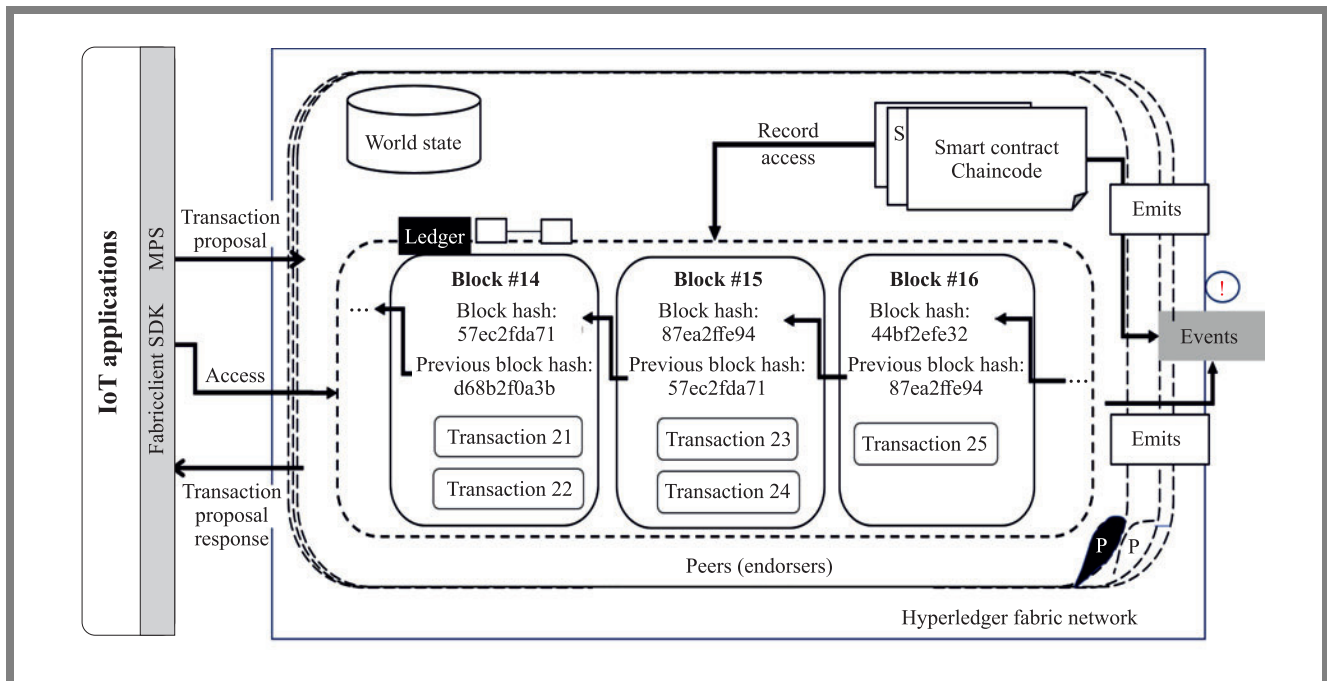


Fig. 1. Main components in a blockchain hyperledger fabric.

component (depicted in Fig. 1) and show how the cryptographic hash function secures the blockchain [17].

The ledger is a tamper-resistant of transitions. It stores the immutable blocks and the current state. The world state is seen as an ordinary database that stores and provides combined outputs of all transactions. Each participant within the blockchain network maintains a copy of the ledger. A cryptographic hash function accepts any binary data as input. A hash function (algorithm) is applied to the input data and generates an output whose length is determined in bytes (hash value), such as 57ec2fd71 (Fig. 1). The hash value serves as a digital fingerprint of that data.

A smart contract (i.e. chaincode) is signed encoded in the programming language. Unlike a traditional contract, SM defines the applicable rules in a blockchain. A chaincode is therefore a program that runs automatically in a blockchain to restrict actions or to transfer assets when specific conditions are met.

A peer network consists of the main component of a HF that belongs to a consortium. It stores the blockchain ledger, runs the SM, validates transactions and adds blocks to the ledger. Thereby, a peer network is seen as an overarching entity of the transactional flow.

Membership services authentication (MSP) is a process that allows to manage identities and authorize participants to access networks in a permissioned blockchain network. Since MSP handles all permissioned identities and knows all of the organization's members, it may recognize and trust each participant. That is why it is considered to be the backbone of the blockchain network.

Event handlers create notifications concerning significant blockchain operations and notifications related to smart contracts. All events include the transaction ID and allow the applications to take action when a transaction is concluded.

Fabric SDK allows the system to create, update and monitor blockchain components. The fabric client (made available through the fabric SDK) provides the `queryByChaincode()` API for developers to transmit a query request to a peer. Such a method is available on an instantiated channel object and takes two inputs. Besides using a chaincode query (queries implemented in a chaincode), a client application can send a request directly to the ledger, which is useful for retrieving metadata (for example instantiated chaincode) or for retrieving a specific transaction or block from the blockchain.

3. Ontological Knowledge Representation

The main difference when compared with a semantic web language, such as OWL, is that the proposed narrative knowledge representation language relies on the ontology of events called HTemp, in addition to the ontology of concepts. A template is an n -ary structure that enables a complex structured dynamic knowledge, e.g. "John falls and pushes the emergency button" to be represented. NKRL also defines the ontology of concept (HClass) that includes more than 3000 concepts. HClass is not different from frame-based or Protégé [18].

The conceptual representation of narrative knowledge is carried out through the HClass ontology of concepts. It is similar to traditional ontology languages, such as OWL or DAML+OIL. NKRL assimilates a concept to the notion of class in the semantic web. It is associated with a set of properties or attributes. NKRL distinguishes two categories of concepts: those that may be instantiated directly (`sortal_concept`) and those that cannot be instantiated directly (`non_sortal_concept`). Instantiable (i.e. instances) concepts, such as `Chair_125`, `Bed_2012`, `Tap_45`, etc., are created from the concepts `chair_`, `bed_`, `tap_` concepts.

The concept of color is an example of a concept that cannot be instantiated directly, as it cannot have a direct instance. Indeed, `Red_120` and `Yellow_1` cannot be considered instances, since they have no meaning if used separately. The solution provided by NKRL is to introduce the `color_appearance` concept, a specialization of the instantiable concept of `physical_appearance`. Thus, it is possible to associate a color with an instantiable concept, such as, for example, `Red_Table`, `Yellow_Cup`, etc.

In contrast, the ontology of events (i.e. HTemp ontology), as stated before, allows expressing knowledge concerning actions and events. We claim, therefore, that NKRL can enhance dynamic knowledge representation in the context of IoT applications. The conceptual narrative knowledge representation is structured into the following components:

Concepts component allows the representation of concepts. A concept is a binary representation of general notions (e.g. human-being, artifacts) or specific notions (doctor, sensor, chair). Formally, the knowledge representation of these notions is known as a concept. A concept is named using lower case symbolic labels with an "underscore", such as `human_being`, `artifact_`, `doctor_`, `sensor_`, `robot`.

Individuals component concerns the formal representation of instances (i.e. individuals) defined in the concepts component. Instances are created by instantiating the properties of concepts. Instances are labeled using the upper case, including the underscore symbol. `Motion_Sensor` and `Camera_1` are instances of the `sensor_concept`, and `Kitchen` is an instance of the `location_concept`.

n -ary component NKRL considers an elementary event as a spatial-temporal knowledge called a template or a predicative occurrence. Each template is expressed with the use of one predicate, specific roles and arguments. Figure 2 presents a hall structure of a predicative occurrence: Predicate that belongs to `MOVE`, `OWN`, `Exist`, `Produce`, `Receive`, `Experience`, `Behave`. Argument represents attributes that can be associated with each generic role, i.e. subject (`Subj`), object (`OBJ`), `Source`, `Modal`, `Topic`, `Context`, `Beneficiary` (`Benf`).

Factual components allow to represent events/actions extracted within a narrative as instances of the n -ary component. Each event allows to describe, for example, a situation and/or a robot-human interaction (e.g. `JOHN_` and the `ROBOT_KOMPAI`). Figs. 3-4 show the `PRODUCE` and `OWN` templates. The `location_concept` indicates the action's lo-

```

PREDICATE
  SUBJ {< argument >: [location]}
  OBJ {< argument >: [location]}
  SOURCE {< argument >: [location]}
  BENF {< argument >: [location]}
  MODAL {< argument >}
  TOPIC {< argument >}
  CONTEXT {< argument >}
           [modulators]
           [temporal attributes]

```

Fig. 2. General structure of an NKRL template.

```

name: Produce:Entity
  PREDICATE: PRODUCE
    SUBJ var1: [(var2)]
    OBJ var3
    [SOURCE var4: [(var5)]]
    [BENF var6: [(var7)]]
    [MODAL var8]
    [TOPIC var9]
    [CONTEXT var10]
    {[modulators], !=abs}
  var1 = < artefact_ > | < human_being >
  var2 = < location_ >
  var3 = < information_content >
  var4 = < human_being >
  var5 = < location_ >
  var6 = < human_being >
  var7 = < location_ >
  var8 = < temporal_development >
  var9 = < situation_ >
  var10 = < situation_ >

```

Fig. 3. Produce template structure.

```

name: Own:SimpleProperty
  PREDICATE: OWN
    SUBJ var1: [(var2)]
    OBJ var3
    [SOURCE var4: [(var5)]]
    [(BENF) var4]
    [MODAL var6]
    TOPIC var7
    [CONTEXT var8]
    {[modulators], != abs}
  var1 != < human_being|property_ >
  var2 = < location_ >
  var3 = < property_ >
  var4 = < human_being >
  var5 = < location_ >
  var6 = < artefact_ > | < temporal_sequence >
  var7 != < spatial_temporal_relation >
  var8 = < situation_ > | < label_ >

```

Fig. 4. Own template structure.

cation. Temporal attributes represented by symbolic labels represent the start or endpoints of the authorized action or

duration of the executed transaction. For this purpose, NKRL supplies two modulators: begin/end. The latter is a time stamp marking the beginning or end of an action/event. The obs modulator is used if no information about the beginning or the end of an action is given.

Figure 3 describes the produce template. A role or variable defined in square brackets is considered optional. The SUBJ, OBJ roles and the var1 and var3 variables are mandatory, while the BENF, MODAL, SOURCE, TOPIC and CONTEXT roles, and var6 and var7 variables, for example, are optional. Variables var1, . . . , var7 describe constraints that make it possible to check whether the values assigned to each variable when creating a predicate occurrence are specific to the terms (concept, instances) used in the Individuals component. Thus, the constraints defined in the templates of the HTemp ontology are associated with the concepts defined in the HClass ontology. Therefore, the knowledge consistency checking process relies on the HClass ontology to establish a hierarchy of concepts and instances based on the generalization/specialization principle.

4. n -ary Knowledge Representation

Formally, to create a predicative occurrence, the system evaluates the expression according to the n -ary structure described by: equation:

$$\text{Label } \text{")"} \text{ Predicat}_{i=1} \text{ (Roles}_{1 \leq j \leq 7} \text{ args)}, \quad (1)$$

where:

- label identifies a predicative occurrence. The label is a sequence of characters that matches the regular expression $[a-z][1-9].[a-z][1-9]$;
- $\text{Predicat}_{i=1}$ is one of the conceptual predicates $\in \{\text{MOVE, PRODUCE, RECEIVE, EXPERIENCE, BEHAVE, OWN, EXIST}\}$;
- $\text{Roles}_{1 \leq j \leq 7}$ is a conceptual role $\in \{\text{BENF, MODAL, TOPIC, CONTEXT, SUBJ, OBJ, SOURCE}\}$
- args – this attribute belongs to concepts and individuals components.

In terms of equivalence between the NKRL representation and the description logic [19], the individuals and concepts components correspond to the ABox and TBox components. However, there are no description language equivalents for the n -ary and factual components.

The OWN predicate allows to describe the type of the ownership notion between The entities or the state of an entity. Therefore, to express the fact that the front door was unlocked at 04/03/2022:9:56:15:362, the predicative occurrence aa11.c18 (Fig. 5) uses the SUBJ role with the FRONT_DOOR_BUTTON as an argument. The property unlocked (opened) is an argument of the TOPIC role, the obs modulator indicates that the starting time belongs to date-1 and is associated with the “front door has been unlocked” action. We highlight that each event requires that its beginning and end be distinguished. However, it is hard to establish or infer the end of an event in many scenarios. Nevertheless, deter-

mining the start of an event, as a minimum, is mandatory for further reasoning.

```

aal1.c18) PREDICATE: OWN
    SUBJ: FRONT_DOOR_BUTTON
    OBJ: property_
    TOPIC: unlocked_
           { obs }
    date-1: 04/03/2022:9:56:15:362
    date-2:
Own:SimpleProperty
aal2.c3) PREDICATE: PRODUCE
    SUBJ: CAMERA_1
    OBJ: detection_: (LOCATION_1)
    TOPIC: activity_
           { obs }
    date-1: 04/03/2022:10:31:20:102
    date-2:
Produce:Assessment/trial
aal3.c6) PREDICATE: OWN
    SUBJ: LIGHT_BUTTON_1: HALL_1
    OBJ: property_
    MODAL: lighting_: (switch_off, switch_on)
           { obs }
    date-1: 04/03/2022:10:31:22:523
    date-2:
Own:SimpleProperty
aal1.c14) PREDICATE: EXPERIENCE
    SUBJ: JOHN_
    OBJ: respiratory_distress
    MODAL: SENSOR_DISTRESS_1
           { obs }
    date-1: 04/03/2022:17:57:35:105
    date-2:
Own:SimpleProperty
    
```

Fig. 5. Examples of predicative occurrences.

The knowledge representation in NKRL allows to specify the date-1 attribute only, while the temporal attribute date-2 is empty. In turn, the predicative occurrence aal2.c3 (Fig. 5) expresses that the camera denoted as CAMERA_1 recorded some activity in LOCATION_1. The predicative occurrence aal3.c6 expresses that the light button localized in the hall, as denoted by HALL_1 changed its state from switch_off to switch_on.

While the embedded_sensor (subclass of artifact_) observes that a patient denoted by JOHN_ (aal1.c14, Fig. 5) displays an acute respiratory deficiency, it does not provide any information about the duration or end of this particular event. Such knowledge is expressed in NKRL using the experience templates. They are mainly used to express the fact that an entity is affected by an action. For example, the system observes a decrease in light intensity in a given space, a human suffering from an illness or an accident (e.g. a fall). While the predicative occurrence aal1.c14 (Fig. 5) expresses that the JOHN_ symbol, representing a human, is used as an argument

of the SUBJ(ect) role, the respiratory_distress property, being an argument of the OBJ(ect) role related to the date-1 attribute, is used to describe the beginning time-stamp of the action. The sensor denoted by SENSOR_DISTRESS_1 signals that John is suffering from a respiratory failure.

5. NKRL and Blockchain

In order to provide an informal example of the paper’s objectives, let us consider a scenario devoted to monitoring elderly persons at home. We assume that John is wearing a fall sensor used to detect the presence of an emergency alarm. Thus, the relevant contextual information considered in this use case includes the accurate location of John and his status (unconscious/conscious). The second piece of contextual information is not directly measurable, hence it is subjected to complex processes. Multiple events/actions must be correlated instantaneously to determine John’s status and assess the current context/situation. The first goal consists in understanding what is happening after an alarm has been triggered. So, the robot moves towards the last location of John and tries to interact with him (i.e. check his status). The robot tries to establish a dialogue-based interaction Fig. 6. If John does not interact with the robot, he is considered unconscious, and then this non-observable context corresponds to an emergency. In this case, the monitoring function should be able to deduce the status. The second goal consists in ensuring secure communication that complies with privacy and authentication mechanisms. In fact, the doctor from the hospital will check the patient’s health by observing the interaction between the robot and John. To do so, the doctor needs to remotely access the robot’s embedded camera. Decision-making is followed by actions, such as allowing the hospital staff to remotely access the indoor home security camera or the robot embedded camera to evaluate the patient’s condition.

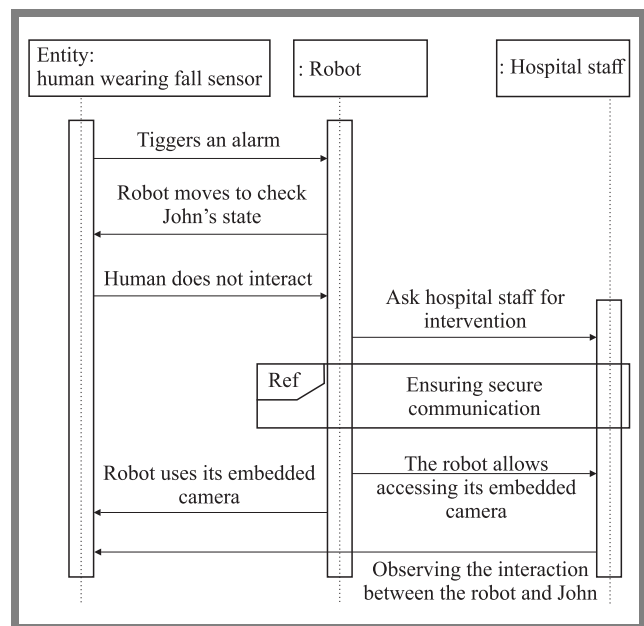


Fig. 6. Scenario sequence diagram.

5.1. HF and Distributed Complex Systems

According to [20], due to the complexity of conditions pertaining to the network serving as a platform for communicating with IoT devices, a robot can modify its internal structure and activity patterns in the self-organization process. IoT devices generate enormous amounts of data and do not have the computational power required. That is why Bitcoin and Ethereum are not suitable for addressing (i.e. managing) actions/contexts described in the scenario above, where an instantaneous reaction is needed. Moreover, IoT devices should mine and create blocks according to the POW-based (proof-of-work) protocol that calls for considerable amounts of energy [14]. However, HF is an open-source distributed ledger platform based on the Linux architecture. It establishes decentralized trust in a network. Only the data we intend to share are shared among the relevant participants, i.e. advanced privacy controls are ensured.

HF is permissioned blockchain and empowered building a consortium, meaning that the participants are identified and may not trust each other. It provides pluggable consensus protocols allowing organizations (multiparty) to customize their consensus protocols. Each participant controls one or more peers (nodes in the chain) and should treat a chaincode as unreliable, since anyone can dynamically deploy a smart contract. Moreover, HF can rely on Byzantine-fault tolerant (BFT) [21] instead of POW consensus algorithms and the execute-order-validate architecture. Therefore, HF enables scalability, i.e., sharing information and permitting IoT devices to execute actions in real-time, since any transaction is endorsed before being added to the chain and validated. Additionally, HF ensures privacy, security, and confidentiality, making it more suitable for meeting the requirements of IoT applications.

Figure 7 shows an overview the layers of an architecture merging blockchain and model ontologies. It comprises three weakly coupled software layers: facade communication com-

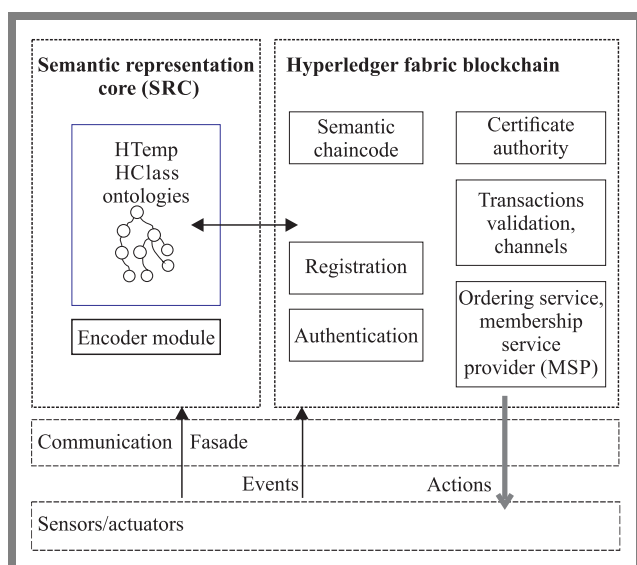


Fig. 7. Layers of an architecture merging blockchain and NKRL ontologies.

ponent, HF module, as well as HTemp and HClass ontologies. The facade communication component, seen as a set of interfaces, provides the unifying concept of service and a semantic description, i.e. hides the heterogeneity of the IoT devices. This layer works in a coherent and homogeneous semantic world linking the concepts defined in the HClass ontology with their real-world counterparts. The communication layer acts as an enterprise service bus (ESB), translating, each time, data generated by an authenticated IoT device to higher-level abstraction or creating commands, i.e. creating actions extracted from the smart contract.

5.2. Development Execution Environment

Thanks to the semantic abstraction level maintained by NKRL ontologies, the semantic representation core (SRC) module and the HF share the exact meaning of knowledge. The experiment conducted assumes that the consortium (participants) network consists of three organizations: robot, hospital, and home. Only the hospital cooperates with one peer (node) and one ordering node, while the robot and the home organization cooperate with two peers and two ordering nodes. The robot is responsible for setting up the blockchain network. It also has the privilege of creating channels and starts the ordering nodes. Only the channels between the robot and the home are private. Each channel has its ledger, which is replicated across other peers. These peers are integrated with the fabric network using certificate authority. Ordering peers receive blocks and generate validated transactions before committing a copy of the ledger to each peer. However, only the ordering peers within the robot and home organizations endorse peers, since they have the chaincode installed. The ordering and membership module (MSP) is the main components. Indeed, the MSP module maintains the cryptographic identities of all participants and links each IoT device to its identity. The ordering node allows the establishment of a consensus on transactions according to BFT algorithms. SMs deployed on channels running within the Docker generate an executable program.

The fabric SDK API is used to invoke the SMs from a client application. It allows to endorse transactions and to interact with the records on the blockchain ledger. The latter is composed of two components: the world state and the transaction log. The former represents a database of the ledger and is used to describe the state of the ledger at a given time. As for the transaction log component, it is the updated history for the world state.

The primary purpose of the authentication component is to ensure secure communication between network peers. This component relies on a public key infrastructure (PKI) to check the peers' cryptographic identities through authentication of a chain of trust and guarantee messages shared between peers involved in the interaction. By contrast, the MSP component uses the peer's public key to check each transaction that the peer should sign with its corresponding private key. Therefore, the identity checking mechanism enables, on the one hand, the node channels to establish MSPs to determine which IoT devices can perform actions. On

the other hand, it permits IoT devices to be trusted by each participant within the blockchain network.

Peers keep any data stored in the ledger. The IoT data are sent to endorsing peers and the facade communication component. After executing the SM, the peer responds to the application client if the transaction is endorsed (valid). The ordering service creates the block, and then the ledger is updated.

6. Implementation and Results

We implement the solution in three different environments of the same network. Fabric 2.3 was deployed as an underlying blockchain application, and we used node.js to write the chaincode and client applications. The proposed architecture ensures the homogeneity of the knowledge base. The relationships between real-world entities and their semantic representations are model-defined, meaning they allows for semantic matching. The model outputs the corresponding predicative occurrence. Therefore, the interface communication layer ensures a coherent representation of the real environment's states and the high-level abstraction. After executing the SM, and if the transaction is endorsed (valid), the peer responds to the application client. The ordering service creates the block, and then the ledger is updated.

After detection of the fall event, the corresponding chaincode installed on the robot peer is launched. Then, an authorization request to access the robot's camera is executed. Therefore, the communication layer enables converting the request into commands to access the robot's camera. Thanks to HF, the robot checks the hospital's identity using an MSP, abstracting all the cryptographic mechanisms, validating certificates and authenticating the user. After completing this process, the visual message action is endorsed, and the robot allows the hospital staff to access its embedded camera.

```

async writeData (ctx, key, time, sender,
                type, event, data) {
    const tmp =
        ID: key,
        SenderName: sender,
        SenderType: type,
        EventName: event,
        time: time,
        Data: data

    const buff = Buffer.from(JSON.stringify (tmp));
    await ctx.stub.putState (key, buff) ;
    return ctx.stub.setEvent (event, buff) ;
}

async readData (ctx,key)
    var response = await
ctx.stub.getState(key):
    return JSON.stringify (response);
}
    
```

Fig. 8. Chaincode implementation.

HF is modular, pluggable, and allows different consensus algorithms (e.g. RAFT and byzantine fault tolerant) to be used. Furthermore, HF relies, by default, on NoSql LevelDB to store public key values, and each entity has its own blockchain identity and registers only once. In the experiment, we used one ordering node only, because it can manage about 100 transactions per block.

In the experiments, we deployed the same chaincode business logic in the three channels connecting the endorsing peers of the network. The chaincodes implement mainly two functions to handle the reading and writing of data (Fig. 8).

6.1. Events Implementation

To test the proposed implementation, we submitted 100 transactions from sensors to the ordering peer which batched and sent all of them to the anchoring peers. We measured the time between submitting a transaction, including the date write in the chaincode and its commitment to the ledger. Figure 7 shows the endorsing time of all the transactions.

The obtained results show that the transactions are endorsed in less than one second in most cases, and the mean duration to endorse a transaction is 819.77 ms. We repeated the same experiment with five sensors emitting 100 transactions at a throughput rate of five transactions per second. In Fig. 10,

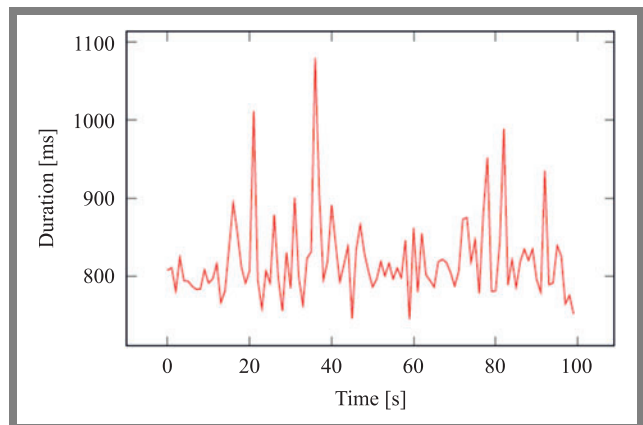


Fig. 9. Endorsing duration of 100 transactions from one sensor.

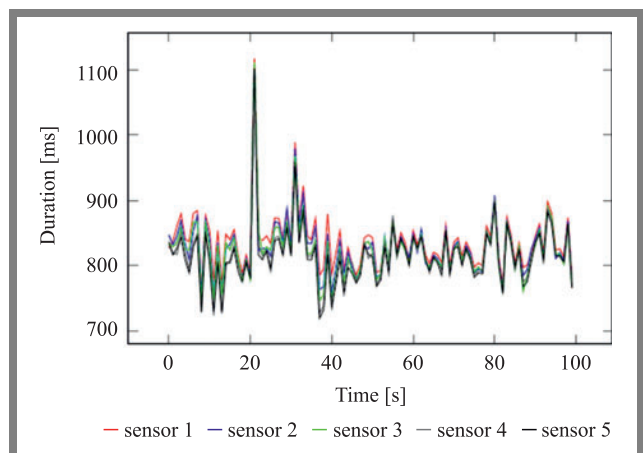


Fig. 10. Endorsing duration of 100 transactions from five sensors.

the resulting duration of the endorsement phase of this test is presented, with mean value duration equaling 824.37 ms. The broadcast chaincode events are captured by the applications connected to the channels using an event listener. The relevant action is inferred after parsing the payload data based on the event description and the sender type. The result of the inference is also submitted to the ledger by the creation of a transaction.

7. Conclusion and Discussion

IoT applications have to enforce security to control the data collected by scattered sensors. Privacy of data and monitoring of physical/virtual access to space or sensitive knowledge are among the many challenges faced by designers of distributed systems. Therefore, security and dynamic knowledge management are at the heart of the IoT application development process. The lack of interoperability and monitoring of physical/virtual access to space or sensitive data may endanger the adoption of the IoT paradigm. Nevertheless, it will be hard to ensure security and privacy for numerous IoT applications without providing a harvest ambient energy mechanism or reducing network latency (e.g. hyperledger fabric). Several studies show that qualitative and quantitative approaches have been adopted over the past years in connection with data processing, action and situation recognition [22]. Furthermore, many frameworks, projects, and techniques rely on a semantic mechanism to annotate and manage sensor data. The ontology web language (OWL) is de facto a solution that is most commonly used to express knowledge in IoT.

Several distributed applications, relying upon OWL, have been implemented. Even if these approaches offered some extensions and a built-in module to enrich the standard web semantic, their main weakness consisted in generating redundant knowledge descriptions. Many scenarios, such as those presented in this paper, need an n -ary structure that expresses actions/events and temporal properties. However, OWL and its variants have failed to address any proposals concerning the notion of n -relations. Thus, the entire semantic web language becomes de facto unsuitable for integrating heterogeneous IoT devices and addressing dynamic knowledge management requirements in IoT applications.

The work presented in this article aims to facilitate the implementation of access control for distributed systems. NKRL innovates by providing a hierarchy ontology of action. Indeed, the formalism we explore allows semantic descriptions of dynamic characteristics of the entities that frequently change overtime involved in IoT applications. Besides, we have proposed a semantic architecture relying on HF to ensure knowledge integrity and authentication while preserving privacy. Finally, the deployment of Fabric 2.3 as the framework's underlying blockchain did not negatively affect the response time. We have performed experiments to validate the time required for an action to take place and have evaluated the system's response time after observing the context, obtaining access to the camera request and executing the access command. The response time includes the processing time in the

communication layer, as well as the time for generating the action and sending the command to an actuator device. This paper demonstrates how this approach ensures message integrity, verification, authentication, security and privacy, and how it allows semantic contextual knowledge to be shared in order to invoke one or more services. The use of a consortium blockchain in which not every peer has equal rights to endorse a proposed transaction is a potential disadvantage of the proposed architecture. Within the HF, only a few peers can validate transactions. Due to the fact that it is an emerging technology, we should explore the usefulness of the public decentralized blockchain principle.

References

- [1] A. Brunete, E. Gambao, M. Hernando, and R. Cedazo, "Smart Assistive Architecture for the Integration of IoT Devices, Robotic Systems, and Multimodal Interfaces in Healthcare Environments", *J. Sensors*, vol. 21, no. 6, 2021 (DOI: 10.3390/s21062212).
- [2] J.M. Byeong, S.K. Sonya, and C. JongSuk, "Organizing the Internet of Robotic Things: The Effect of Organization Structure on Users' Evaluation and Compliance toward IoRT Service Platform", *IROS*, pp. 628–629, 2020 (DOI: 10.1109/IROS45743.2020.9340834).
- [3] A. Kumari and S. Tanwar, "Secure data analytics for smart grid systems in a sustainable smart city: Challenges, solutions, and future directions", *J. Sustainable Computing: Informatics and Systems*, vol. 28, 2020 (DOI: 10.1016/j.suscom.2020.100427).
- [4] B. Bhushan, P. Sinha, K.M. Sagayam, and J.A. Onesimu, "Untangling blockchain technology: A survey on state of the art, security threats, privacy services, applications and future research directions", *J. Computers Electrical Engineering*, vol. 90, 2021 (DOI: 10.1016/j.compeleceng.2020.106897).
- [5] S. Pal, A. Dorri, and E. Jurdak, "Blockchain for IoT Access Control: Recent Trends and Future Research Directions", *CoRR*, 2021 (DOI: 10.1016/j.jnca.2022.103371).
- [6] H.C. Chen, "Collaboration IoT-Based RBAC with Trust Evaluation Algorithm Model for Massive IoT Integrated Application", *J. Mob. Networks Appl.*, vol. 24, pp. 839–852, 2021 (DOI: 10.1007/s11036-018-1085-0).
- [7] S. Christos, E.P. Kostas, K. Byung-Gyu, and G. Brij, "Secure integration of IoT and Cloud Computing", *J. Future Generation Computer System*, vol. 78, pp. 964–975, 2018 (DOI: 10.1016/j.future.2016.11.031).
- [8] –, (<https://www.hyperledger.org/use/fabric>)
- [9] C. İozgü and D. Yilmazer, "Improving privacy in health care with an ontology-based provenance management system", *J. Expert Systems, Special Issue: eHealth and Staying Smarter*, vol. 37, 2020 (DOI: 10.1111/exsy.12427).
- [10] P. Gonzalez-Gil, J.A. Martínez, and A.F. Skarmeta, "Lightweight Data-Security Ontology for IoT", *J. Sensors*, vol. 20, 2020 (DOI: 10.3390/s20030801).
- [11] L. Sabri and A. Boubetra, "Narrative Knowledge Representation and Blockchain: A Symbiotic Relationship", *Advanced Information Networking and Applications Proceedings of the 34th International Conference on Advanced Information Networking and Applications (AINA-2020)*, pp. 320–332, 2020 (DOI: 10.1007/978-3-030-44041-1_30).
- [12] B. Sejdiu, I. Florije, and L. Ahmedi, "A Management Model of Real-time Integrated Semantic Annotations to the Sensor Stream Data for the IoT", *WEBIST*, pp. 59–66, 2020 (DOI: 10.5220/00101115005900066).
- [13] G.P. Zarri, "A knowledge representation tool for encoding the 'meaning' of complex narrative texts", *Natural Language Engineering*, vol. 3, pp. 231–253, 1997 (DOI: 10.1017/S1351324997001794).
- [14] A. Reyna, M. Cristian, J. Chen, E. Soler, and M. Díaz, "On blockchain

and its integration with IoT. Challenges and opportunities”, *J. Future Generation Computer Systems*, vol. 388, pp. 173–190, 2018 (DOI: 10.1016/j.future.2018.05.046).

- [15] F. Tschorsch and B. Scheuermann, “Bitcoin and beyond: a technical survey on decentralized digital currencies”, *J. IEEE Communications Surveys & Tutorials*, vol. 18, pp. 2084–2123, 2016 (DOI: 10.1109/COMST.2016.2535718).
- [16] D.D. Fiergbor, “Blockchain Technology in Fund Management”, *Communications in Computer and Information Science; Springer*, vol. 899, pp. 310–319, 2018 (DOI: 10.1007/978-981-13-2035-4_27).
- [17] –, (<https://www.ibm.com/topics/hyperledger>)
- [18] –, (<https://protege.stanford.edu/>).
- [19] F. Baader, D.L. McGuinness, D. Nardi, and P.F. Patel–Schneider, “The Description Logic Handbook: Theory, implementation, and applications”, *Cambridge University Press*, 2010 (DOI: 10.1017/CBO9780511711787).
- [20] J. Kwapien and S. Drozd, “Physical approach to complex systems”, *Physics Reports*, vol. 515, no. 34, pp. 115–226, 2012 (DOI: 10.1016/j.physrep.2012.01.007).
- [21] Y. Li, L. Qiao, and Z. Lv, “An Optimized Byzantine Fault Tolerance Algorithm for Consortium Blockchain”, *Peer-to-Peer Netw. Appl.*, vol. 18, pp. 2826–2839, 2021 (doi: 10.1007/s12083-021-01103-8).
- [22] D. Hooda and R. Rani, “Ontology driven human activity recognition in heterogeneous sensor measurements”, *J. Ambient Intelligence and Humanized Computing*, vol. 11, pp. 5947–5960, 2020 (DOI: 10.1007/s12652-020-01835-0).



Manal Lamri is a Ph.D. Student at the Faculty of Mathematics and Informatics, University of Mohamed El Bachir El Ibrahimi, Bordj Bou Arreridj, 34030, Algeria. Her research interests are in ontology, knowledge representation, reasoning, the Internet of Things paradigm, and blockchain technology.

E-mail: manal.lamri@univ-bba.dz

Faculty of Mathematics and Informatics, University of Mohamed El Bachir EL Ibrahimi, Algeria



Lyazid Sabri is an Assistant Professor at the Bachir Ibrahimi University. He holds a Ph.D. in Computer Science, Images, Signals and Intelligent Systems from University Paris-Est Créteil (UPEC). He is a Senior Consultant on Artificial Intelligence, Information Systems Architectures and Cybersecurity. While at UPEC, he worked on several collaborative research projects (e.g.

SembySem, Web of Objects, Predykot, A2nets) dealing with IoT and ambient assisted living, robotics, artificial intelligence, and access management. His research interests are in ontology, knowledge representation and reasoning, robotics, artificial intelligence, deep learning, cognitive psychology, the Internet of Things paradigm, and blockchain technology. E-mail: sabrilyazid@gmail.com, lyazid.sabri@univ-bba.dz, sabri@lissi.fr

Faculty of Mathematics and Informatics, University of Mohamed El Bachir EL Ibrahimi, Algeria

Laboratory of Images, Signals and Intelligent Systems, University of Paris-Est, France

Multi-operator Differential Evolution with MOEA/D for Solving Multi-objective Optimization Problems

Sakshi Aggarwal and Krishn K. Mishra

Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

<https://doi.org/10.26636/jtit.2022.161822>

Abstract — In this paper, we propose a multi-operator differential evolution variant that incorporates three diverse mutation strategies in MOEA/D. Instead of exploiting the local region, the proposed approach continues to search for optimal solutions in the entire objective space. It explicitly maintains diversity of the population by relying on the benefit of clustering. To promote convergence, the solutions close to the ideal position, in the objective space are given preference in the evolutionary process. The core idea is to ensure diversity of the population by applying multiple mutation schemes and a faster convergence rate, giving preference to solutions based on their proximity to the ideal position in the MOEA/D paradigm. The performance of the proposed algorithm is evaluated by two popular test suites. The experimental results demonstrate that the proposed approach outperforms other MOEA/D algorithms.

Keywords — differential evolution, multi-objective, mutation-operators, weighted-aggregation

1. Introduction

Multi-objective evolutionary algorithms (MOEAs) are applied for decoding various multi-objective optimization problems (MOP) [1]–[3]. To develop an effective and efficient MOEA, one cannot overlook some serious concerns such as the selection of solution for the offspring in order to evolve the population. Another concern is related to how diversity of the population may be maintained while choosing the solutions for the successive generations. And finally, it is very hard to balance the diversification-intensification relationship in MOP, since the objectives might be conflicting in nature. Depending upon the selection criteria for new solutions, MOEAs are broadly classified into three categories: Pareto-dominance-based MOEAs [4]–[7], performance indicator-based approaches [8]–[11], and decomposition-based algorithms [12]–[15]. However, a general approach is to transform the MOP into multiple single-objective problems, i.e. to transform a decision-space into an objective space for developing MOP frameworks.

In recent years, the decomposition-based MOEA technique (MOEA/D) has gained attention for solving MOP [16]. The popular examples are MOEA/D-DE [17] and MOEA/D-CMA [18], utilizing the single-search mutation operator of differential evolution (DE) to converge the entire pop-

ulation towards the Pareto front. Likewise, MOEA/D with a distance update strategy (MOEA/D-DU) [19] motivates researchers to measure the distance between the value of weighted-aggregation function and its corresponding vector in MOEA/D. Despite their valuable results, the aforementioned frameworks suffer from the following disadvantages:

- the solutions are selected either randomly or from the local region. In MOEA/D-DE, the parent vector is either chosen from the neighbor or randomly from the entire population. This type of selection is likely to mislead the search process and confine it to a certain area of the Pareto front;
- similarly, in MOEA/D-CMA, few solutions are mutated through CMA-ES [20]–[22] and most of them are expected to converge through DE. This study may enhance diversity of the population, but lacks in faster convergence towards the Pareto front;
- in the existing studies, offspring is generated by means of conventional approaches (either by DE or GA [19]). These are not capable of producing reliable results for all the sub-problems and, hence, may be stuck in the local minima.

To cope with this, we propose a multi-operator based differential evolution with MOEA/D (MOEA/D-MODE) that alleviates, to certain extent, the shortcomings in the area of diversity-preservation and convergence rate.

This paper relies on the clustering-based MOEA/D that has the advantages of multiple-mutation strategies of the established evolutionary approach (DE), to ensure the equilibrium between exploration and exploitation. Furthermore, the difficulty to pull the solutions to the Pareto front is taken care of to some extent as well. Our contribution is described below:

- a novel multi-operator DE variant (MOEA/D-MODE) is proposed for ensuring a better trade-off between diversity and convergence in the MOEA/D multi-objective optimizer. To implement this idea, three diverse mutation strategies of DE are employed;
- clustering-based evolution is emphasized which can explicitly facilitate better diversification. The clusters are of varying sizes and each cluster is operated with a distinct mutation operator;
- contemporary ideas are combined in order to select the solution vector for the generating mutant solution;

- to ensure maximum diversity, we have incorporated polynomial mutation followed by standard crossover techniques to yield novel solutions in the sub-population. Then, we compared the proposed algorithm with three existing solutions: MOEA/D-CMA, MOEA/D-DE, and MOEA/D-DU, and also discussed potential reasons behind the failure of these methods proposed in MOP.

The remaining sections of the paper are organized as follows. Section 2 illustrates the fundamentals of MOP and is followed by a presentation of the related work in Section 3. In Section 4, the crucial components of MOEA/D-MODE are discussed. In Section 5 comprehensive implementation of the proposed algorithm with the aim to solve MOP is presented. Section 6 describes the experimental studies in terms of benchmark functions, parameter settings, and evaluation metrics for comparison purposes. In Section 7, performance of the improved MOEA/D-MODE algorithm is analyzed (with respect to two aspects) and statistical results comparing the solution with, other algorithms are verified. Finally, the paper is concluded in Section 8.

2. Background

Any MOP can be defined as:

$$\text{minimize: } F(x) = f_1(x), \dots, f_m(x) \quad \text{subject to: } x \in \Omega, \quad (1)$$

where Ω represents the decision (solution) space of a d -dimensional vector $x = (x_1, x_2, \dots, x_d)$, and $F : \Omega \rightarrow \mathbb{R}^m$ contains m continuous objective values in the \mathbb{R}^m objective space. Moreover, if Ω is a connected and closed region in the objective space \mathbb{R}^m and the corresponding objective solutions are continuous of x , Eq. (1) is referred to as a continuous MOP.

Let $u = (u_1, \dots, u_m)$ and $v = (v_1, \dots, v_m) \in \mathbb{R}^m$ be two solutions, u dominates v if and only if:

$$\begin{aligned} u_i &\leq v_i, \quad \text{for all } i = 1, \dots, m, \\ u_i &< v_i, \quad \text{for any } i \in 1, \dots, m. \end{aligned}$$

Solution $x^* \in \Omega$ is said to be a Pareto optimal solution if there does not exist $x \in \Omega$ such that $f(x)$ dominates $f(x^*)$. All the reliable Pareto solutions together form a set, known as a Pareto set (PS):

$$PS = \{x \in \Omega | x \text{ is Pareto optimal}\}. \quad (2)$$

The set of all the Pareto objective vectors, known as the Pareto front (PF), is given as:

$$PF = \{f(x) \in \mathbb{R}^m, x \in PS\}. \quad (3)$$

For a given MOP, the ideal solution z^* is the best solution vector $z^* = (z_1^*, z_2^*, \dots, z_m^*)$, where z_i^* represents the best solution (here the minimum value) of f_i , for every $i = 1, 2, \dots, m$.

The prime objective of any MOP technique is to guide the population of worthwhile solutions toward the PF, ensuring convergence and, simultaneously, maximum distribution over the PF for diversity related purposes.

3. Related Works

Three categories of MOEAs may be used in MOP. So, this section is devoted to discussing the literature based on the aforementioned categories. In the majority of literature focusing on MOEAs assistance of the Pareto dominance is relied upon [4]–[7], [23], [24]. In these studies, the effectiveness of a solution is measured by the Pareto dominance relations with the remaining solutions encountered in the last search. It is an iterative process that runs for each individual element in the objective-space. Since the dominance feature alone could hamper the diversity of the solutions, some alternatives may be combined in MOEAs, such as crowding and fitness sharing [24], [25]. One of the most popular Pareto-dominance MOEA schemes is NSGA-II [6]. The crucial characteristic of NSGA-II is its rapid nondominated sorting to rank the solutions for further selection.

Indicator-driven MOEAs are another category, as they endeavor to optimize performance metrics as an indicator [8], [9]. They ensure the desired ordering sequence of the optimal sets that will be used to approximate the Pareto front. The most widely adopted performance indicator is hypervolume (HV), which possess significant theoretical characteristics. In the literature, we have few canonical performance indicator-based MOEAs [8], [9] that disguise HV as the selection factor. One of the suggestions is to rank the solutions yielded by the HV indicator rather than estimating their exact values [9]. Another alternative strategy is to find other indicators that are computationally less expensive and offer fair theoretical characteristics, e.g. Λ_p [28]. Such an approach has been embraced in a few MOEAs.

The category of decomposition-based MOEAs exploits the aggregation function in which the objectives of a MOP are aggregated using randomly distributed weight-vectors. This set of weight-vectors will eventually create multiple weighted-aggregation functions, each of them representing a single-objective problem. Diversity of the population is maintained by ensuring fair distribution of the weight-vectors in the objective space. MOEA based on the decomposition (MOEA/D) [16] is a scheme that is most widely adopted in the domain of multi-objective optimization. New frameworks based on MOEA/D and relevant to the study performed in this paper are reviewed in the following subsections.

3.1. MOEA/D with DE

The general practice in MOEA/D is to decompose PF approximations of a problem (1) into several scalar-optimization functions. Li and Zhang in [17] extended the work by implementing DE and polynomial mutation for maintaining the diversity of the population in MOEA/D. In such an approach, three parent solutions are selected having a low probability of $1 - \delta$. In such a way, a wide range of offspring could be produced and, thus, the exploration capability was enhanced. Furthermore, there is a restriction on replacing the maximum number of solutions with a new child solution. Instead of relying upon the neighborhood of size T , parameter

n_r is introduced. It limits the size of the solution-vector to be replaced by the new offspring.

The differences between MOEA/D-MODE and MOEA/D-DE can be summarized in the following manner:

- in MOEA/D-DE, a single mutation strategy is incorporated that utilizes three parent solutions only. The standard DE technique is used to produce new offspring. Multi-operator DE often outperforms single mutation DE in the case of single-objective problems. However, to enhance the search capabilities of MOP, multiple mutation strategies are ensembled in MOEA/D-MODE;
- the extra measure taken in MOEA/D-MODE is the implementation of the crossover technique after the polynomial mutation. The crossover technique is useful for exploiting regions formed by mutant vectors. This allows to strengthen the trade-off between exploration and exploitation.

3.2. MOEA/D with CMA-ES

Working on MOEA/D frameworks, Li *et al.* [18] introduced the covariance matrix adaptation evolution strategy (CMA-ES) into MOP in order to balance CMA-ES and DE efficiently. CMA-ES is an evolutionary approach which allows to generate novel solutions using the Gaussian distribution model. To lower the cost of computation, the problem domain is organized into a group of sub-problems where only one sub-problem is optimized through CMA-ES and others are evolved by applying the DE approach. The best solutions optimized by CMA-ES are always carried forward in the distribution mean update. This leads to faster convergence.

The differences between MOEA/D-MODE and MOEA/D-CMA are such that MOEA/D-CMA involves clustering of sub-problems, with only a few of them being optimized by the Gaussian distribution model of CMA-ES. It seems the algorithm is more focused on DE, as the majority of sub-problems are evolved by means of the DE mutation strategy. Unlike MOEA/D-CMA, MOEA/D-MODE allows different mutation strategies to be applied in the clusters of the sub-problems, thus maintaining diversity and working in a single flow.

3.3. MOEA/D with Distance Update Strategy

Another MOEA variant based on the decomposition technique, as proposed by Yuan *et al.* in [19], uses the aggregation function to speed up the convergence in multiple-objective optimization. As the number of objectives increases exponentially, it becomes difficult to maintain diversity and to approach the PF uniformly. To cope with this challenge, researchers have performed extensive analyses on the aggregation functions by estimating the perpendicular distance from the weight-vector of the solution in the high-dimension objective space. The performance of such an approach in the case of a 2-objective optimization problem, (and with more than 2 objectives) has been analyzed as well. The differences between our approach and MOEA/D with the distance update strategy include the following:

- in MOEA/D-MODE, the worst neighbor is used as the solution according to its distance from the weight-vector and the best solution according to the better aggregation function value corresponding to the sub-population;
- DE generally offers better results compared with the genetic operators in the case of single-optimization problem. The Cr parameter sets the number of new solutions to be exploited. With the low value of Cr , a wide range of child solutions will be covered, while a high value of Cr is focused on the parent vector only. Due to the above-mentioned reasons, DE search operators are incorporated in MOEA/D-MODE to solve MOP.

4. Pivotal Components of MOEA/D-MODE

The single-mutation strategy is incorporated into decomposition-based multi-objective optimization for population evolution-related purposes. In the proposed algorithm, we adopt a novel approach involving multiple-mutation operators. Each of them is applied uniquely to evolve the sub-populations, leading to stronger exploration and better convergence. As illustrated in Fig. 1, the dashed lines represent the contour of the sub-problems decomposed by Eq. (5). The clusters are organized based on the weight vectors $\lambda_1, \lambda_2, \dots, \lambda_8$. In each generation, one solution at a time is taken from the cluster and the assigned mutation strategy is applied.

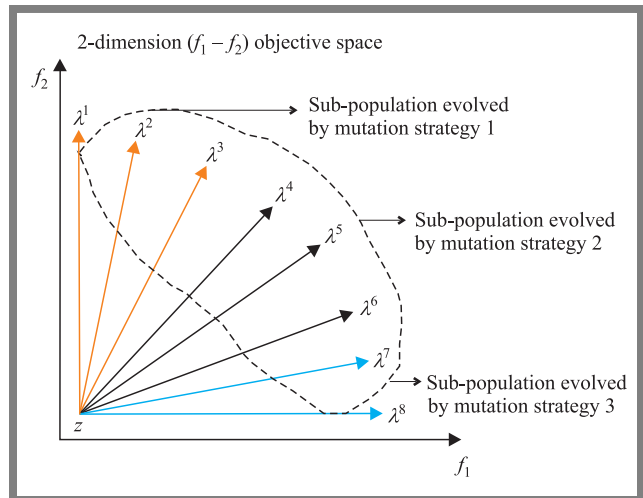


Fig. 1. Illustration on the clusters of the sub-problems decomposed by Eq. (5). Each cluster is assigned a unique mutation operator for generating novel solutions. Here, there are three clusters: $1 = \{1, 2, 3\}$, $2 = \{4, 5, 6\}$ and $3 = \{7, 8\}$.

4.1. Neighbor Selection and Clustering

The common practice in decomposition-based MOEAs is to transform the MOP into many single-objective problems, with each objective being a weighted combination of different objectives. This is achieved by initializing the weight-vectors in the objective-space.

Let $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,m})^T$, for $i = 1, 2, \dots, N$, be uniformly distributed weight-vectors for N solutions, such that $\sum_{j=1}^m \lambda_{i,j} = 1$. Under such an assumption, the neighbors of each unique solution are identified according to their similarity. This is achieved by computing the $N \times N$ Euclidean distance metric:

$$\text{dist}(u, v) = \sqrt{\sum_{i=1}^m (\lambda_{u_i} - \lambda_{v_i})^2}, \quad (4)$$

where $\text{dist}(u, v)$ represents the Euclidean distance between two solutions u and v . The closer the distance, the higher the neighborhood relationship. Therefore, in MOEA/D-MODE, we construct a best-neighbors vector B of size T for further processing such that $B = \{x_1, x_2, \dots, x_T\}$.

To achieve maximum diversity even in the later stages of the population, the objective is to initially disintegrate the entire population and cluster the solutions based on the assigned weight vectors. All sub-populations have different sizes. In conjunction, multiple-mutation techniques of DE have been applied that ensure better coverage of the search space. This practice is likely to explicitly maintain the diversity of solution during evolution of the population.

To accomplish the task k -means clustering [29] is applied with $k = 3$, since three diverse mutant operators are considered in this algorithm to process three sub-populations.

4.2. Parent Selection and Offspring Generation

Another major concern is the selection of parent solutions for offspring generation. It is important that the selection criteria be driven not only by the distant vector λ_i but also by proximity to the ideal position in the objective space, i.e. using the aggregation-function value $G_i(x)$ given in Eq. (5). Such an approach is driven by the likelihood that, a solution which is inferior in terms of the λ_j may contribute to a better $G_i(x)$ value.

Therefore, in this paper, we consider the weighted-aggregation function value $G_i(x)$ which underlines the best solution in the sub-population, while selecting the parent solutions for the respective mutation operators.

In the proposed algorithm, three diverse mutation operators are applied to turn on the novel solutions. Additionally, polynomial mutation and crossover techniques have been incorporated that are rarely applied in existing MOEA/D variants. The crossover techniques employ either a binomial crossover or an exponential crossover for the new solution u .

4.3. Updating Solutions in the Sub-population

The most common scheme for using aggregation functions in updating neighbors of the solution is the Tchebycheff function [30]. In this function, the scalar optimization sub-problem is given by:

$$G_i(x) = \max_{1 \leq i \leq m} \{\lambda_i |f_i(x) - z_i^*|\} \quad \text{subject to } x \in \Omega, \quad (5)$$

where m denotes the number of objectives, λ is a uniformly distributed weight vector across each objective, and z_i^* represents best the solution found so far for each objective i .

The problem of converging the entire solution-set towards PF is remodeled into N scalar sub-problems requiring optimization. Eventually, the spread of the final solutions could be evenly distributed if $G(x)$ and λ are appropriately determined. Once the new offspring u is achieved, the solutions in the sub-population get updated if:

$$G(x) > G(u), \quad (6)$$

where x denotes the solution-vector in the cluster (i.e. sub-population). Otherwise, the same parent solution will be carried forward to the next generation. Table 1 summarizes the concepts exploited in the proposed MOEA/D-MODE algorithm.

Tab. 1. MOEA/D-MODE concept.

No.	Stages	Technique used in MOEA/D-MODE
1	Selection of neighbors for each solution	Led by the distant vector λ
2	Solution clustering	3-means clustering based on factor λ
3	Parent solutions are selected for offspring	According to the best $G_i(x)$ Eq. (5) in the sub-population
4	Offspring generation	Three diverse mutation strategies have been incorporated
5	Maximal diversity	Enhanced by polynomial mutation and crossover techniques
6	Update solutions in the sub-population	Using Eq. (6)

5. MOEA/D-MODE Algorithm

This section focuses on the mathematical model of multi-operator DE for solving MOP. Additionally, we describe our approach consisting in exploiting three diverse mutation operators along with a polynomial mutation and standard crossover techniques.

Initially, the population is initialized randomly with N number of candidate solutions as:

$$x_{i,j} = x_{i,j}^{\text{lower}} + (x_{i,j}^{\text{upper}} - x_{i,j}^{\text{lower}}) \times \text{rand} \quad i \in N \text{ and } j = 1, 2, \dots, D, \quad (7)$$

where rand is a function that generates random numbers between $[0 \dots 1]$ [31]. The terms lower and upper represent lower and upper boundaries of variable x in the D -dimension.

After generating the sub-populations, solutions in the sub-populations are evolved via multi-mutation operators, where each sub-population is assigned a unique mutation operator. This allows to maintain diversity of the internal population.

Algorithm 1. MOEA/D-MODE**Parameter initialization:**

$MAX_{FES}, N, K, T, FES \leftarrow 0$

Controlling parameters initialization:

F, p_m, η

Weight vectors Λ :

initialize a set of weight vectors $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_N\}$

Population initialization:

random population X of size N as $\{x_1, x_2, \dots, x_N\}$

instantiate a ideal point $z^* = z_1^*, z_2^*, \dots, z_m^*$

 T -neighbors initialization:

for $i \leftarrow 1$ to N **do**

$B(i) \leftarrow \{i_1, i_2, \dots, i_T\}$

end

Clustering:

$C \leftarrow k\text{-means}(\Lambda, K)$

while $FES \leq MAX_{FES}$ **do**

for $s \leftarrow 1$ to N **do**

$P \leftarrow B(s)$

if $s \in$ any C **then**

 Generate mutant vector from three defined mutation operators as:

$$\bar{y}_s = \begin{cases} \text{Eq. (8)} & \text{if } s \in C(1) \\ \text{Eq. (9)} & \text{if } s \in C(2) \\ \text{Eq. (10)} & \text{if } s \in C(3) \end{cases}$$

$y_s \leftarrow \text{PolynomialMutation}(x^s, \bar{y}_s)$

$u_s \leftarrow \text{Crossover}(x^s, y_s)$

$z^* = \min(z^*, z(u_s)^*)$

 UpdateSubPopulation(u_s, z^*, C)

end

end

$FES \leftarrow FES + N$

end

For each unique solution x^s chosen from the respective sub-population, a mutant vector \bar{y}_s is generated as follows:

– sub-population 1: DE/parent-to-worst/1

$$\bar{y}_s = x^s + F \times (x_{p1} - x^s + x_{r1} - x_{\text{worst}}), \quad (8)$$

– sub-population 2: DE/parent-to-worst/1

$$\bar{y}_s = x^s + F \times (x_{p2} - x^s + x_{r2} - x_{\text{worst}}), \quad (9)$$

– sub-population 3: DE/weighted-rand-to-worst/1

$$\bar{y}_s = x^s + F \times (x_{r3} + x_{p3} - x_{\text{worst}}). \quad (10)$$

x^s denotes the target vector, x_{p1} , x_{p2} , and x_{p3} are 40%, 16% and 25% of the best solutions chosen from sub-populations 1, 2, and 3, respectively. Additionally, the topmost solutions are extracted from the respective clusters and marked as x_{r1} , x_{r2} , and x_{r3} , respectively. In the propounded multi-operator DE for MOP, the objective is to filter the solutions that cannot be converged to PF and, hence, maintain the maximum distance from the Pareto optimal solutions. This is implemented as $x_{\text{worst}} \in B$ which is the worst neighbor of x^s since their distance $\lambda_s - \lambda_{x_{\text{worst}}}$ differs significantly.

The three mutation strategies presented above have their own advantages, such as:

- sub-population-based evolution is used where each of them holds a variable number of solutions. This practice is likely to explicitly maintain population diversity throughout the evolution;
- the solutions that achieve significantly close proximity to the ideal position are exploited to improve the selection procedures not only from the neighborhood, but that paves the way for the maximum space coverage;
- each mutant operator tries to maintain the maximum distance from the solution that seems less promising at the time. Hence, it brings all the solutions close to the PF.

Polynomial mutation is adopted widely in evolutionary approaches in order to allow variation in the solutions. The above mutation strategies are followed by polynomial mutation in which y is generated from \bar{y} in the following manner:

$$y_k = \begin{cases} \bar{y}_k + \sigma_k \times (\text{upper}_k - \text{lower}_k) & p_m \\ \bar{y}_k & 1 - p_m \end{cases}, \quad (11)$$

where

$$\sigma_k = \begin{cases} (2 \times \text{rand})^{\frac{1}{\eta+1}} - 1 & \text{if } \text{rand} < 0.5 \\ 1 - (2 - 2 \times \text{rand})^{\frac{1}{\eta+1}} & \text{otherwise} \end{cases}. \quad (12)$$

The rand function produces a random number between $[0 \dots 1]$. There are two controlling parameters: p_m which defines the expectation of the number of mutated variables and η representing the distribution index of the polynomial mutation. The terms upper_k and lower_k are the upper and lower boundaries of the k -th decision variable of solution s , respectively.

In order to find Pareto optimal solutions, a crossover technique is employed. In this approach, maximum exploitation could be maintained along with the evolution of new solutions u yielded from y . Either binomial or exponential crossover is applied according to:

$$u_k = \begin{cases} \text{if } \text{rand} < 0.4 \\ x_k^s \\ \text{otherwise} \\ \begin{cases} y_k & \text{for } k = \langle l \rangle_D, \langle l+1 \rangle_D, \dots, \langle l+L-1 \rangle_D, \\ x_k^s & \text{for rest of } k \in [1, D] \end{cases} \end{cases} \quad (13)$$

where $\langle \rangle$ is a modulo operator in the exponential crossover.

After evolving the solutions in a sub-population, the ideal position is changed. Therefore, we get a new $z^* = \min[z(x)^*, z(u)^*]$. Subsequently, $G(u)$ is computed as:

$$G_i(u) = \max_{1 \leq i \leq m} \{\lambda_i | f_i(u) - z_i^* \}. \quad (14)$$

Once the weighted function $G(u)$ has been obtained, next generation solutions are decided. To ensure the better solutions, the solutions in the sub-population are updated as:

$$x_k = \begin{cases} u_k & \text{if } G(x) > G(u) \\ x_k & \text{otherwise} \end{cases}. \quad (15)$$

where x is the target solution in the sub-population, u depicts a new solution corresponding to x , for each component $k \in 1, \dots, D$. Similarly, the entire mechanism is implemented for the solutions in the remaining sub-populations. Algorithm 2 shows the sub-population updating criteria in the propounded variant of DE for multi-objective optimization.

Algorithm 2. UpdateSubPopulation(u_s, z^*, C)

Compute $G(x_C)$ according to Eq. (5)

Compute $G(u_s)$ according to Eq. (14)

if $G(x_C) > G(u_s)$ **then**

 | Update the solutions x of cluster C where $s \in C$

end

6. Experimental Setup

The implementation of MOEA/D-MODE is executed with Matlab R using the PlatEMO framework [32]. Its performance is evaluated with the use of two test suites, with respect to three well-known decomposition-based MOEAs frameworks for solving MOP: MOEA/D-CMA, MOEA/D-DE, and MOEA/D-DU.

First, MOP benchmark functions are tethered with the bias difficulties as well as BT1-BT9 instances [18] included. For BT1 to BT8, there are two objectives, whereas BT9 alone is a many-objective problem defined with the use of three objectives.

In the second step, the behavior of MOEA/D-MODE on the ZDT series [33] is evaluated. Such a method is conceived purely for two-objective test problems. However, ZDT5 is excluded from the experimental study, since it involves binary computations. Both of the test suites having diverse function problems of dimension $D \in \{10, 30\}$ and objectives $M \in \{2, 3\}$.

The control parameters and other relevant data of proposed algorithm MOEA/D-MODE are provided in Table 2. The other common parameters are:

- **number of runs and MAX_{FES}**. MOEA/D-MODE and the remaining competing algorithms participating in the comparison are run 30 times, independently in each of the test suites. The termination criterion for all the algorithms is set to 10,000 for all test problems;
- **weight-vector Λ** . Weight-vector $\Lambda = \{\lambda_1, \dots, \lambda_N\}$ is a set of uniformly distributed random values and has the size of $N \times M$, where N shows the population size and M denotes the total number of objectives;
- **population size N** . To promote a fair comparison, the MOEA/D-MODE framework, and other algorithms assume the population size to be 100 for each test problem;
- **neighborhood size T** . In the proposed MOEA/D-MODE framework, and in other algorithms (MOEA/D-CMA, MOEA/D-DE, and MOEA/D-DU), T is initialized to 10;
- **mutation parameters (p_m and η)**. All respective algorithms rely on polynomial mutations for introducing new

Tab. 2. Parameters settings of MOEA/D-MODE.

Parameter	Symbol	Value
Maximum function evaluations	MAX _{FES}	10,000
Population size	N	100
Neighbors size	T	10
Number of clusters	C	3
Scaling factor	F	0.5
Crossover probability	Cr	1
Expectation of the mutated variables	p_m	1
Distribution index	η	20

solutions. Mutation probability p_m is set to 1 with a large distribution index (with its value equaling 20) is used for mutation η .

Some algorithms are characterized by particular parameter settings. In MOEA/D-DE and MOEA/D-DU, δ is the probability of selecting parents from local regions, and is set to 0.9. n_r is used by MOEA/D-DE to determine the maximum number of solutions replaced by the offspring. The value chosen is 2. On the other hand, parameter K holds different meanings in MOEA/D-DU and MOEA/D-CMA respectively. In MOEA/D-DU, K denotes the number of the nearest weight vectors, whereas in MOEA/D-CMA, K represents the number of groups. Both algorithms assume that this value equals 5.

6.1. Evaluation Metrics

Inverted generational distance (IGD) [34] is used as a performance evaluation metric. IGD is a metric that is widely adopted in the multi-objective domain and it allows to obtain collective information on the convergence and distribution of solutions. In the objective-space, we need a significant number of uniformly distributed variables that converge to PF in order to efficiently estimate IGD.

Along with IGD, we incorporate another well-known metric, namely hyper-volume (HV) [11], as the predominant comparison factor. HV is crucially cooperative to PF, and its encouraging theoretical characteristics turn it into a fair metric [35]. It can represent both convergence and distribution of the solutions. The larger the HV value, the better the level of quality.

Selection of the reference point is the main concern encountered while computing HV. In this paper, following the recommendation from [36] and [37], we assumed the reference point to be $1.1z^{\text{nad}}$, where z^{nad} is analytically computed against each function instance. Besides, according to the setup used in [38] and [39], the solutions that do not converge to the reference point are ignored for HV computation.

To understand the difference for statistical significance of function instances, we performed the Wilcoxon Rank-Sum test [40] with normal approximation, tie-breaking, and with the significance level set to 1%. It was performed on the HV metric scores yielded by algorithms other than the proposed solution.

Tab. 3. Comparison of algorithms based on HV results, for an average and standard deviation (in brackets). The best results are highlighted in bold print.

Function	M	D	MOEA/D-MODE	MOEA/D-CMA	MOEA/D-DE	MOEA/D-DU
BT1	2	30	0 (0)	0 (0)	0 (0)	0 (0)
BT2	2	30	0 (0)	0 (0)	0 (0)	0 (0)
BT3	2	30	0 (0)	0 (0)	0 (0)	0 (0)
BT4	2	30	0 (0)	0 (0)	0 (0)	0 (0)
BT5	2	30	0 (0)	0 (0)	0 (0)	0 (0)
BT6	2	30	0.121 (0.040)	0 (0)	0 (0)	0 (0)
BT7	2	30	0.092 (7.25 × 10⁻³)	0 (0)	0.013 (0.027)	6.659 10 ⁻³ (0.015)
BT8	2	30	0.096 (0.021)	0 (0)	0 (0)	0 (0)
BT9	3	30	0 (0)	0 (0)	0 (0)	0 (0)
ZDT1	2	30	0.532 (0.062)	0.516 (0.03)	0.26 (0.075)	0.264 (0.103)
ZDT2	2	30	0.238 (0.075)	0.224 (0.034)	0.017 (0.06)	0 (0)
ZDT3	2	30	0.605 (0.127)	0.423 (0.057)	0.288 (0.091)	0.462 (0.052)
ZDT4	2	10	0.38 (0.106)	0 (0)	0 (0)	0 (0)
ZDT6	2	10	0.271 (0.033)	0.371 (0.048)	3.595 (0.053)	0 (0)

7. Result Analysis

First the convergence and distribution of MOEA/D-MODE solutions obtained with the use of the two test suites, i.e. BT and ZDT, are analyzed (Table 3). The set of non-dominated solutions found by the proposed algorithm in 30 independent runs is depicted in Fig. 2 and 3. Based on these illustrations, the following observations may be made.

In the BT test suite, BT6–BT8, Fig. 2f–h, shows the convergence of the solutions to the PF across the objective space. Only a few of the candidate solutions try to reach the PF. This indicates that the embedded mutation strategy requires a greater ability to deal with the variations in MOP.

From BT1–BT8 (Fig. 2a–g), one may conclude that the solution set is distributed in the objective space, but does not converge to the optimal PF. This may be due to the early termination of the algorithm. Further iterations are needed for the evolution, so that it may converge very well, since the optimization problem involves tough biases.

The result of the only function problem based on 3-objectives is depicted in Fig. 2i. It illustrates the distribution of the solutions along the PF but the results shown are not encouraging. It seems that the normal population size, taken for MOP, e.g. 100, is not suitable for a problem that involves more than 2-objectives.

As far as the analysis of the ZDT series (Fig. 3) is concerned, the proposed algorithm shows far better results. It is clearly seen that the solution-set becomes converged to the PF (Fig. 3a–e). MOEA/D-MODE shows a better convergence rate in ZDT3 (Fig. 3c). However, there is still some room for improvement in the convergence rate in order to optimize different classes of problems.

7.1. Statistical Analysis

Table 4 shows a statistical comparison between MOEA/D-MODE and of other algorithms. Table 5, in turn, contains the IGD results. W^+ stands for the number of test instances in the case of which MOEA/D-MODE is significantly superior. $W^=$ means there are no significant differences between the obtained scores, and W^- is the number of instances for which existing solutions perform significantly better than MOEA/D-MODE.

The comments concerning MOEA/D-MODE and covering all 14 test instances are as follows:

In the BT test suite, MOEA/D-MODE shows a certain advantage over MOEA/D variants, i.e. MOEA/DE, MOEA/D-CMA and MOEA/D-DU. In the majority of test problems, MOEA/D-MODE achieves results that are comparable with those of the three remaining algorithms. However, it also shows an improvement in three function instances that are overlooked by the other alternatives.

When comparing results for the ZDT test series, one may clearly observe that MOEA/D-MODE remains competitive in the majority of test instances. It has shown that the multi-

Tab. 4. Summary of statistical results on HV metrics between MOEA/D-MODE and the rival algorithms.

Test suite	Algorithm	W^+	$W^=$	W^-
BT	MOEA/D-CMA	3	6	0
	MOEA/D-DE	3	6	0
	MOEA/D-DU	3	6	0
ZDT	MOEA/D-CMA	4	0	1
	MOEA/D-DE	4	0	1
	MOEA/D-DU	5	0	0

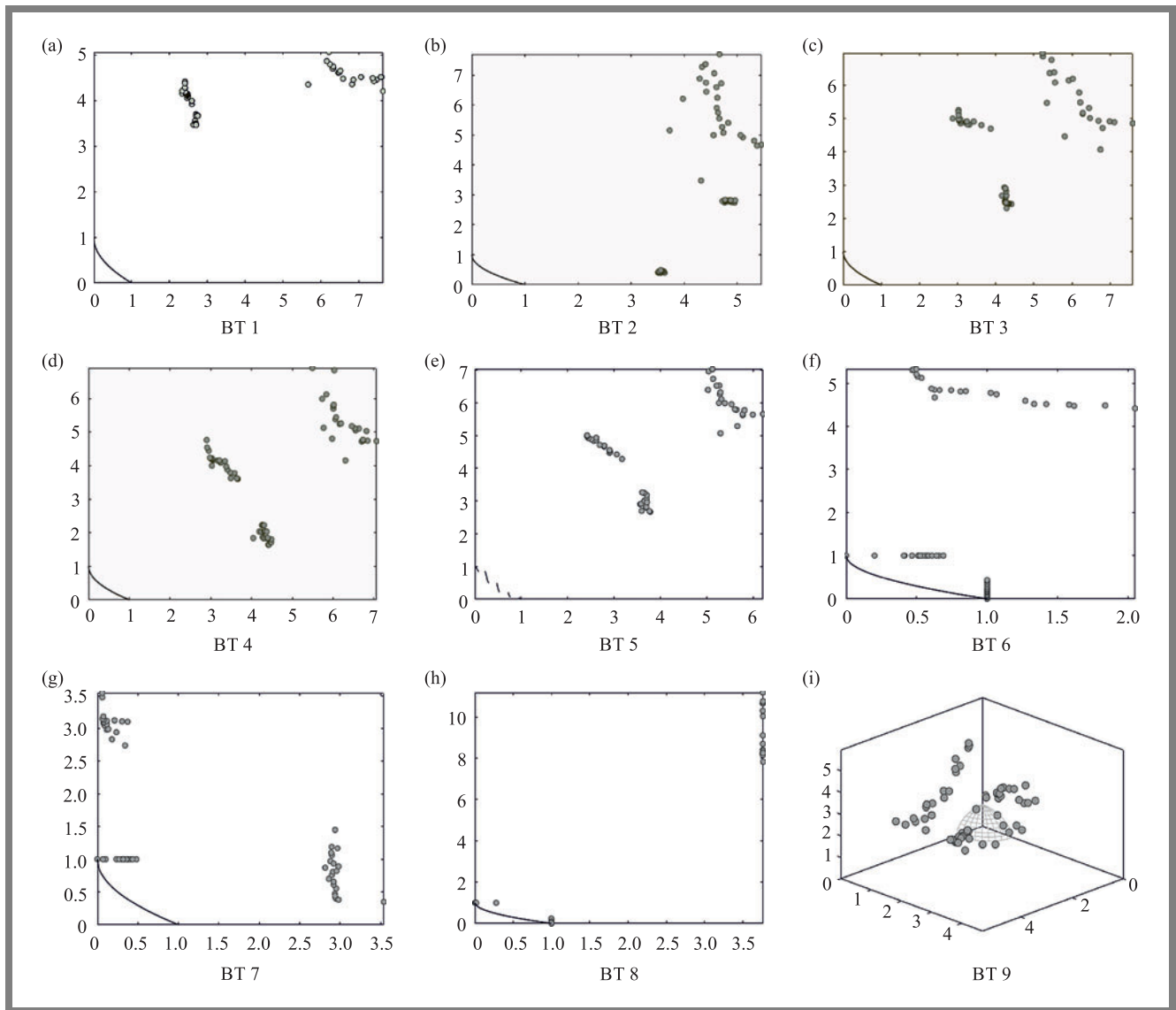


Fig. 2. Pareto front of BT-test suite. The axes are the objective values for BT1-BT8 test problems that are defined based on 2-objectives. Since BT9 is a 3-objective problem, the Pareto front has a 3-dimensional geometry. The solid curve represents the Pareto optimal front whereas the solid points depict the regions estimated by MOEA/D-MODE.

operator procedure in MOEA/D-MODE is superior or equivalent to state-of-the-art MOEA/D methods.

The proposed MOEA/D-MODE is specifically competitive when compared with two MOEA/D variants, i.e. MOEA/D-CMA and MOEA/D-DE. Test results verify that the crucial components of MOEA/D-MODE, i.e. multi-operator DE and parent selection schemes, facilitate reliable results to a greater extent than in other DE variants. However, the proposed algorithm has some room for improvement in handling functions with bias difficulties.

7.2. Further Discussion

The first concern is why the existing algorithms, i.e. MOEA/D-DE and MOEA/D-CMA are outperformed by MOEA/D-MODE. In fact, they fail to exhibit performance that would be on par with the proposed MOEA/D variant. We

suspect two potential reasons. Firstly, both state-of-the-art methods overly, emphasize the weight vectors that may be confined by only one solution or particular region. So it is likely to mislead from the corresponding area of PF and fail to preserve diversity. Secondly, normal parent selection criteria are applied. The procedures are biased towards preferring solutions from the local area in order to produce offspring. It is more likely that other regions in the objective-space may be overlooked. On the other hand, MOEA/D-MODE achieves better results in terms of selecting those solutions that have a fair aggregation score, but may be far from the weight vector. This has been even experimentally verified by using multiple mutation strategies during the evolutionary task.

The second concern is why MOEA/D-MODE fails to be better than the other solution when dealing with 3-objective optimization. Population size may be one of the critical reasons here. In the analysis, a normal population size of 100 is

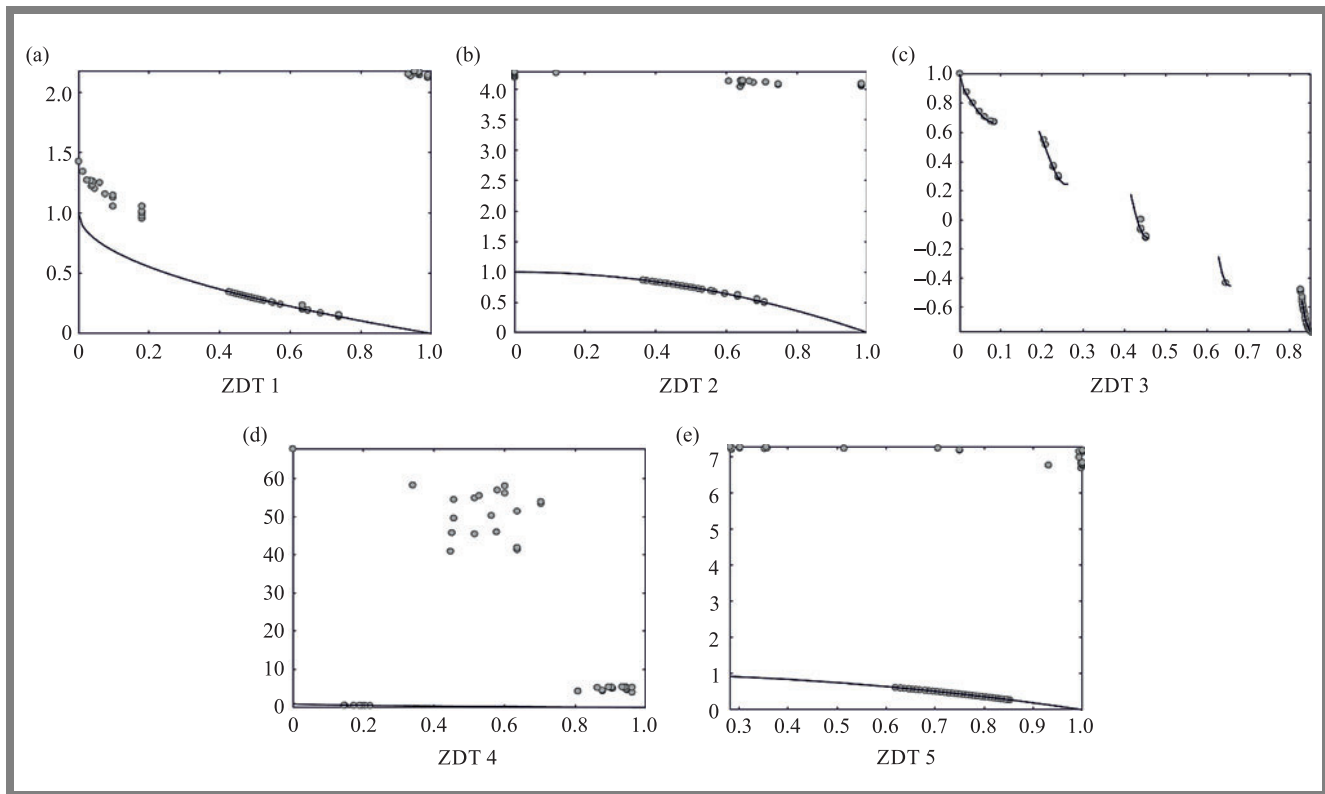


Fig. 3. Pareto front of ZDT-test suite. The axes are the objective values for test problems confined to a 2-dimensional space. The solid curve represents the Pareto optimal front whereas the solid points depict the regions estimated by MOEA/D-MODE.

Tab. 5. Summary of average IGD results and standard deviation (in brackets) compared between the MOEA/D-MODE algorithm and other algorithms. The best results are highlighted in bold print.

Function	M	D	MOEA/D-MODE	MOEA/D-CMA	MOEA/D-DE	MOEA/D-DU
BT1	2	30	3.955 (0.158)	3.851 (0.023)	3.894 (0.052)	3.996 (0.128)
BT2	2	30	2.262 (0.603)	1.61 (0.05)	1.73 (0.127)	1.405 (0.097)
BT3	2	30	3.844 (0.342)	3.939 (0.064)	3.957 (0.091)	3.947 (0.136)
BT4	2	30	3.913 (0.225)	3.796 (0.08)	3.845 (0.108)	3.715 (0.137)
BT5	2	30	3.929 (0.15)	3.87 (0.042)	3.929 (0.069)	3.97 (0.13)
BT6	2	30	0.676 (0.219)	2.341 (0.372)	1.844 (0.174)	2.041 (0.386)
BT7	2	30	0.819 (0.056)	1.555 (0.262)	1.023 (0.24)	1.323 (0.464)
BT8	2	30	0.81 (0.115)	5.254 (0.457)	4.32 (0.378)	3.834 (0.413)
BT9	3	30	3.711 (0.292)	3.085 (0.074)	3.43 (0.162)	3.206 (0.074)
ZDT1	2	30	0.226 (0.112)	0.152 (0.024)	0.419 (0.088)	0.405 (0.133)
ZDT2	2	30	0.274 (0.128)	0.173 (0.035)	0.697 (0.181)	1.087 (0.172)
ZDT3	2	30	0.194 (0.119)	0.299 (0.055)	0.459 (0.097)	0.219 (0.048)
ZDT4	2	10	0.513 (0.196)	6.035 (1.94)	3.941 (1.33)	42.143 (13.0)
ZDT6	2	10	0.161 (0.071)	0.023 (0.056)	0.029 (0.052)	3.7 (0.739)

used to converge the solutions to the PF. Perceptively, more solutions are required to bring the entire population to the PF in a higher-dimensional space. A smaller population size distributes the solutions sparsely in a high-order objective space. Thus, the sparse solutions fail to capture some areas from the entire PF, and this leads to a slow population convergence rate.

Poor performance of MOEA/D-MODE in ensuring faster convergence in the case of biased optimization problems, i.e. in the BT test suite, is the third concern. Despite its encouraging results concerning the evaluation of metrics (HV and IGD), it fails to show any superiority in terms of the convergence rate in BT test functions. We suspect that an

early termination of the algorithm is the reason here. As biases may cause large-scale changes objective vectors, the search operators need to remain strong. To achieve this, MAX_{FES} must be greater than 10,000, so that enough time is ensured for better exploitation of the regions. Apart from this, normal function problems, such as ZDT, have shown successful convergence with the standard procedures, as shown in Fig. 3.

8. Conclusion and Future Work

In this paper, a MOEA/D-MODE algorithm is proposed for solving multi-objective optimization problems and for improving the exploration-exploitation equilibrium. The concept is to put forth a multi-operator DE variant with complicated MOEA/D that ensures the distribution of the solutions throughout the evolutionary process. Specifically, in MOEA/D-MODE, the entire population is divided into multiple sub-populations, which are thereafter evolved by the assigned mutant operators of DE. In MOEA/D-MODE, we argue that the solution involves in the preference with respect to the proximity to the ideal position in the objective-space could improve the optimal results rather than relying upon the weight-vectors only.

We have analyzed the influence of multiple operators on the quality of MOEA/D-MODE, and several discussions have been conducted. We have shown that MOEA/D-MODE outperforms MOEA/D alternatives in terms of maintaining the convergence rate and distribution of solutions while solving MOP. Well-known test suites (BT and ZDT) with a total of 14 function instances have been employed to evaluate the algorithm's superiority. The results show that multiple mutation may achieve unprecedented results when coupled with MOEA/D.

In the future, we would extend our work to the high-dimension objective space. It would be interesting to address the problem of multiple-objective optimization with the concern of multi-operator evolutionary approach. We also would like to improve the outcomes of studies concerned with optimization problems involving bias difficulties.

References

- [1] C.A.C. Coello, D.A.V. Veldhuizen, and G.B. Lamont, "Evolutionary Algorithms for Solving Multi-Objective Problems", *Kluwer Academic*, 2007 (DOI: 10.1007/978-0-387-36797-2).
- [2] K.C. Tan, E.F. Khor, and T.H. Lee, "Multiobjective Evolutionary Algorithms and Applications (Advanced Information and Knowledge Processing)", *Springer*, 2005 (DOI: 10.1007/1-84628-132-6).
- [3] A. Zhou *et al.*, "Multiobjective evolutionary algorithms: A survey of the state of the art", *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 32–49, 2011 (DOI: 10.1016/j.swevo.2011.03.001).
- [4] J.D. Knowles and D.W. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy", *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, 2000 (DOI: 10.1162/106365600568167).
- [5] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization", *Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems*, K.C. Giannakoglou, D.T. Tsahalis, J. Périaux, K.D. Papailiou, and T. Fogarty, Eds. Athens, Greece: *Int. Center Numer. Methods Eng.*, pp. 95–100, 2001.
- [6] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002 (DOI: 10.1109/4235.996017).
- [7] Q. Zhang, A. Zhou, and Y. Jin, "RM-MEDA: A regularity model-based multiobjective estimation of distribution algorithm", *IEEE Trans. Evol. Comput.*, vol. 12, no. 1, pp. 41–63, 2008 (DOI: 10.1109/TEVC.2007.894202).
- [8] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search", *Parallel Problem Solving from Nature (PPSN VIII)*, vol. 3242, pp. 832–842, 2004 (DOI: 10.1007/978-3-540-30217-9_84).
- [9] J. Bader and E. Zitzler, "HypE: An algorithm for fast hypervolume-based many-objective optimization", *Evol. Comput.*, vol. 19, no. 1, pp. 45–76, 2011 (DOI: 10.1162/EVCO_a_00009).
- [10] D. Brockhoff, T. Wagner, and H. Trautmann, "R2 indicator-based multiobjective search", *Evol. Comput.*, vol. 23, no. 3, pp. 369–395, 2015 (DOI: 10.1162/EVCO_a_00135).
- [11] S. Jiang, J. Zhang, Y.-S. Ong, A.N. Zhang, and P.S. Tan, "A simple and fast hypervolume indicator-based multiobjective evolutionary algorithm", *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2202–2213, 2015 (DOI: 10.1109/TCYB.2014.2367526).
- [12] A. Jaszkievicz, "On the performance of multiple-objective genetic local search on the 0/1 knapsack problem – a comparative experiment", *IEEE Trans. Evol. Comput.*, vol. 6, no. 4, pp. 402–412, 2002 (DOI: 10.1109/TEVC.2002.802873).
- [13] Y. Jin, T. Okabe, and B. Sendho, "Adapting weighted aggregation for multiobjective evolution strategies", *Evolutionary Multi-Criterion Optimization*, pp. 96–110, 2001 (DOI: 10.1007/3-540-44719-9_7).
- [14] E.J. Hughes, "Multiple single objective Pareto sampling", *Proc. Congr. Evol. Comput.*, pp. 2678–2684, 2003 (DOI: 10.1109/CEC.2003.1299427).
- [15] H. Ishibuchi, T. Yoshida, and T. Murata, "Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling", *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 204–223, 2003 (DOI: 10.1109/TEVC.2003.810752).
- [16] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition", *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, 2007 (DOI: 10.1109/TEVC.2007.892759).
- [17] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II", *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 284–302, 2009 (DOI: 10.1109/TEVC.2008.925798).
- [18] H. Li, Q. Zhang and J. Deng, "Biased Multiobjective Optimization and Decomposition Algorithm", *IEEE Transactions on Cybernetics*, vol. 47, no. 1, pp. 52–66, 2017 (DOI: 10.1109/TCYB.2015.2507366).
- [19] Y. Yuan, H. Xu, B. Wang, B. Zhang, and X. Yao, "Balancing Convergence and Diversity in Decomposition-Based Many-Objective Optimizers", *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 2, pp. 180–198, 2016 (DOI: 10.1109/TEVC.2015.2443001).
- [20] N. Hansen and S. Kern, "Evaluating the CMA evolution strategy on multimodal test functions", *Parallel Problem Solving from Nature-PPSN VIII*, pp. 282–291, 2004 (DOI: 10.1007/978-3-540-30217-9_29).
- [21] A. Auger and N. Hansen, "A restart CMA evolution strategy with increasing population size", *Proc. Congr. Evol. Comput.*, pp. 1769–1776, 2005 (DOI: 10.1109/CEC.2005.1554902).
- [22] I. Loshchilov, "CMA-ES with restarts for solving CEC 2013 benchmark problems", *Proc. Congr. Evol. Comput.*, pp. 369–376, 2013 (DOI: 10.1109/CEC.2013.6557593).
- [23] N. Srinivas and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms", *Evol. Comput.*, vol. 2, no. 3, pp. 221–248, 1994.
- [24] J. Horn, N. Nafpliotis, and D.E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization", *Proc. 1st Int. Conf. Evol. Comput.*, pp. 82–87, 1994 (DOI: 10.1109/ICEC.1994.350037).
- [25] P.A.N. Bosman and D. Thierens, "The balance between proximity and diversity in multiobjective evolutionary algorithms", *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 174–188, 2003 (DOI: 10.1109/TEVC.2003.810761).
- [26] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach", *IEEE*

- Trans. Evol. Comput.*, vol. 3, no. 4, pp. 257–271, 1999 (DOI: 10.1109/4235.797969).
- [27] M. Fleischer, “The measure of Pareto optima applications to multi-objective metaheuristics”, *Proc. Evol. Multi-Criterion Optim.*, pp. 519–533, 2003 (https://doi.org/10.1007/3-540-36970-8_37).
- [28] O. Schütze, X. Esquivel, A. Lara, and C.A.C. Coello, “Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization”, *IEEE Trans. Evol. Comput.*, vol. 16, no. 4, pp. 504–522, 2012 (DOI: 10.1109/TEVC.2011.2161872).
- [29] J. Qi, Y. Yu, L. Wang, and J. Liu, “K*-Means: An Effective and Efficient K-Means Clustering Algorithm”, *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, pp. 242–249, 2016 (DOI: 10.1109/BDCloud-SocialCom-SustainCom.2016.46).
- [30] X. Ma, Q. Zhang, G. Tian, J. Yang, and Z. Zhu, “On Tchebycheff Decomposition Approaches for Multiobjective Evolutionary Optimization”, *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 2, pp. 226–244, 2018 (DOI: 10.1109/TEVC.2017.2704118).
- [31] R. Tanabe and A.S. Fukunaga, “Improving the search performance of shade using linear population size reduction”, *2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1658–1665, 2014 (DOI: 10.1109/CEC.2014.6900380).
- [32] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, “PlatEMO: A Matlab platform for evolutionary multi-objective optimization [educational forum]”, *IEEE Computational Intelligence Magazine*, vol. 12, no. 4, pp. 73–87, 2017 (DOI: 10.1109/MCI.2017.2742868).
- [33] E. Zitzler, K. Deb, and L. Thiele, “Comparison of multiobjective evolutionary algorithms: Empirical results”, *Evolutionary computation* 8.2, pp. 173–195, 2000 (DOI: 10.1162/106365600568202).
- [34] A. Zhou, Q. Zhang, Y. Jin, and B. Sendhoff, “Adaptive modelling strategy for continuous multi-objective optimization”, *Proc. Congr. Evol. Comput.*, pp. 431–437, 2007 (DOI: 10.1109/CEC.2007.4424503).
- [35] N. Beume, B. Naujoks, and M. Emmerich, “SMS-EMOA: Multiobjective selection based on dominated hypervolume”, *Eur. J. Oper. Res.*, vol. 181, no. 3, pp. 1653–1669, 2007 (DOI: 10.1016/j.ejor.2006.08.008).
- [36] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, “Theory of the hypervolume indicator: Optimal μ -distributions and the choice of the reference point”, *Proc. 10th ACM SIGEVO Workshop Found. Genet. Algorithms*, pp. 87–102, 2009 (DOI: 10.1145/1527125.1527138).
- [37] H. Ishibuchi, Y. Hitotsuyanagi, N. Tsukamoto, and Y. Nojima, “Many-objective test problems to visually examine the behavior of multiobjective evolution in a decision space”, *Proc. Int. Conf. Parallel Prob. Solv. Nat.*, pp. 91–100, 2010 (DOI: 10.1007/978-3-642-15871-1_10).
- [38] T. Wagner, N. Beume, and B. Naujoks, “Pareto-, aggregation-, and indicator-based methods in many-objective optimization”, *Proc. Evol. Multi-Criterion Optim.*, pp. 742–756, 2007 (DOI: 10.1007/978-3-540-70928-2_56).
- [39] X. Zou, Y. Chen, M. Liu, and L. Kang, “A new evolutionary algorithm for solving many-objective optimization problems”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 5, pp. 1402–1412, Oct. 2008 (DOI: 10.1109/TSMCB.2008.926329).
- [40] S. García, A. Fernández, J. Luengo, and F. Herrera, “Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power”, *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010 (DOI: 10.1016/j.ins.2009.12.010).



Sakshi Aggarwal is a Research Scholar (pursuing a Ph.D. degree) at the Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, India. She completed her M.Tech. degree in Software Engineering, from Galgotias University Greater Noida, India. Her research interests cover evolutionary algorithms, global optimization, and machine learning. Currently, she is working on multi-operator differential evolution variants used in feature selection applications.

E-mail: sakshiaggarwal@mnnit.ac.in

Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India



Krishn K. Mishra is presently working as an Assistant Professor at the Department of Computer Science and Engineering, MNNIT Allahabad, India. He has successfully organized around 6 IEEE conferences in India (ICCT Series) acting in the capacity of a conference secretary and has worked as a program chair for many other conferences. He is a regular reviewer of the *Journal of Supercomputing* (Springer), *Applied Intelligence*, *Applied Soft Computing*, *IEEE Transaction on Cybernetics*, *IEEE System Journal*, *Neural computing and application* and *IETE journals*.

E-mail: kkm@mnnit.ac.in

Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Multimodal Sarcasm Detection via Hybrid Classifier with Optimistic Logic

Dnyaneshwar Madhukar Bavkar¹, Ramgopal Kashyap¹, and Vaishali Khairnar²

¹Department of Computer Science and Engineering, Amity University, Raipur, Chhattisgarh, India,

²Department of Information Technology, Terna Engineering College, Nerul, Navi Mumbai, India.

<https://doi.org/10.26636/jtit.2022.161622>

Abstract — This work aims to provide a novel multimodal sarcasm detection model that includes four stages: pre-processing, feature extraction, feature level fusion, and classification. The pre-processing uses multimodal data that includes text, video, and audio. Here, text is pre-processed using tokenization and stemming, video is pre-processed during the face detection phase, and audio is pre-processed using the filtering technique. During the feature extraction stage, such text features as TF-IDF, improved bag of visual words, n-gram, and emojis as well on the video features using improved SLBT, and constraint local model (CLM) are extraction. Similarly the audio features like MFCC, chroma, spectral features, and jitter are extracted. Then, the extracted features are transferred to the feature level fusion stage, wherein an improved multilevel canonical correlation analysis (CCA) fusion technique is performed. The classification is performed using a hybrid classifier (HC), e.g. bidirectional gated recurrent unit (Bi-GRU) and LSTM. The outcomes of Bi-GRU and LSTM are averaged to obtain an effective output. To make the detection results more accurate, the weight of LSTM will be optimally tuned by the proposed opposition learning-based aquila optimization (OLAO) model. The MUSTARD dataset is a multimodal video corpus used for automated sarcasm discovery studies. Finally, the effectiveness of the proposed approach is proved based on various metrics.

Keywords — Bi-GRU, improved CCA, LSTM, multimodal sarcasm detection

Tab. 1. Nomenclature used.

Abbreviation	Description
AAM	Active appearance model
ALO	Ant lion optimization
AO	Aquila optimizer
BiGRU	Bi-directional gated recurrent unit
CAT	Convolution and attention
CCA	Canonical correlation analysis
CDVaN	Contextual dual-view attention network
CLM	Constraint local model
CMBO	Cat mouse-based optimization
CNN	Convolutional neural network
DL	Deep learning
DT	Decision tree
FDR	False discovery rate

FNR	False negative rate
FPR	False positive rate
HC	Hybrid classifier
IWAN	Incongruity-aware attention network
LBF	Local binary feature
LBP	Local binary pattern
LSTM	Long short term memory
MCC	Matthews's correlation coefficient
MFCC	Mel frequency cepstral coefficient
ML	Machine learning
NB	Naïve Bayes
NLP	Natural language processing
NN	Neural network
NPV	Net predictive value
OLAO	Opposition learning based aquila optimization
PCA	Principal component analysis
PRO	Poor and rich optimization
RF	Random forest
RNN	Recurrent neural network
SDS	Self-deprecating sarcasm
SLBT	Shape local binary texture
SSO	Social spider optimization
SVM	Support vector machine
TF-IDF	Term frequency-inverse document frequency

1. Introduction

Sarcasm is described as the use of remarks that imply the reverse of what one says, either to damage someone's feelings or to criticize something spectacularly [1]–[3]. It is a metaphorical language that is frequently used to communicate on social media, verbally and also with the use of the written text format. In the sarcasm sentiment, negative emotions are expressed via positive words found in the text, in order to expose their sarcasm [4], [5]. Tempo and speech time, variation, pitch level, and acoustic characteristics are all available in verbal sarcasm [6]. To demonstrate its sarcastic characteristics, this type of communication relies also

on tones and gestures, including eye and hand movements. Since no tone or gestures are available in sarcastic utterances represented in the text form, an ordinary person cannot recognize them. To detect sarcasm [7], an effective NLP approach is required for categorizing sarcastic features and properties within a sentence available in the text format [8]–[10].

Sarcasm was already characterized by NLP methods, where identification is described as the process of classifying a word or sentence sequence with sarcastic features and qualities by using NLP techniques [11]. It is also known as a system that learns and identifies ordinary and sarcastic sentences at the semantic level. Sentiment categorization is the basic goal of processes detecting sarcasm in a sentence. Due to its durability and ability to monitor itself based on specific datasets and requirements, the ML model [12]–[14] is frequently used for sarcasm detection [15], [16]. Sarcasm detection has proven to be useful in a variety of situations, as it allows businesses to analyze customers' reactions to their items, thus improving product quality [17]. It also aids in the elimination of incorrect categorization of customer views on problems, goods, and services. In human-computer interactions, sarcasm detection is also effective in conversation, system review rating, and summarization. For example, ML-based sarcasm identification [18] is used, relying on higher entropy, SVM, NN, window class, statistics, semantics, etc. In addition, an in-depth survey is conducted on automatic sarcasm detection methods, with a comparison of the scale of a given study, including the features, classification techniques, as well as performance parameters used. The survey is beneficial in identifying the newest trends in sarcasm detection [19]–[22].

The major contribution of this work is:

- BoW is newly defined along with other text-based features, like TF-IDF, n-gram,
- during the feature-level fusion phase, an improved multi-level CCA fusion technique is performed,
- OLAO model is implemented for weight optimization in LSTM.

In this work, a review of multimodal sarcasm detection methods is presented in Section 2. An overall description of the adopted multimodal sarcasm detection model is portrayed in Section 3. Pre-processing, feature extraction and level fusion processes are presented in Section 4. Section 5 describes a classification methods based on hybrid classifiers. Section 6 depicts the weight optimization of LSTM via an OLAO algorithm. The results are presented and discussed in Section 7. Section 8 concludes the paper, while Table 1 summarizes the nomenclature and abbreviations used.

2. Literature Review

Basavaraj *et al.* [23] suggested a method for detecting sarcasm in human words. The approach captures three types of data: voice, text, and temporal facial expressions to exploit the basic cognitive properties of human utterances. The data was unstructured because it contained dimensions of feelings and emotions that were used to produce sarcasm, with fa-

cial expressions being impacted by glottal and facial organs. The main effort focused on creating natural judgments in the prediction processes by employing cognitive data lineage information. It was difficult to identify sarcasm in genuine human conversations. Utilizing cloud resources, the multi-class NN model was applied as a soft cognition technique for detecting sarcasm. Voice cues and eye motions were examples of cognitive traits identified that might impact sarcasm detection.

Deepak *et al.* in [24] utilized DL in code-switch tweets to identify sarcasm, particularly in an Indian native language being a mixture of Hindi and English. The suggested system combined a softmax attention layer with Bi-LSTM and CNN for detecting real-time sarcasm. The SentiHindi feature vector was created employing pre-trained GloVe word embeddings and handmade features. The suggested softAttBiLSTM-feature-rich CNN model was compared and validated using performance assessment. With a classification accuracy of ~ 0.93 as well as an F-measure of ~ 0.89 , the system from [24] surpasses baseline DL techniques.

Wu *et al.* [25] created IWAN – an approach which uses a scoring method to identify sarcasm by concentrating on word-level incongruity among modalities. This scoring process might give words with incongruent modality a higher weight. The approach could capture word-level incongruity, resulting in greater performance and interpretability. The authors have added word-level characteristics for detecting multimodal sarcasm. In the MUsTARD dataset, they performed comprehensive comparison trials with 7 baseline models, but the model produced traditional outcomes. The benefits of the suggested IWAN algorithm were presented based on experimental findings that not only offered traditional performance on the MUsTARD dataset but also provided interpretability benefits.

Kamal *et al.* [26] demonstrated a DL strategy for identifying SDS on Twitter. They suggested a new CAT-BiGRU framework that comprises input, embedding, convolutional, two attention layers, and BiGRU. The SDS-based semantic and syntactic features in the embedding layers are extracted by the convolutional layer. Amazon word embedding as well as affective space and two SenticNet-based computing resources were determined to test the effectiveness of the suggested system. The authors concluded that DL-based techniques can reliably detect SDS in social media content based on the experimental results.

Eke *et al.* [27] conducted an analysis of sarcasm identification and classification strategies based on performance standards, datasets, classification models, feature engineering, and pre-processing. Text articles were studied during the research, with an emphasis placed on context and content-based language elements. Accuracy and precision metrics of such classification techniques as SVM, NB, RF, maximum entropy, and DT algorithm were measured and evaluated.

Kumar *et al.* [28] analyzed an empirical investigation of DL and shallow methods for detecting sarcasm used in text datasets. Using three predictive learning models, over 20,000

postings from Reddit and Twitter from the benchmark SemEval 2015 Task 11 were identified as sarcastic or non-sarcastic in this study. To generate the output, the first framework was developed based on TF-IDF weighted, which was trained through three classifiers, including gradient boosting, multinomial NB, and RF, as well as ensemble voting. The investigation compared the three learning approaches to classifying sarcasm into two datasets. It was discovered that the Bi-LSTM scheme achieved the maximum score for Reddit and Twitter datasets.

Ren *et al.* [29] suggested a CDVaN sarcasm identification model based on the sarcasm creation process. They used CDVaN for capturing contextual semantic information as well as for making the distinction between positive and negative situations in sarcasm. In contrast to the sarcasm-generating process, a multi-hop attention network was used to acquire contextual semantic information. Investigations on IAC-V2 as well as IAC-V1 datasets have shown that the suggested CDVaN system was capable of efficiently discriminating sarcasm. The model achieved state-of-the-art or equivalent performance, as per the findings.

Zheng *et al.* [30] identified sarcasm and irony on Twitter using several NLP and ML approaches. They discussed several research projects concentrating on irony and sarcasm to evaluate and clarify the meanings of such terms. The experiment was carried out by comparing several types of classification algorithms relying on some well-known text classification classifiers. The findings of this experiment suggest that ML approaches, particularly DL methods, were on the rise as the most promising for classification-related tasks. The F-score of the result was 0.89 and is comparable to the F-score of the sarcastic dataset.

Table 2 summarizes research projects focusing on multimodal sarcasm detection. The NN model determined in [23] offers a lower mean error rate, a high accuracy level and higher sensitivity. However, experiments involving benchmark datasets were not conducted in this work. SoftArt BiLSTM-feature-rich CNN model from [24] offers a higher classification accuracy level, a better recall rate, higher precision, and higher F-scores, but this model could not overfit based on dropout regularization. Moreover, the IWAN model deployed in [25] offers better precision, a higher recall rate, the best

Tab. 2. Review of multimodal sarcasm detection systems.

Paper	Adopted scheme	Features	Limitations	Dataset used	Effectiveness values
[23]	NN model	Better accuracy, lower mean error, higher sensitivity	Experiments on benchmark datasets were not conducted	Multi-modal sarcasm detection dataset	Overall accuracy is 78.57%
[24]	SoftArt BiLSTM-feature-rich CNN method	Superior classification accuracy, higher recall, better precision, higher F-score	This model could not overfit based on dropout regularization	The randomly sampled dataset contains 3000 sarcastic and 3000 non-sarcastic bilingual Hinglish (Hindi English) tweets	Classification accuracy is 92.71%
[25]	IWAN model	Better precision, higher recall, best F1-score, improved interpretability	Context incongruity was not investigated	Multi-modal sarcasm detection dataset	Overall accuracy is 93%
[26]	CAT-BiGRU model	Higher precision, better recall, improved F-score, higher accuracy	Multilingual data operation was not performed on multimodal platforms	Six benchmark datasets including Twitter dataset	Overall accuracy is 90%
[27]	ML algorithm	Best classifier accuracy, increased precision, higher recall, maximum F-score	Lack of a standard dataset was an issue in sarcasm identification	Sarcasm identification dataset	F-score is 73.5%
[28]	Multinomial NB model	Highest accuracy, higher recall, better precision, increased F1-score	Crowd-sourced or self-tagging datasets provide novel limitations for detecting the sarcastic tone	SemEval 2015 Task 11 and Kaggle's Reddit dataset	Overall accuracy is 86.32%
[29]	CDVaN model	Good effectiveness, better performance, lower error rate	Sarcasm related work was not continued owing to multi-modal data	IAC-V1 dataset and IAC-V2 dataset	Precision level is 76.32%
[30]	CNN model	Higher F-score, higher accuracy, larger correct rate	Different pre-processing approaches were not explored based on irony as well as sarcasm recognition	Semantic evaluation 2018 task 3: irony detection in English tweets	F1-score is 0.99%

F1-score, and improved interpretability. However, it failed to investigate context incongruity. Likewise, the CAT-BiGRU model from [26] offers higher precision, a better recall rate, an improved F-score, and higher accuracy. However, no multilingual data operations were performed on multimodal platforms. The ML algorithm was exploited in [27] and it has been determined that it offers the best classifier accuracy, an increased precision level, a higher recall rate and a maximum F-score. However, the lack of a standard dataset was an issue in sarcasm identification. The multinomial NB model from [28] offers the highest accuracy level. However, crowd-sourced or self-tagging datasets provide novel limitations related to detecting the sarcastic tone. The CDVaN model proposed in [29] is characterized by a lower error rate and ensures better performance and effectiveness. However, the sarcasm work was not continued due to multimodal data. Finally, the CNN model presented in [30] ensures better results, but different pre-processing approaches were needed to assure the quality of input data.

3. Multimodal Sarcasm Detection Model Adopted

This work introduces a new multimodal sarcasm detection model that comprises pre-processing, feature extraction, feature level fusion, and classification stages. First, the input text, video, and audio are subjected to the pre-processing stage. Next, the text content is pre-processed using tokenization and stemming. Video is pre-processed via face detection (Viola-Jones), and audio is pre-processed using the filtering technique (Butterworth filtering). Subsequently, the pre-processed text, video, and audio inputs are transferred to the feature extraction stage, where text features are extracted using TF-IDF, improved bag of words, n-gram, and emojis. Video features are extracted via improved SLBT and CLM. Audio features are extracted using MFCC, chroma, spectral features, and jitter. The extracted features are transferred to the feature level fusion phase, wherein an improved fusion technique is adopted. Classification is performed using a hybrid classifier

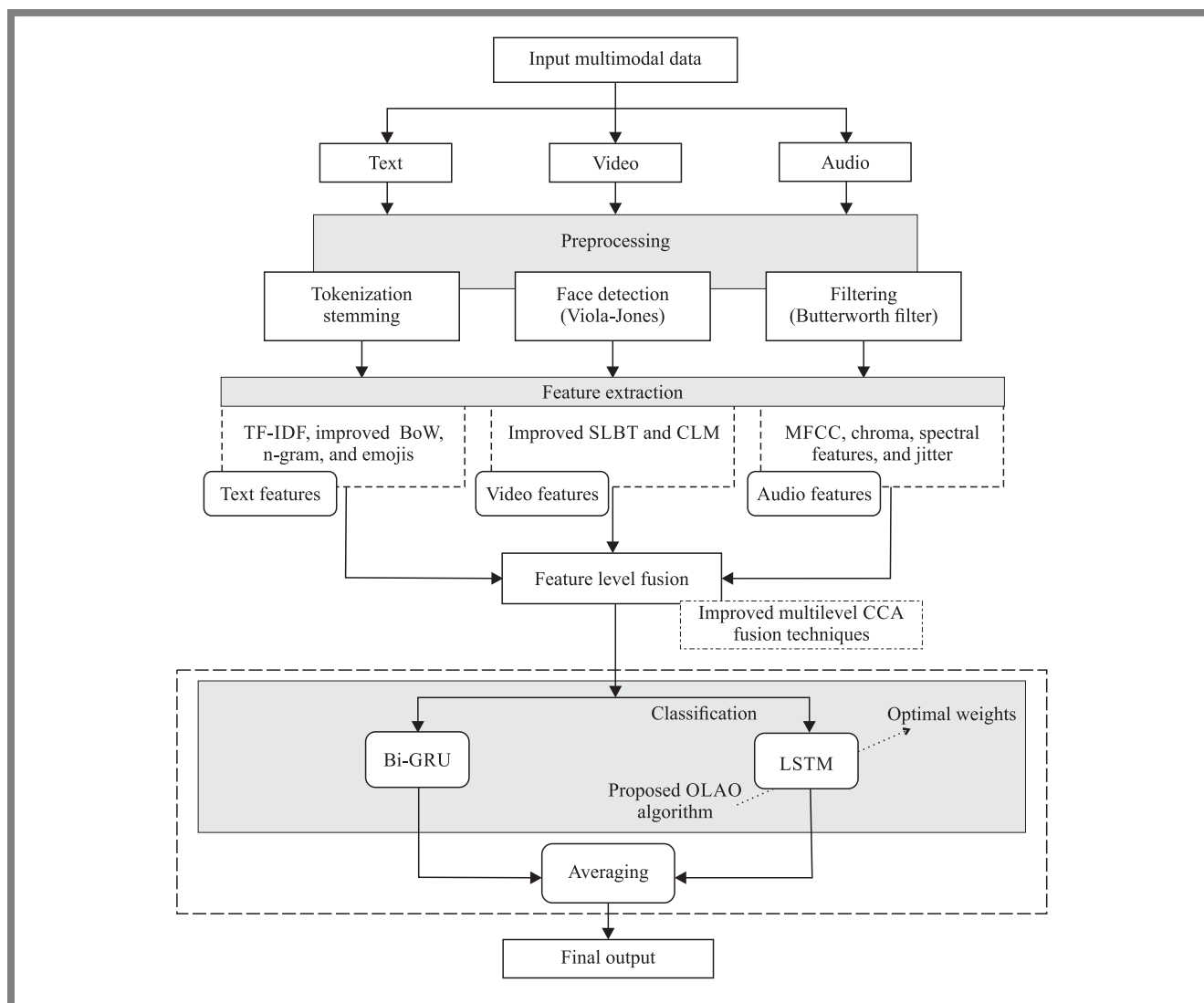


Fig. 1. Overall framework of the adopted model.

that combines LSTM and Bi-GRU by averaging the output of LSTM and Bi-GRU. To make the detection more precise and accurate, the weights of LSTM are tuned by a self-improved AO algorithm. The results show the presence of sarcasm in the given input.

Figure 1 illustrates the overall architecture of the adopted multimodal sarcasm detection model.

4. Model Details

4.1. Pre-processing Stage

Pre-processing is the initial and crucial process for successful learning. The input data is the multimodal data that includes text, audio, and video. Text is pre-processed using tokenization and stemming. Video is pre-processed via the face detection (Viola-Jones) model, and audio is pre-processed using the filtering technique (Butterworth filtering).

Tokenization [31] is the method of transforming text into tokens prior to vectorization. Undesirable tokens may be easily filtered off. For instance, a document may be divided into paragraphs or phrases broken down into words. This method consists in dividing large amounts of text into smaller chunks. Raw texts are broken down into words and phrases during the tokenization process as well. As a consequence, the tokens might aid in determining the NLP framework or understanding the context. By analyzing the word sequence, tokenization aids in determining the meaning of the text.

Tokenization may be accomplished using a variety of libraries and approaches. This task is carried out using such libraries as Keras, NLTK, and Gensim.

Stemming [31] is one of the normalizing strategies that reduce the number of calculations. It is a strategy for removing suffixes and retrieving the original words. During the stemming procedure, libraries such as Snowball Stemmer, Porter Stemming, and others are employed. Furthermore, stemming is mostly used to reduce data dimensionality. Stemming-related errors include under- and over-stemming.

Under stemming is characterized as false negatives which occur if 2 words are stemmed from the same stems and their roots are similar. Over stemming is viewed as a false-positive case that occurs when two words are stemmed from the same root but have different stems. Stemming is relied upon up information retrieval systems (i.e. Internet search engines) and other applications. It is also used in domain analysis to identify the existing domain vocabularies.

Face detection is difficult due to the numerous differences in the appearance of individual images, including facial expressions, pose variations, image orientation, occlusion, as well as lighting conditions. The Viola-Jones face detection method [32] is employed in this study.

It is an object detection approach capable of operating in real-time. Full view frontal upright faces are required for Viola-Jones. The approach, at its most basic level, reads an input image via a window, seeking human facial characteristics. When more characteristics are detected, the window in an

image is classified as a face. Further, the window should be resized and the process should be repeated to produce different size faces. For each window scale, the procedure is applied separately from other scales. To reduce the number of features, each window must be checked using a series of levels. Earlier levels have fewer features to verify and are thus simpler to pass, whereas later levels have more features and therefore are more difficult. The examination of features is performed at each level, and if the collected value does not meet the threshold, the level is considered failed and the specific window is not recognized as a face. The Viola-Jones face detection approach is divided into three key stages (integral image, classifier learning with AdaBoost, and attentional cascade structure) that allow for successful face detection in real-time applications.

The Butterworth filter [33] has a frequency response in the pass band that would be as flat as feasible. A maximally flat magnitude filter is another name of this particular filter. The Butterworth family of filters is very simple and useful. The cutoff frequency and filter order are the two key parameters used. Frequency response is monotonic and filter order affects the sharpness of the transition from the pass band to the stop band.

The poles linked with the squares of the frequency response magnitude are uniformly distributed in angle on concentric with the origin circle in the s-plane and containing a radius equal to the cut-off frequency for continuous time Butterworth filter. The poles that characterize the system function are easily acquired once the cutoff frequency and the filter order are established. One may easily design a differential equation that characterizes the filter after the poles have been determined. The squared magnitude function for an m -th order Butterworth low pass filter is:

$$|C(j\omega)|^2 = C(j\omega) \times C^*(j\omega) = \frac{1}{1 + \left(\frac{j\omega}{j\omega_c}\right)^{2m}}. \quad (1)$$

The first $2m - 1$ derivatives of $C(j\omega)^2$ at $\omega = 0$ are equal to 0 and the Butterworth response is maximally flat at $\omega = 0$. The derivative of the magnitude response is always negative for $+\omega$ and the magnitude response is minimized with ω . For $\omega \gg \omega_c$ the magnitude response is determined by:

$$|C(j\omega)|^2 = \frac{1}{\left(\frac{j\omega}{j\omega_c}\right)^2}. \quad (2)$$

4.2. Feature Extraction

The pre-processed text, video, and audio obtained are subjected to the feature extraction phase. From the text, such features as TF-IDF, n-grams, improved BoW, and emojis are extracted. TF-IDF [34] is a significant text demonstration format and includes a longer history when compared with the 3 well-known depiction techniques. It depends upon the BOW method, where a text is characterized by a compilation of words deployed in the document. The TF_{pq} constraint describes how many times word p appears in the document q . The better the value, the more noteworthy the word. The DFP constraint signifies the count of documents where p appears

once. If p is significant for q , it must comprise a higher TF_{pq} and lower DF_p . Hence, TF-IDF is determined as:

$$TD-IDF_{pq} = TF_{pq} \log \frac{M}{DF_p + 1}. \quad (3)$$

The extracted TF-IDF features are denoted as TF-IDF.

Any sequence of n tokens or words is called an n -gram. Moreover, an n -gram model [44] is defined as “a method of including sequences of words or characters that permits us to maintain richer pattern discovery in text, i.e. it attempts to captivate patterns of sequences (words or characters subsequent to one another) while being responsive to appropriate relations (words or characters subsequent to one another)”. The extracted n -gram-based features are denoted as $Ngram$.

BoW is the simplest technique used to transform the text into features. It separates words in the reviewed text into word count data and calculates the number of times a phrase appears in the corpus of a given text. It only cares about the order in which words appear in the text, not the sequence in which they appear frequently. The existing BoW evaluation does not consider the semantics of visual words, which is considered in the improved evaluation.

In the improved BoW, histograms of the visual words are used as a feature vector, such as:

$$K(P, Q) = \sum_{l=1}^L k(P_l, Q_l), \quad (4)$$

Where P and Q are images and l is visual word number. Then:

$$K(P_l, Q_L) = J_l^I + \sum_{i=0}^{I-1} \frac{1}{2^{I-i}} (J_l^i - J_l^{i+1}) S, \quad (5)$$

where I denotes the count and i indicates the current levels. Each level is weighted using $\frac{1}{2^{I-i}}$, J_l^i indicates the histogram intersection function, and S is the scaling factor.

$$J(H_{P_l}^i, H_{Q_L}^i) = \sum_{k=1}^{4^i} \min(H_{P_l}^i(k), H_{Q_L}^i(k)), \quad (6)$$

where the $H_{P_l}^i(k)$ denotes l -th count of visual words in the k -th subregion of image P at level i . The improved BoW characteristics are denoted by the $IBoW$ symbol.

Similarly, for texts with emojis a sentiment score based on unicode is extracted together with the text features, with regard to the emoji’s lexicon.

The position of an emoji is determined by its sentiment score as well as neutrality. The emotion score is between -1 and $+1$. Positive emojis are on the right-hand side of the map, while negative emojis are on the left-hand side. The most prevalent negative emoji is a sad face. The most common positive emojis include trophies, celebration symbols, hearts and a wrapped present – in addition to joyful smiles. Neutral emojis are classified using the neutrality range of 0 to 1 and all emojis have a sentiment score of 0. The extracted text with emoji features is denoted as EMO.

The overall extracted text features are denoted as TF = TF-IDF + $Ngram$ + $IBoW$ + EMO.

4.2.1. Video-based Features

From video content, features like improved SLBT and CLM are extracted. SLBT [10] is a feature that merges texture and shape characteristics. SLBT is identical to AAM, because it analyzes texture modeling using LBP texture features rather than intensity values. Consider $IM = [IM_1, IM_2, \dots, IM_{NO}]$ symbolizing a training set of pictures NO with $XP = [XP_1, XP_2, \dots, XP_{NO}]$ as shape landmark points. By matching these landmark points and then performing PCA on those points, shape variants may be achieved. Equation (7) determines any shape vector XP in the training set:

$$\begin{aligned} XP &\approx \overline{XP} + RS_{l_s} BD_{l_s}, \\ BD_{l_s} &= RS_{l_s}^T (XP - \overline{XP}), \end{aligned} \quad (7)$$

where \overline{XP} denotes the mean shape, RS_{l_s} includes the eigenvectors of the largest eigenvalues Ω_{l_s} and BD_{l_s} denotes weights or shape model parameters (e.g. l_s denotes the shape in BD_{l_s}). Such an approach may capture shape model parameters matching a given image by modifying Eq. (7).

To generate a shape-free patch, each training set image is warped into a mean shape for texture modeling. Computational complexity, efficiency, as well as processing time are mostly influenced by the size of the shape-free patch. For texture modeling in AAM, direct intensity values from a shape-free patch are required. To acquire illumination and noise invariant features, SLBT conducts LBP on a shape-free patch. Feature extraction using LBP is easier and faster than with Gabor wavelets.

Moreover, the shape vector and LBP vectors are used in SLBT. Unlike the LBP evaluation used in the conventional technique, improved LBP (geometric mean-LBP) is based on the comparison with neighboring pixels after comparison of the regional average RM of the image with the center pixel. Here, G_g indicates the neighboring pixel, s indicates the number of neighbors. The operation logic of ILBP is:

$$ILBP = \sum_{g=1}^{\delta} t 2^{g-1}. \quad (8)$$

In improved LBP, function t is determined by:

$$t = \begin{cases} 1, & \text{if } GM \geq G_b \text{ and } RM \leq G_g \\ 0, & \text{else if } RM \geq G_b \text{ and } RM > G_g \\ 1, & \text{else if } RM < G_b \text{ and } G_b \leq G_g \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

$$GM = \left(\prod_{g=1}^s G_g \right)^{\frac{1}{s}}, \quad (10)$$

where G_b indicates the center pixel and G_g denotes the neighboring pixel. The improved SLBT characteristics are denoted by the ISLBT symbol.

Consider $LI = [LI_1, LI_2, \dots, LI_{NO}]$ as the LBP feature histogram of all training sample images. Texture modeling (same as shape modeling) is accomplished with PCA given in Eq. (11). Here, $BD_{\tilde{t}}$ denotes the weights or texture mod-

eling parameter (\hat{t} refers to texture in $BD_{\hat{t}}$), $RS_{\hat{t}}$ indicates the eigenvectors and LI refers to the mean vector.

$$BD_t = RS_t^T (LI - \overline{LI}). \quad (11)$$

Using Eq. (12), a mixed shape and texture parameter vector are generated. Because shape (distance) and texture (intensity values) are measured using separate units, a diagonal matrix of weights WE_{ls} is computed for each shape parameter. By using PCA on the combined parameter vector as in Eq. (13), the shape texture parameter determining the texture, and local shape may be achieved.

$$BD_{lst} = \begin{pmatrix} WE_{ls} & BD_{ls} \\ & BD_t \end{pmatrix}, \quad (12)$$

$$CZ = RS_{lst}^T (BD_{lst} - \overline{BD_{lst}}). \quad (13)$$

In Eqs. (12), (13) RS_{lst} denotes the eigenvectors, $\overline{BD_{lst}}$ specifies the mean vector and CZ refers to the shape texture parameter. The LBF feature histogram is derived from a shapeless patch with five divisions along each row or column (e.g. 25 blocks).

If an annotated test image XP_{test} is provided as input, Eq. (7) is used to convert it into the shape model parameter BD_{test} which is then multiplied by WE_{ls} . The test image is distorted into a shapeless patch, by which LBP features are extracted as LI_{test} and the texture model parameter BD_{test} is computed by Eq. (11). From Eq. (13) CZ_{test} is employed for classification purposes and is generated by combining $BD_{ls\ test}$ and $BD_{t\ test}$ results in $BD_{lst\ test}$ as well as the shape texture parameter CZ_{test} .

CLM [36] is a group of approaches for identifying collections of points on a target picture that are bound by a statistical shape model. The main procedure aims to:

- sample a section of the image surrounding the current estimate and project it into a reference frame,
- create a “response image” for each point that shows the cost of having the point at each pixel,
- use the shape model parameters to identify a combination of points that minimizes the cost.

The optimum fit is discovered by minimizing the shape and pose parameters:

$$B(a, d) = \sum_{r=1}^{r=e} R_r [\hat{T}_d(X_r + Y_r a)]. \quad (14)$$

The term CLM is mainly referred to as a model used for creating response images using normalized correlation with a local patch, with the model patches being updated to match the current face while simultaneously being constrained by a global texture model. The CLM features are denoted by CLM.

The overall extracted video features are denoted as $VF = ISLBT + CLM$.

4.2.2. Audio-based Features

Features such as MFCC, chroma, spectral features, and jitter are extracted from audio content.

MFCCs [37] are frequently employed in speech recognition systems that can automatically recognize digits spoken into a phone. MFCCs are rapidly being used in music-related information retrieval applications, such as genre categorization apps and audio similarity measurements. The way you use the oral anatomy to produce each sound determines how it sounds. As a consequence, creating a description that encapsulates the physical mechanics of spoken language is one technique capable of uniquely identifying sounds. The method of encoding this data is to use MFCC features. The basic MFCC properties of the signal are provided by cepstral coefficients. On the other hand, additional characteristics, such as delta, acceleration, and energy can typically increase the accuracy. MFCC-based audio features are denoted by the MFCC symbol.

The 12 various pitch classes are referred to as chroma [38] features or chromagrams. Chroma-based characteristics, referred to as “pitch class profiles”, are useful for evaluating music with usefully classified pitches (typically in 12 groups) and for tuning which approximates the equal-tempered scale. Chromatic and melodic features of music are captured by chroma features which are responsive to changes in timbre as well as accompaniment. Chroma audio-based features are denoted by the CHR symbol.

Frequency and power characteristics of the signal are extracted using the spectral features block [39]. Filters may be used to remove undesirable frequencies. Such an approach is ideal for analyzing repeating signal patterns, including motions or vibrations from an accelerometer. Spectral characteristics are denoted by the SP symbol.

Jitter defines time distortions of phase and amplitude of the signal caused by clock deviation introduced during the analog-to-digital conversion. The effect of jitter increases with transmission distance and with the number of signal conversion stages. Jitter features are represented as Jitter.

The extracted audio features are denoted as $AF = MFCC + CHR + SRP + Jitter$. All extracted features combined are denoted by the FE symbol:

$$FE = TF + VF + AF. \quad (15)$$

4.3. Feature Level Fusion

The extracted text, video, and audio features are subjected to the feature-level fusion process. First, the audio and video features are transferred to CCA1 that produces an output and then the text features are transferred to CCA2. Next, the combined outcome of CCA1 and CCA2 is handed over to CCA3 to produce the final feature level fusion output. Figure 2 illustrates the feature-level fusion process.

Multilevel CCA [40] is a technique used for performing multivariate statistical analysis. The goal of CCA is to project two groups of multivariate data into an ordinary space with the highest possible correlation among them. The purpose of CCA in this situation is to discover a couple of column projection vectors $u_V \in \mathcal{R}^d$ and $u_Z \in \mathcal{R}^d$ in which the correlation among $u_V^T V$ and $u_Z^T Z$ is maximized, given

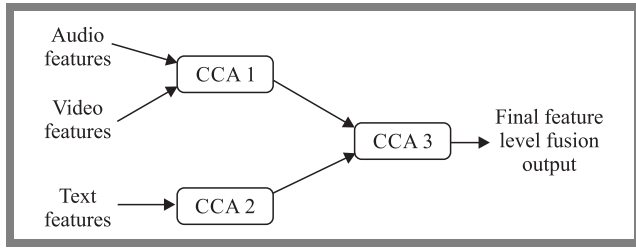


Fig. 2. Feature-level fusion process using multilevel CCA.

by two data matrices $V = \{V_v \in \mathbb{R}^d, v = 1, 2, \dots, \hat{K}\}$, $Z = \{Z_v \in \mathbb{R}^d, v = 1, 2, \dots, \hat{K}\}$, and \hat{K} pairings of data from two modalities. The objective function defined as the maximization function in this case is:

$$\arg \max_{u_L, u_S} \frac{u_V^T \hat{C}_{ZZ} u_Z}{\sqrt{u_V^T \hat{C}_{VV} u_V} \sqrt{u_Z^T \hat{C}_{ZZ} u_Z}}. \quad (16)$$

The data covariance matrices are $\hat{C}_{VV} = E[\mathbf{V}\mathbf{V}]^T$, $\hat{C}_{ZZ} = E[\mathbf{Z}\mathbf{Z}]^T$, and $\hat{C}_{VZ} = E[\mathbf{V}\mathbf{Z}]^T$.

$$\begin{aligned} & \text{maximize } u_V^T \hat{C}_{VZ} u_Z \\ & \text{Subject to } u_V^T \hat{C}_{VV} u_V = 1. \\ & \quad \quad \quad u_Z^T \hat{C}_{ZZ} u_Z = 1 \end{aligned} \quad (17)$$

Equation (17) is solved via generalized eigenvalue issues:

$$\begin{bmatrix} 0 & \hat{C}_{VZ} \\ \hat{C}_{ZV} & 0 \end{bmatrix} \begin{bmatrix} u_V \\ u_Z \end{bmatrix} = \lambda \begin{bmatrix} \hat{C}_{VV} & 0 \\ 0 & \hat{C}_{ZZ} \end{bmatrix} \begin{bmatrix} u_V \\ u_Z \end{bmatrix}, \quad (18)$$

where u_V denotes an eigenvector of $\hat{C}_{VV}^{-1} \hat{C}_{VZ} \hat{C}_{ZZ}^{-1} \hat{C}_{ZV}$ and u_Z indicates an eigenvector of $\hat{C}_{ZZ}^{-1} \hat{C}_{ZV} \hat{C}_{VV}^{-1} \hat{C}_{VZ}$. The projection matrices U_V and U_Z are attained via stacking u_V and u_Z as column vectors, respectively to various eigenvalue issues.

An improved correlation is determined in multi-level CCA as:

$$I_{corr} = 1 - \left(\frac{\sum_{\hat{c}=1}^{\hat{Q}} (\tilde{P}_{\hat{c}} - \bar{P}) (\tilde{R}_{\hat{c}} - \bar{R})}{\sqrt{\sum_{\hat{c}=1}^{\hat{Q}} (\tilde{P}_{\hat{c}} - \bar{P})^2} \sqrt{\sum_{\hat{c}=1}^{\hat{Q}} (\tilde{R}_{\hat{c}} - \bar{R})^2}} \right)^2. \quad (19)$$

5. Classification via Hybrid Classifiers

After the feature-level fusion, the classification process is performed using an optimized hybrid classifier (Bi-GRU and LSTM). Then, the outputs of both classifiers are averaged to determine the final outcome.

Through the use of a linear connection and a gate control unit, the LSTM network offers an efficient way to solve gradient desertion-related difficulties. As a consequence, the LSTM network caught the time-series data’s significant dependence. The sequences of persistent LSTM cells are included in LSTM [41] development. The input, output, and forget gates were all represented by three units in the LSTM cells. This feature enables the LSTM memory cells to suggest and store information for long periods of time.

Consider \tilde{H} and \tilde{C} as the hidden and cell state. Then $\tilde{H}_{\hat{t}}, \tilde{C}_{\hat{t}}$ and $F_{\hat{t}}, \tilde{C}_{\hat{t}-1}, \tilde{H}_{\hat{t}-1}$ represent the output and input layers, respectively. At time \hat{t} the output, input and forget gates are denoted as $O_{\hat{t}}, \tilde{I}_{\hat{t}}, \tilde{G}_{\hat{t}}$ respectively. The LSTM cell is primarily used $\tilde{G}_{\hat{t}}$ to filter the data. The modeling of $\tilde{G}_{\hat{t}}$ is:

$$\tilde{G}_{\hat{t}} = \kappa (W_{\tilde{L}} F_{\hat{t}} + h_{\tilde{L}} + W_j \tilde{H}_{\hat{t}-1} + h_j), \quad (20)$$

where W_j, h_j and $W_{\tilde{L}}, h_{\tilde{L}}$ specify the bias parameters and the weight matrix, respectively. The bias parameter and the weight factor are chosen randomly, while the weight factor is tuned optimally by the proposed OLAO model. The activation function of gate κ is elected as a sigmoid operation. Next, the LSTM cell makes use of the input gate to combine the proper data, as determined in Eqs. (21)–(23). $W_{\tilde{X}}, h_{\tilde{X}}$ and $W_{\tilde{Y}}, h_{\tilde{Y}}$ denote the weight matrices and the bias parameters which map the input and the hidden layers to the cell gate. W_x, h_x and W_y, h_y represent the weight and bias parameters that map the hidden and input layers to $\mathbf{IL}_{\hat{t}}$:

$$\tilde{U}_{\hat{t}} = \tanh (W_{\tilde{Y}} F_{\hat{t}} + h_{\tilde{Y}} + W_{\tilde{X}} \tilde{H}_{\hat{t}-1} + h_{\tilde{X}}), \quad (21)$$

$$\mathbf{IL}_{\hat{t}} = \kappa (W_y F_{\hat{t}} + h_y + W_x \tilde{H}_{\hat{t}-1} + h_x), \quad (22)$$

$$\tilde{C}_{\hat{t}} = \tilde{G}_{\hat{t}} \tilde{C}_{\hat{t}-1} + \mathbf{IL}_{\hat{t}} \tilde{f}_{\hat{t}}, \quad (23)$$

Finally, the LSTM obtains a hidden layer (output) from the output gate as:

$$o_{\hat{t}} = \kappa (W_{\hat{e}} F_{\hat{t}} + h_{\hat{e}} + W_{\hat{r}} \tilde{H}_{\hat{t}-1} + h_{\hat{r}}), \quad (24)$$

$$\tilde{H}_{\hat{t}} = o_{\hat{t}} \tanh (\tilde{C}_{\hat{t}}), \quad (25)$$

where $W_{\hat{e}}, h_{\hat{e}}$ and $W_{\hat{r}}, h_{\hat{r}}$ indicate the weight and bias parameters used for mapping the input and hidden layers to $o_{\hat{t}}$ respectively. The output of LSTM is denoted as CL_{LSTM} .

For organizing the sequential data stream, learning a continuous representation might be beneficial. An RNN is dedicated to encoding sequential data. Here, a Bi-GRU for learning the features from a sentence sequence, with GCN appending the outputs for DDI extraction afterward, is used. Bi-GRU [42] is broken down into two sections for calculation: forward and reverse sequence information transfers. The forward GRU for a given sentence $\tilde{Z} = (z_1, z_2, \dots, z_n), z \in \mathbb{S}^k, z$ signifies the current word concatenating vector. The forward GRU is:

$$\hat{i} = \sigma (w_{\hat{y}\hat{i}} \tilde{y}_{\hat{g}} + w_{\hat{h}\hat{i}} \hat{h}_{\hat{g}-1} + \tilde{a}_{\hat{i}}), \quad (26)$$

$$\tilde{l} = \sigma (w_{\tilde{y}\tilde{l}} \tilde{y}_{\hat{g}} + w_{\tilde{h}\tilde{l}} \hat{h}_{\hat{g}-1} + \tilde{a}_{\tilde{l}}), \quad (27)$$

$$\tilde{s} = \tanh (w_{\tilde{y}\tilde{s}} \tilde{y}_{\hat{g}} + w_{\tilde{h}\tilde{s}} (\hat{i} \Theta) \hat{h}_{\hat{g}-1} + \tilde{a}_{\tilde{s}}), \quad (28)$$

$$\hat{h} = (1 - \tilde{l}) \Theta \hat{h}_{\hat{g}-1} + \tilde{l} \Theta \tilde{s}, \quad (29)$$

where w_* and \tilde{a}_* are the weight matrix and the bias vector, respectively, σ denotes the sigmoid function, $\hat{h}_{\hat{g}}$ is the hidden state of the current time step \hat{g} , Θ is element-wise multiplication, and $\tilde{y}_{\hat{g}}$ is the input word vector at time step \hat{g} , $\tilde{h}_{\hat{g}}$ and $\tilde{h}_{\hat{g}}$ indicate the forward GRU and backward GRU output, respectively. The Bi-GRU output is indicated as $\hat{h}_{\hat{g}}^{\text{Bi-GRU}} = [\tilde{h}_{\hat{g}}; \tilde{h}_{\hat{g}}]$. The final classification output is:

$$\text{Out} = \frac{\text{CL}_{\text{LSTM}} + \hat{h}_{\hat{g}}^{\text{Bi-GRU}}}{2}. \quad (30)$$

6. LSTM Weight Optimization via OLAO

The weights of LSTM are tuned to optimal levels via the OLAO method adopted. Figure 3 illustrates the input solution to the adopted OLAO model. The total count of weights in LSTM is indicated as N . The final outputs of both Bi-GRU and LSTM are averaged to obtain the overall outcome. The error function is determined as $\text{error} = (1 - \text{accuracy})$. The objective function Obj of the implemented scheme is:

$$Obj = \min(\text{error}). \quad (31)$$

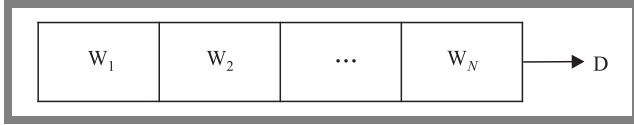


Fig. 3. Solution encoding.

6.1. Proposed OLAO Algorithm

Despite AO [43] offering better exploration capabilities, a good chance of reaching the optimal solution, and good exploitation-related abilities, it suffers from insufficient local exploitation ability. To overcome this problem, the OLAO model is proposed as an enhancement of the existing optimization models [44], [45]–[48]. AO is inspired by the behavior of hunting Aquila birds. The proposed OLAO concept deploys an OBL solution [49] that is modeled for generating opposite solutions. Specific points and their opposites are calculated simultaneously to select the best solution. OBL-based initialization guarantees an improved convergence rate, thus quickly reaching enhanced solutions.

6.2. Initialization of Solutions

The optimization rule relied upon in AO is a population-based method that starts with a population of candidate solutions D , as shown in Eq. (32). The said population is created stochastically between the lower LB and upper UB bounds of the specific issue. In each iteration, the best answer obtained is selected as the roughly optimal solution.

$$D = \begin{bmatrix} \tilde{q}_{1,1} & \dots & \tilde{q}_{1,j} & \tilde{q}_{1,Dim-1} & \tilde{q}_{1,Dim} \\ \tilde{q}_{2,1} & \dots & \tilde{q}_{2,j} & \dots & \tilde{q}_{2,Dim} \\ \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{q}_{A-1,1} & \dots & \tilde{q}_{A-1,j} & \dots & \tilde{q}_{A-1,Dim} \\ \tilde{q}_A & \dots & \tilde{q}_{A,j} & \tilde{q}_{A,Dim-1} & \tilde{q}_{A,Dim} \end{bmatrix}, \quad (32)$$

where D indicates the group of current candidate solutions that are created randomly in Eq. (33), $D_{\tilde{i}}$ represents the position of the \tilde{i} -th solution, A denotes the entire count of candidate solutions, and Dim refers to the dimension of the issue.

$$D_{\tilde{i}\tilde{j}} = \text{rand} \times (UB_{\tilde{j}} - LB_{\tilde{j}}) + LB_{\tilde{j}}, \quad (33)$$

$$\tilde{i} = 1, 2, \dots, A, \quad \tilde{j} = 1, 2, \dots, Dim,$$

where rand denotes a random number, $LB_{\tilde{j}}$ indicates the \tilde{j} -th lower bound, and $UB_{\tilde{j}}$ refers to the \tilde{j} -th upper bound of the issue.

6.3. AO mathematical Model

The proposed AO method imitates the behavior of Aquila's during each stage of the hunting at process. If $\tilde{l} \leq \frac{2}{3}\tilde{L}$ the exploration phases are exciting, it could move from exploration to exploitation steps using different behaviors, else, the exploitation phases are done.

The behavior of Aquilas is represented as a mathematical optimization framework whose goal is to find the optimum solution taking into consideration a given set of constraints. The mathematical model of the AO algorithm is determined as follows.

Step 1. Expanded exploration D_1 . In the first model D_1 , the Aquila identifies the its and chooses the optimal hunting location by soaring high in a vertical stoop. The AO requires high explorers to determine the area of the search space in which the prey is located. This behavior is represented as:

$$D_1(\tilde{l} + 1) = D_{\text{best}}(\tilde{l}) \cdot \left(1 - \frac{\tilde{l}}{\tilde{L}}\right) + [D_{\tilde{M}}(\tilde{l}) - D_{\text{best}}(\tilde{l}) \cdot \text{rand}], \quad (34)$$

where $D_1(\tilde{l} + 1)$ denotes the next iteration of the t solution that is produced by the initial search technique D_1 . This indicates the approximate location of the prey and $D_{\text{best}}(\tilde{l})$ is the best solution obtain until the \tilde{l} -th iteration. Expression $\frac{1-\tilde{l}}{\tilde{L}}$ is often used to regulate the number of iterations in the extended search (exploration). $D_{\tilde{M}}(\tilde{l})$ indicates the mean value of the current solutions linked at the time of iteration \tilde{l} -th, as given by Eq. (35). \tilde{l} and \tilde{L} represent the current iteration as well as the higher number of iterations, respectively.

$$D_{\tilde{M}}(\tilde{l}) = \frac{1}{A} \sum_{\tilde{i}=1}^A D_{\tilde{i}}(\tilde{l}) \quad \forall \tilde{j} = 1, 2, \dots, Dim, \quad (35)$$

where Dim denotes the dimension of the issue and A indicates the count of candidate solutions (population size).

Step 2. Narrowed exploration D_2 . Whenever the prey location is determined by a higher soar, the Aquila circles around the target, surveys the land and attacks using the second method D_2 . In anticipation of the attack, AO carefully investigates the specific region of the targeted prey. This behavior is formulated as:

$$D_2(\tilde{l} + 1) = D_{\text{best}}(\tilde{l}) \times \text{Levy}(\beta) + D_{\tilde{R}}(\tilde{l}) + (\vec{u} - \vec{v}) \cdot \text{rand}, \quad (36)$$

where $D_2(\tilde{l} + 1)$ denotes the next iteration of \tilde{l} solution, as determined by the second search procedure D_2 . β indicates the dimension space, the Levy flight distribution $\text{Levy}(\beta)$ functions given by Eq. (37), while $D_{\tilde{R}}(\tilde{l})$ denotes a random solution from the $1, \dots, A$ at \tilde{l} -th iteration.

$$\text{Levy}(\beta) = \bar{s} \cdot \frac{\bar{h} \times \vartheta}{|\bar{g}|^{\frac{1}{\rho}}}, \quad (37)$$

where \bar{s} denotes a constant value of 0.01, \hat{h} and \bar{g} denote random numbers between 0 and 1, ϑ is determined as:

$$\vartheta = \frac{\Gamma(1 + \rho) \cdot \sin e^{\frac{\pi\rho}{2}}}{\Gamma\frac{1+\rho}{2} \cdot \rho \cdot 2^{\frac{\rho-1}{2}}}, \quad (38)$$

where ρ denotes a constant value of 1.5. In Eq. (36), \bar{u} and \bar{v} present the spiral shape in the search, as determined in Eqs. (39), (40).

$$\bar{u} = \bar{d} \cos(\theta), \quad (39)$$

$$\bar{v} = \bar{d} \sin(\theta), \quad (40)$$

where:

$$\bar{d} = \bar{d}_1 + \bar{Z}\beta_1, \quad (41)$$

$$\theta = -\psi\beta_1 + \theta_1, \quad (42)$$

$$\theta_1 = \frac{3\pi}{2}. \quad (43)$$

\bar{Z} indicates a low value fixed at 0.00565, β_1 denotes a minor integer of 0.005 and ψ refers to an integer number from 1 to the length of the search area (Dim). \bar{d} and \bar{d}_1 assume a value between 1 and 20 for fixing the range of the search cycles. However, as per the proposed OLAO method, \bar{d} and \bar{d}_1 are randomly generated with $\tau = 2, 414$ as:

$$\bar{x}_{\bar{n}+1} = 2\tau |\bar{x}_{\bar{n}}| (1 - 2|\bar{x}_{\bar{n}}|), \quad 0 < \bar{x}_{\bar{n}} < 1. \quad (44)$$

Step 3. Expanded exploitation (D_3). Whenever the prey area is identified and the Aquila is ready to land and attack, the third method is used D_3 . This behavior is represented as:

$$D_3(\bar{l} + 1) = [D_{\text{best}}(\bar{l}) - D_{\bar{M}}(\bar{l})] \cdot \alpha - \text{rand} + [(UB - LB) \cdot \text{rand} + LB] \times \mu. \quad (45)$$

Here $D_3(\bar{l} + 1)$ denotes the solution of the next iteration \bar{l} , $D_{\text{best}}(\bar{l})$ indicates the prey's approximate location until the \bar{l} -th iteration (the greatest solution), and $D_{\bar{M}}(\bar{l})$ signifies the mean value of the current solution at the t -th iteration. α and μ are the exploitation modification parameters set to a minimum value of 0.1. LB indicates the problem's lower bound, and UB signifies the problem's upper bound.

Step 4. Narrowed exploitation D_4 . While the Aquila model prey in the 4-th method D_4 , it strikes over land depending on their stochastic motions. Such an approach is referred to as "walk and grab prey". AO attacks the prey at the last location. This behavior is described as:

$$D_4(\bar{l} + 1) = QF \cdot D_{\text{best}}(\bar{l}) - [\tilde{G}_1 \cdot D(\bar{l}) \cdot \text{rand}] - \tilde{G}_2 \cdot \text{Levy}(\beta) + \text{rand} \cdot \tilde{G}_1. \quad (46)$$

$D_4(\bar{l} + 1)$ denotes the solution of the fourth search method's iteration \bar{l} , and QF represents a quality function from Eq. (47), used to equalize the search techniques. \tilde{G}_1 represents different AO movements that are utilized to track the prey during the flight and is derived using Eq. (48). \tilde{G}_2 provides decreasing numbers from 2 to 0, reflecting the AO's flight slope used to follow the prey during its elope from the 1-st to the last (\bar{l}) location, as created by Eq. (49). $D(\bar{l})$ denotes the present

iteration's \bar{l} -th solution.

$$QF(\bar{l}) = \frac{2 \cdot \text{rand} - 1}{\bar{l}^{(1-\bar{L})^2}}, \quad (47)$$

$$\tilde{G}_1 = 2 \cdot \text{rand} - 1, \quad (48)$$

$$\tilde{G}_2 = 2 \cdot \left(1 - \frac{\bar{l}}{\bar{L}}\right). \quad (49)$$

$QF(\bar{l})$ is the \bar{l} -th iteration's quality function value, \bar{l} and \bar{L} show the current iteration as well as the higher count of iterations. Algorithm 1 illustrates the pseudo-code of the proposed OLAO model.

Algorithm 1. OLAO scheme adopted

Initialization phase

Population initialization D in AO

Initialize the AO parameters

As per the proposed OLAO model the OBL concept is deployed

while (the end condition is not met) do

 Compute the fitness function values:

$D_{\text{best}}(\bar{l})$

 for $\bar{i} = 1, 2, \dots, A$ do

 Mean value update $D_{\bar{M}}(\bar{l})$.

 Update \bar{v} , \bar{u} , \tilde{G}_1 , \tilde{G}_2 , $\text{Levy}(\beta)$, etc.

 if $\bar{l} \leq \frac{2}{3} \cdot \bar{L}$ then

 if $\text{rand} \leq 0.5$ then

 Step 1. Expanded exploration (D_1)

 Current solution update in Eq. (34)

 if $\text{Fitness}[D_1(\bar{l} + 1)] < \text{Fitness}[D(\bar{l})]$ then

$D(\bar{l}) = D_1(\bar{l} + 1)$

 if $\text{Fitness}[D_1(\bar{l} + 1)] < \text{Fitness}[D_{\text{best}}(\bar{l})]$ then

$D_{\text{best}}(\bar{l}) = (D_1(\bar{l} + 1))$

 end if

 end if

 else

 Step 2. Narrowed exploration (D_2)

 Current solution update in Eq. (36)

 if $\text{Fitness}[D_2(\bar{l} + 1)] < \text{Fitness}[D(\bar{l})]$ then

$D(\bar{l}) = D_2(\bar{l} + 1)$

 if $\text{Fitness}[D_2(\bar{l} + 1)] < \text{Fitness}[D_{\text{best}}(\bar{l})]$ then

$D_{\text{best}}(\bar{l}) = D_2(\bar{l} + 1)$

\bar{d} and \bar{d}_1

 are randomly generated as in Eq. (44)

 end if

 end if

 end if

 else

 if $\text{rand} \leq 0.5$ then

 Step 3. Expanded exploitation (D_3)

 Current solution update in Eq. (45)

 if $\text{Fitness}[D_3(\bar{l} + 1)] < \text{Fitness}[D(\bar{l})]$ then

$D(\bar{l}) = (D_3\bar{l} + 1)$

```

if Fitness[ $D_3(\tilde{l} + 1)$ ] < Fitness[ $D_{best}(\tilde{l})$ ] then
   $D_{best}(\tilde{l}) = (D_3\tilde{l} + 1)$ 
end if
end if
else
  Step 4. Narrowed exploitation ( $D_4$ )
  Current solution update in Eq. (46)
  if Fitness( $D_4[\tilde{l} + 1]$ ) < Fitness[ $D(\tilde{l})$ ] then
     $D(\tilde{l}) = (D_4\tilde{l} + 1)$ 
    if Fitness[ $D_4(\tilde{l} + 1)$ ] < Fitness[ $D_{best}(\tilde{l})$ ] then
       $D_{best}(\tilde{l}) = (D_4\tilde{l} + 1)$ 
    end if
  end if
end if
end if
end for
end while
Return ( $D_{best}$ )

```

7. Results and Discussions

The adopted multimodal sarcasm detection with HC + OLAO scheme was implemented in Python. The outcomes were computed for the extant schemes such as HC + PRO [50], HC + AO [43], HC + SSO [51], HC + CMBO [52], HC + ALO [53], CNN [54], RNN [55], RF [56], NB [57], Bi-GRU [26], and NN [23]. Furthermore, its performance was evaluated by varying the learning percentage metrics, such as precision, sensitivity, accuracy, specificity, FNR, FDR, F-score, MCC, FPR, rand index, and NPV, correspondingly.

Next, the authors extracted a representative sample from the collection of 6,365 annotated videos. The dataset obtained contained 690 movies with an equal amount of sarcastic and non-sarcastic classifications. The sample images are shown in Fig. 4.

7.1. Dataset Description

The dataset is taken from [58]. The MUsTARD dataset is a multimodal video corpus used for automated sarcasm discovery studies. The Big Bang Theory, The Golden Girls, Friends, and Sarcasmaholics Anonymous are just a few of the well-known TV programs that are included in the dataset. MUsTARD is a collection of sarcastic label-annotated audiovisual utterances accompanied by their context, which offers more details about the situation in which the utterance is made. A novel dataset (MUsTARD) comprises short videos that have been carefully annotated for their sarcastic feature, allowing researchers to investigate the topic. They chose to work with a balanced sample of sardonic as well as non-sarcastic clips to enable us to conduct our tests that expressly focus on the multimodal components of sarcasm.

7.2. Performance Analysis

The performance analysis of the presented HC + OLAO model is illustrated in Figs. 5–7. The adopted HC + OLAO scheme attains higher accuracy (0.86) for the learning rate of 60 percent (compared to the learning rate of 80 percent) than other existing schemes, as shown in Fig. 5a. This demonstrates the impact of the improved features on the text and video data, and the contribution of the optimization strategy to tuning the weights for better training results.

The HC + OLAO scheme attains higher sensitivity (0.98) (for a learning rate of 80 percent) than other extant schemes – see Fig. 5b. The proposed HC + OLAO scheme has shown a maximum precision value, ensuring better performance than other conventional models at the learning rate of 80 percent, as shown in Fig. 5c. This proves the impact of HC which gets trained with the suitable features. As the weights of LSTM are tuned optimally, the proposed HC + OLAO technique paves the way for better detection of the presence of sarcasm from multimodal inputs.

The metrics of the developed HC + OLAO scheme that are worse than those of the traditional approaches, including FPR, FNR, and FDR, are represented in Fig. 6. Similarly, the adopted HC + OLAO model attains the minimum FDR value for a learning rate of 80 percent, when compared with the learning rate of 80 percent, as shown in Fig. 6c. The HC + OLAO model proves that the lower FPR value offers better performance for the learning rate of 60 percent than the conventional models, as shown in Fig. 6b. The lower FNR (0.2) value of the proposed model means it is less prone to outcome errors at the learning rate of 70 percent, as depicted in Fig. 6a. The performance analysis has proven that the HC + OLAO scheme has converged with the objective (lower error).

Figure 7 represents other metrics analyzed, such as MCC, NPV, rand index, and F-score. The graph clearly illustrates that MCC of the HC + OLAO model attains a higher value (0.71) for learning the learning rate of 70 percent. However, existing models attain lower values, as shown in Fig. 7c. Similarly, the proposed model achieves the maximum NPV value (0.8) for a learning rate of 60 percent, compared to the learning rate of 80 percent, as shown in Fig. 7a. Likewise, the F-score for the learning rate of 70 percent is superior to other traditional schemes (Fig. 7b). The rand index for the learning rate of 60 percent achieves a higher value (0.99). Consequently, it has been proven that the presented HC + OLAO model surpasses other solutions in terms of performance.

7.3. Overall Performance Analysis

The overall performance analysis of the developed HC + OLAO scheme, comparing it with other models, is summarized in Tables 3–5 for learning rates of 60, 70 and 80 percent, respectively. The learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration. The proposed scheme achieves maximum accuracy values (0.86) to the extant approaches at the learning rate of 60 percent, and superior F-measure outcomes for the learn-

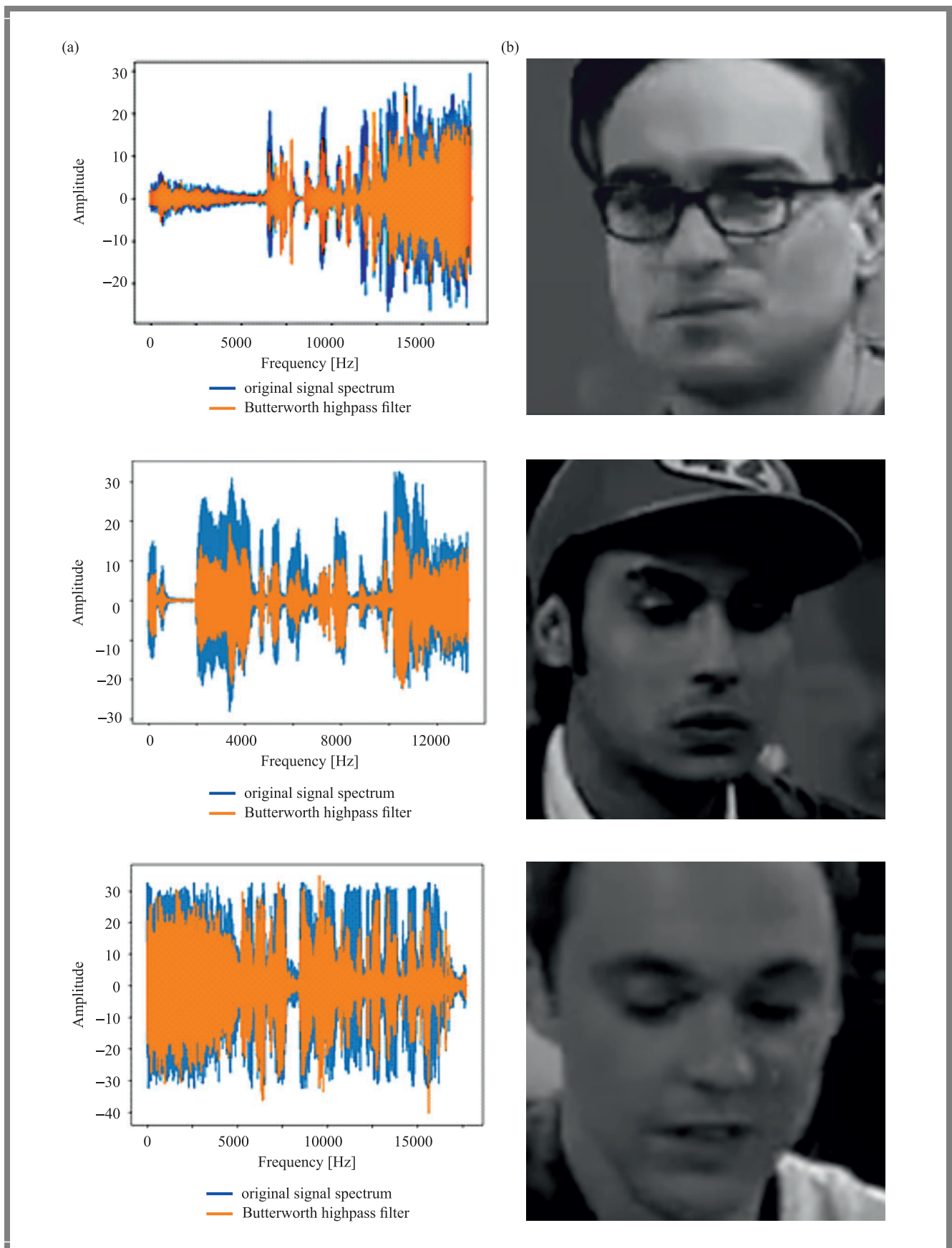


Fig. 4. Representation of: (a) audio preprocessing and (b) image preprocessing.

Tab. 3. Overall performance analysis for the learning rate of 60 percent.

Metric	Method											
	HC + PRO [50]	HC + AO [43]	HC + SSO [51]	HC + CMBO [52]	HC + ALO [53]	CNN [54]	RNN [55]	RF [56]	NB [54]	Bi-GRU [26]	NN [23]	HC + OLAO
FDR	0.40	0.41	0.30	0.45	0.47	0.15	0.21	0.45	0.21	0.33	0.49	0.11
Sensitivity	0.74	0.83	0.75	0.75	0.95	0.56	0.38	0.78	0.76	0.80	0.77	0.79
MCC	0.26	0.29	0.43	0.17	0.23	0.57	0.42	0.19	0.68	0.43	0.09	0.70
Precision	0.62	0.62	0.72	0.57	0.56	0.96	0.92	0.58	0.89	0.70	0.55	0.92
FPR	0.48	0.55	0.32	0.59	0.78	0.13	0.13	0.59	0.27	0.37	0.69	0.09
F-measure	0.68	0.71	0.74	0.65	0.71	0.71	0.54	0.66	0.82	0.75	0.64	0.85
Specificity	0.55	0.48	0.71	0.44	0.25	0.99	0.99	0.44	0.91	0.66	0.34	0.94
FNR	0.29	0.40	0.28	0.28	0.25	0.47	0.65	0.25	0.24	0.23	0.26	0.24
NPV	0.68	0.73	0.75	0.63	0.78	0.70	0.63	0.66	0.80	0.77	0.59	0.82
Accuracy	0.61	0.60	0.73	0.59	0.60	0.77	0.69	0.61	0.84	0.73	0.56	0.86
Rand index	0.82	0.82	0.87	0.78	0.78	0.89	0.84	0.79	0.92	0.87	0.76	0.94

Tab. 4. Overall performance analysis for the learning rate of 70 percent.

Metric	Method											
	HC + PRO [50]	HC + AO [43]	HC + SSO [51]	HC + CMBO [52]	HC + ALO [53]	CNN [54]	RNN [55]	RF [56]	NB [54]	Bi-GRU [26]	NN [23]	HC + OLAO
FDR	0.16	0.19	0.17	0.19	0.13	0.19	0.36	0.22	0.30	0.25	0.54	0.13
Sensitivity	0.96	0.95	0.92	0.27	0.69	0.36	0.72	0.23	0.81	0.86	0.67	0.83
MCC	0.26	0.24	0.23	0.37	0.61	0.49	0.64	0.33	0.70	0.66	0.02	0.72
Precision	0.57	0.56	0.57	0.97	0.90	1.03	0.91	0.96	0.88	0.83	0.49	0.91
FPR	0.77	0.77	0.74	0.15	0.29	0.14	0.28	0.15	0.33	0.21	0.65	0.11
F-measure	0.71	0.71	0.70	0.42	0.78	0.54	0.80	0.37	0.84	0.85	0.57	0.87
Specificity	0.26	0.26	0.29	1.01	0.94	1.03	0.94	1.01	0.89	0.83	0.38	0.92
FNR	0.21	0.25	0.34	0.76	0.34	0.67	0.31	0.80	0.39	0.34	0.36	0.20
NPV	0.81	0.78	0.74	0.59	0.76	0.65	0.78	0.58	0.82	0.86	0.56	0.85
Accuracy	0.61	0.60	0.60	0.64	0.81	0.72	0.83	0.63	0.85	0.84	0.52	0.87
Rand index	0.79	0.79	0.79	0.81	0.92	0.86	0.92	0.80	0.92	0.93	0.73	0.95

Tab. 5. Overall performance analysis for the learning rate of 80 percent.

Metrics	Methods											
	HC + PRO [50]	HC + AO [43]	HC + SSO [51]	HC + CMBO [52]	HC + ALO [53]	CNN [54]	RNN [55]	RF [56]	NB [54]	Bi-GRU [26]	NN [23]	HC + OLAO
FDR	0.42	0.40	0.22	0.38	0.22	0.33	0.20	0.20	0.36	0.38	0.48	0.19
Sensitivity	0.88	0.91	0.17	0.89	0.17	0.16	0.32	0.26	0.70	0.92	0.56	0.72
MCC	0.30	0.36	0.26	0.38	0.26	0.29	0.40	0.35	0.55	0.42	0.47	0.57
Precision	0.61	0.63	0.92	0.65	0.92	1.03	0.95	0.95	0.82	0.65	0.56	0.84
FPR	0.61	0.59	0.17	0.53	0.17	0.25	0.21	0.17	0.62	0.53	0.56	0.16
F-measure	0.72	0.74	0.29	0.75	0.29	0.28	0.48	0.40	0.76	0.76	0.56	0.78
Specificity	0.42	0.44	1.01	0.50	1.01	1.03	1.00	1.01	0.84	0.50	0.47	0.87
FNR	0.35	0.36	0.86	0.32	0.86	0.87	0.71	0.77	0.67	0.33	0.48	0.31
NPV	0.75	0.80	0.56	0.80	0.56	0.52	0.60	0.58	0.74	0.84	0.47	0.76
Accuracy	0.65	0.68	0.59	0.69	0.59	0.56	0.66	0.63	0.77	0.71	0.52	0.80
Rand index	0.82	0.83	0.78	0.85	0.78	0.76	0.83	0.81	0.88	0.85	0.73	0.91

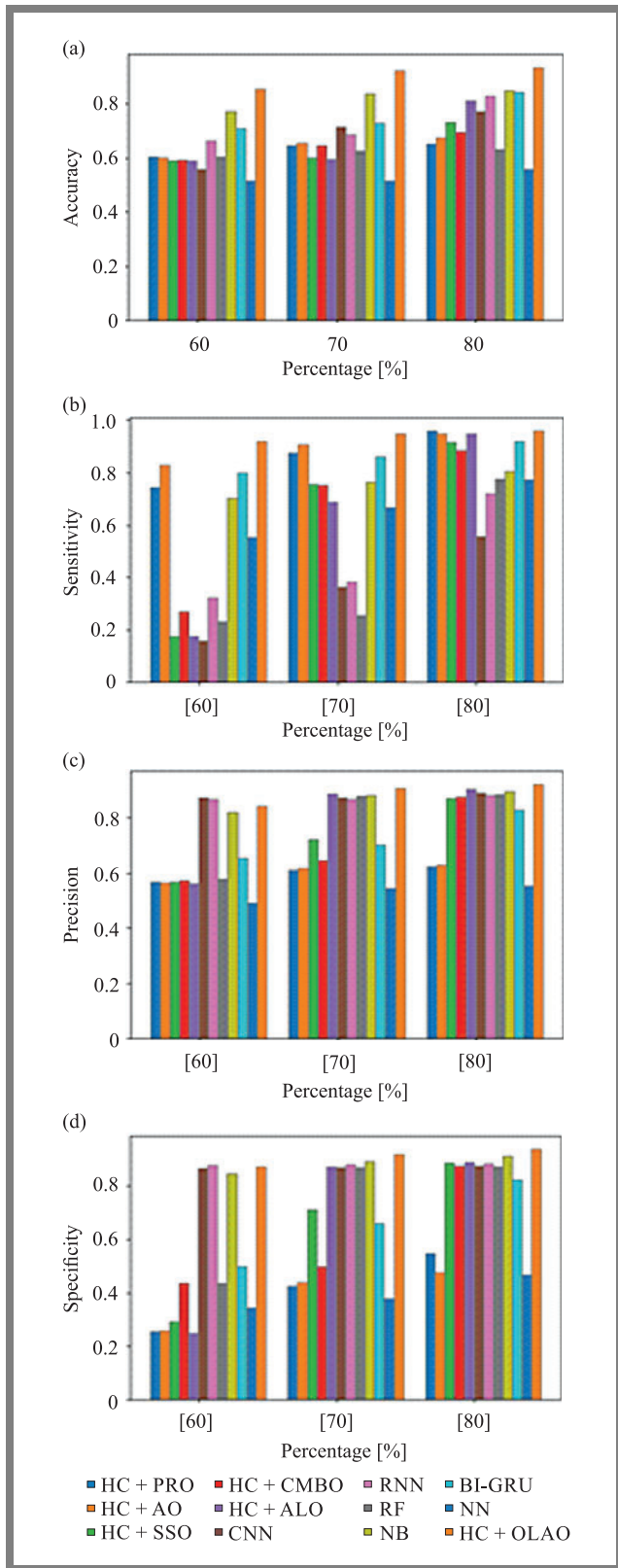


Fig. 5. Performance analysis of the adopted scheme to the extant approaches for: (a) accuracy, (b) sensitivity, (c) precision, and (d) specificity.

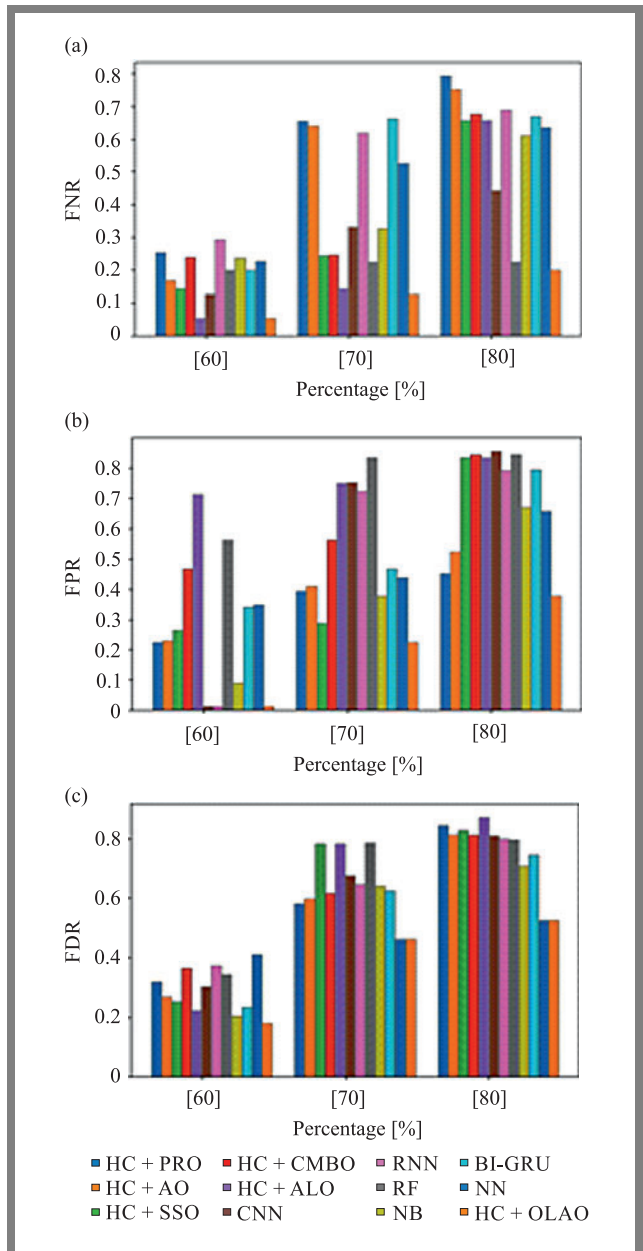


Fig. 6. Performance analysis of the adopted scheme to the traditional approaches for: (a) FNR, (b) FPR, and (c) FDR.

Tab. 6. Statistical analysis with respect to accuracy.

Metric	Std Dev.	Mean	Median	Best	Worst
HC + PRO [50]	0	1.21	1.21	1.21	1.21
HC + AO [43]	0.01	1.17	1.17	1.19	1.16
HC + SSO [51]	0.03	1.20	1.19	1.31	1.19
HC + CMBO [52]	0	1.32	1.32	1.32	1.32
HC + ALO [53]	0.04	1.18	1.16	1.26	1.16
HC + OLAO	0.01	1.16	1.15	1.21	1.15

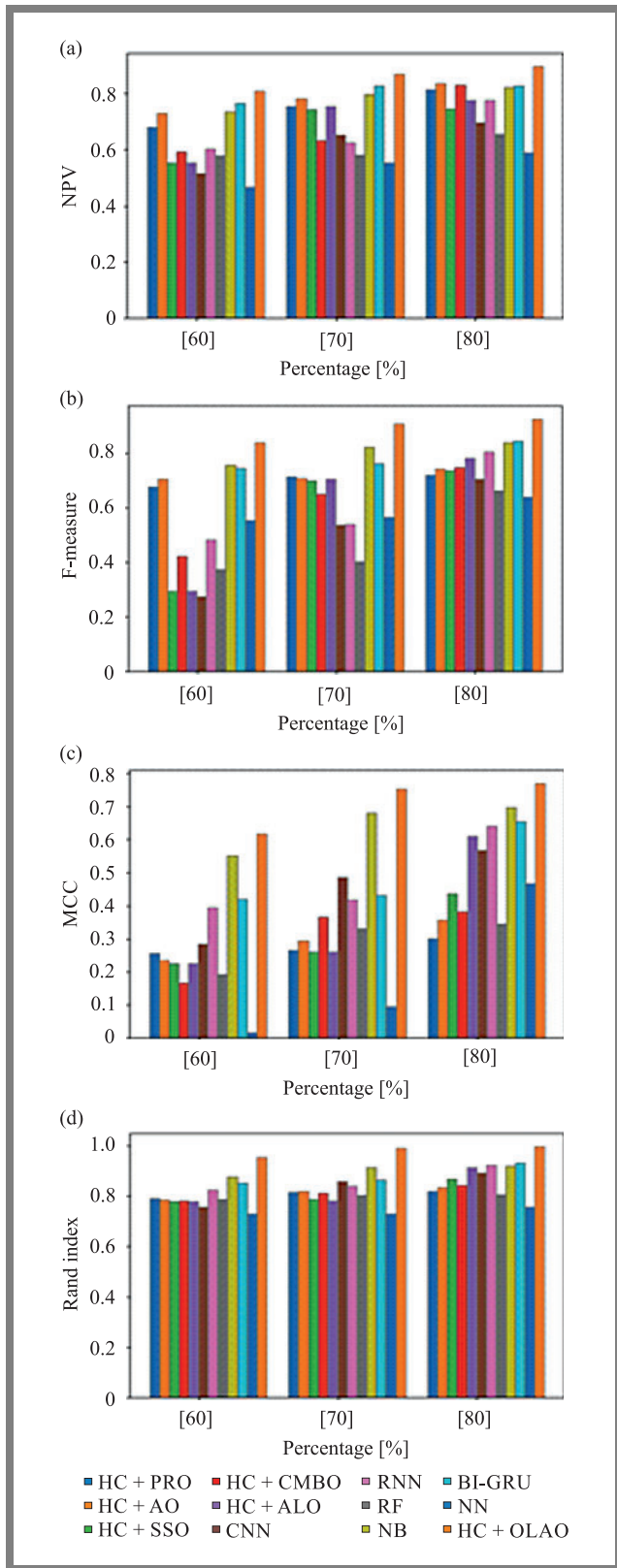


Fig. 7. Performance analysis of the adopted scheme to the traditional approaches for: (a) NPV, (b) F-measure, (c) MCC, and (d) rand index.

ing rate of 70 percent. However, the existing models show the worst performance, as they suffer from lower convergence speed for error minimization purposes.

7.4. Statistical Analysis

The statistical analysis of the proposed approach versus the existing scheme, based on the accuracy metric, is presented in Table 5. The best-case scenario proves an enhancement of the accuracy of results achieved by the proposed HC + OLAO model (1.21), with the said results surpassing the values obtained with the use of other models. Mean performance shows better outcomes for accuracy-related metrics. Therefore, the proposed model has proved to be more effective in multimodal sarcasm detection, almost in all scenarios.

7.5. Features Analysis

The feature-based analysis of the proposed model, with an without relevant comparisons, is illustrated in Table 7.

Also in this case the proposed HC+OLAO model offers better accuracy than the with conventional BoW, without optimization, model with conventional SLBT, and without feature level fusion, respectively. Further, the proposed HC+OLAO model holds lower FNR (0.24) with better performance. This im-

Tab. 7. Analysis based on features type of proposed model.

Metric	Without optimization	With conventional BoW	With conventional SLBT	Without feature level fusion	HC + OLAO
Accuracy	0.62	0.85	0.75	0.70	0.86
Sensitivity	0.80	0.77	0.82	0.97	0.79
Specificity	0.45	0.92	0.68	0.43	0.94
Precision	0.60	0.90	0.72	0.69	0.92
FNR	0.26	0.24	0.24	0.32	0.24
F-measure	0.68	0.83	0.77	0.81	0.85
MCC	0.20	0.69	0.45	0.12	0.70
FPR	0.61	0.09	0.38	0.87	0.09
NPV	0.67	0.80	0.79	0.74	0.82
FDR	0.46	0.10	0.34	0.61	0.11
Rand	0.81	0.92	0.89	0.96	0.94

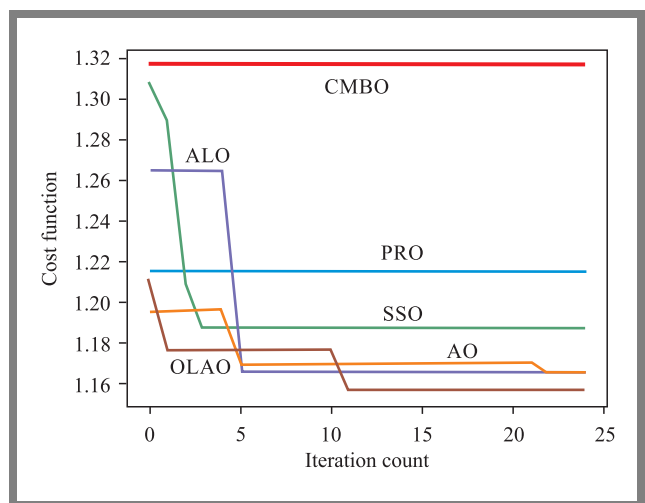


Fig. 8. Convergence analysis of the proposed and other approaches.

plies that the combination proposed in the system is suitable for multimodal sarcasm detection.

7.6. Convergence Analysis

The convergence of the adopted OLAO model is examined and compared with that of the traditional schemes by varying the iteration count between 0, 5, 10, 1, 20, and 25, respectively. Figure 8 illustrates the convergence analysis of the presented method, compared with the traditional schemes. The cost function of the OLAO model is minimized as the count of iterations increases. In addition, the cost function began to decrease from 10–12 iterations. The cost function provides a lower constant value (1.15) for 12–25 iterations than other existing models, such as PRO, AO, SSO, CMBO, and ALO. The proposed OLAO approach achieves the minimum cost function as per the objectives defined in Eq. (27).

Therefore, it is proven that the adopted OLAO approach returns a lower cost function with superior outcomes.

8. Conclusion

This work has identified a new multimodal sarcasm detection method that includes four stages: pre-processing, feature extraction, feature level fusion, and classification. The extracted features were subjected to feature level fusion. In this phase, an improved multilevel CCA fusion technique was applied. The classification was performed using HC solutions, such as LSTM and Bi-GRU. Finally, the outputs of LSTM and Bi-GRU were averaged to obtain an effective output. In order to render the detection method more accurate and precise, the weight of LSTM was tuned using the proposed OLAO model. The final result showed whether any sarcasm was present or not in the analyzed sample. Finally, the results of the adopted technique were compared with the extant methods, with various metrics, including F-score, FDR, specificity, FPR, accuracy, FNR, sensitivity, precision, NPV, rand index, and MCC, taken into consideration. The mean performance of the adopted HC + OLAO approach is better in terms of accuracy-related metrics, when compared with traditional schemes, such as HC + PRO, HC + AO, HC + SSO, HC + CMBO, and HC + ALO.

References

- [1] K. Nimala, R. Jebakumar, and M. Saravanan, "Sentiment topic sarcasm mixture model to distinguish sarcasm prevalent topics based on the sentiment bearing words in the tweets", *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 6801–6810, 2021 (DOI: 10.1007/s12652-020-02315-1).
- [2] Y. Kumar and N. Goel, "AI-Based Learning Techniques for Sarcasm Detection of Social Media Tweets: State-of-the-Art Survey", *SN Comput. Sci.*, vol. 1, no. 6, 2020, (DOI: 10.1007/s42979-020-00336-3).
- [3] A. Banerjee, M. Bhattacharjee, K. Ghosh *et al.*, "Synthetic minority oversampling in addressing imbalanced sarcasm detection in social media", *Multimed. Tools Appl.*, vol. 79, pp. 35995–36031, 2020 (DOI: 10.1007/s11042-020-09138-4).
- [4] R. Justo, J.M. Alcaide, M.I. Torres *et al.*, "Detection of Sarcasm and Nastiness: New Resources for Spanish Language", *Cogn. Comput.*, vol. 10, pp. 1135–1151, 2018 (DOI: 10.1007/s12559-018-9578-5).
- [5] R.A. Potamias, G. Siolas, and A. Stafylopatis "A transformer-based approach to irony and sarcasm detection", *Neural Comput. & Applic.*, vol. 32, pp. 17309–17320, 2020 (DOI: 10.1007/s00521-020-05102-3).
- [6] Y. Du, T. Li, M.S. Pathan *et al.*, "An Effective Sarcasm Detection Approach Based on Sentimental Context and Individual Expression Habits", *Cogn. Comput.*, vol. 14, pp. 78–90, 2021 (DOI: 10.1007/s12559-021-09832-x).
- [7] L. Ren, B. Xua, H. Lin, X. Liu, and L. Yang, "Sarcasm Detection with Sentiment Semantics Enhanced Multi-level Memory Network", *Neurocomputing*, vol. 401, pp. 320–326, 2020 (DOI: 10.1016/j.neucom.2020.03.081).
- [8] M.S. Razali, A.A. Halin, L.S.Y. Doraisamy, and N.M. Norowi, "Sarcasm Detection Using Deep Learning With Contextual Features", *IEEE Access*, vol. 9, pp. 68609–68618, 2021 (DOI: 10.1109/ACCESS.2021.3076789).
- [9] S. Rathod, "Hybrid Metaheuristic Algorithm for Cluster Head Selection in WSN", *Journal of Networking and Communication Systems*, vol. 3, no. 4, 2020 (DOI: 10.46253/jnacs.v3i4.a1).
- [10] N.S. Lakshmi Prabha and S. Majumder, "Face recognition system invariant to plastic surgery", *12th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 258–263, 2012 (DOI: 10.1109/ISDA.2012.6416547).
- [11] A. Onan and M.A. Toçoğlu, "A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification", *IEEE Access*, vol. 9, pp. 7701–7722, 2021 (DOI: 10.1109/ACCESS.2021.3049734).
- [12] Meherkandukuri, "Deep Convolutional Neural Network for Emotion Recognition via EEG Signal", *Journal of Computational Mechanics, Power System and Control*, vol. 4, no. 2, 2021 (DOI: 10.46253/jcmps.v4i2.a3).
- [13] S. Rajeyyagari, "Automatic speaker diarization using deep LSTM in audio lecturing of e-Khool platform", *Journal of Networking and Communication Systems*, vol. 3, no. 4, 2020 (DOI: 10.46253/jnacs.v3i4.a3).
- [14] J. Russel Fernandis, "ALOA: Ant Lion Optimization Algorithm-based Deep Learning for Breast Cancer Classification", *Multimedia Research*, vol. 4, no. 1, (DOI: 10.46253/j.mr.v4i1.a5).
- [15] C.I. Eke, A.A. Norman, and L. Shuib, "Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model", *IEEE Access*, vol. 9, pp. 48501–48518, 2021 (DOI: 10.1109/ACCESS.2021.3068323).
- [16] Y. Diao, *et al.*, "A Multi-Dimension Question Answering Network for Sarcasm Detection", *IEEE Access*, vol. 8, pp. 135152–135161, 2020 (DOI: 10.1109/ACCESS.2020.2967095).
- [17] A. Kumar, V.T. Narapareddy, V. Aditya Srikanth, A. Malapati, and L.B.M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM", *IEEE Access*, vol. 8, pp. 6388–6397, 2020 (DOI: 10.1109/ACCESS.2019.2963630).
- [18] Y. Zhang *et al.*, "CFN: A Complex-Valued Fuzzy Network for Sarcasm Detection in Conversations", *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 12, pp. 3696–3710, 2021 (DOI: 10.1109/TFUZZ.2021.3072492).
- [19] K. Rothermich, A. Ogunlana, and N. Jaworska, "Change in humor and sarcasm use based on anxiety and depression symptom severity during the COVID-19 pandemic", *Journal of Psychiatric Research*, vol. 140, pp. 95–100, 2021 (DOI: 10.1016/j.jpsychires.2021.05.027).
- [20] P. Parameswaran, A. Trotman, and D. Eysers, "Detecting the target of sarcasm is hard: Really?", *Information Processing and Management*, vol. 58, no. 4, 2021 (DOI: 10.1016/j.ipm.2021.102599).
- [21] N.Z.Z. Wang, "The paradox of sarcasm: Theory of mind and sarcasm use in adults", *Personality and Individual Differences*, vol. 163, 2020 (DOI: 10.1016/j.paid.2020.110035).
- [22] R. Pandey, A. Kumar, J.P. Singh, and S. Tripathi, "Hybrid attention-based Long Short-Term Memory network for sarcasm identification", *Applied Soft Computing*, vol. 106, 2021 (DOI: 10.1016/j.asoc.2021.107348).
- [23] N. Basavaraj Hiremath, and M.M. Patil, "Sarcasm Detection using Cognitive Features of Visual Data by Learning Model", *Expert Systems with Applications*, vol. 184, 2021 (DOI: 10.1016/j.eswa.2021.115476).
- [24] D. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-

- up language using soft-attention based bi-directional LSTM and feature-rich CNN”, *Applied Soft Computing*, vol. 91, 2020 (DOI: 10.1016/j.asoc.2020.106198).
- [25] Y. Wu *et al.*, “Modeling Incongruity between Modalities for Multimodal Sarcasm Detection”, *IEEE MultiMedia*, vol. 28, no. 2, pp. 86–95, 2021, (DOI: 10.1109/MMUL.2021.3069097).
- [26] A. Kamal and M. Abulaish “CAT-BiGRU: Convolution and Attention with Bi-Directional Gated Recurrent Unit for Self-Deprecating Sarcasm Detection”, *Cogn. Comput.*, vol. 14, pp. 91–109, 2022 (DOI: 10.1007/s12559-021-09821-0).
- [27] C.I. Eke, A.A. Norman, S. Liyana, and H.F. Nweke, “Sarcasm identification in textual data: systematic review, research challenges and open directions”, *Artif. Intell. Rev.*, vol. 53, pp. 4215–4258, 2020 (DOI: 10.1007/s10462-019-09791-8).
- [28] A. Kumar and G. Garg, “Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets”, *Journal of Ambient Intelligence and Humanized Computing*, 2019 (DOI: 10.1007/s12652-019-01419-7).
- [29] L. Ren, H. Lin, B. Xu, *et al.*, “Learning to capture contrast in sarcasm with contextual dual-view attention network”, *Int. J. Mach. Learn. and Cyber.* vol. 12, pp. 2607–2615, 2021 (DOI: 10.1007/s13042-021-01344-2).
- [30] Z.L. Chia, M. Ptaszynski, and M. Wroczynski, “Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection”, *Information Processing and Management*, vol. 58, no. 4, 2021, (DOI: 10.1016/j.ipm.2021.102600).
- [31] A.F. Hidayatullah and M.R. Ma’arif, “Pre-processing Tasks in Indonesian Twitter Messages”, *Journal of Physics: Conference Series*, vol. 801, 2017 (DOI: 10.1088/1742-6596/801/1/012072).
- [32] N. Hazim Barnouti, *et al.*, “Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance”, *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016 (DOI: 10.14569/IJACSA.2016.070550).
- [33] H. Pandey and R. Tiwari, “An Innovative Design Approach of Butterworth Filter for Noise Reduction in ECG Signal Processing based Applications”, *Progress In Science in Engineering Research Journal PISER 12*, vol. 2, pp. 332–337, 2014.
- [34] D. Kim, D. Seo, S. Cho, and P. Kang, “Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec”, *Information Sciences*, vol. 477, pp. 15–29, 2019 (DOI: 10.1016/j.ins.2018.10.006).
- [35] C. Cheng, L. Chunping, H. Yan, and Y. Zhu, “A semi-supervised deep learning image caption model based on Pseudo Label and N-gram”, *International Journal of Approximate Reasoning*, vol. 131, pp. 93–107, 2021 (DOI: 10.1016/j.ijar.2020.12.016).
- [36] D. Cristinacce and T. Cootes, “Automatic feature localisation with constrained local models”, *Pattern Recognition*, vol. 41, no. 10, pp. 3054–3067, 2008 (DOI: 10.1016/j.patcog.2008.01.024).
- [37] O.C. Ai, M. Hariharan, S. Yaacob, and L.S. Chee, “Classification of speech dysfluencies with MFCC and LPCC features”, *Expert Systems with Applications*, vol. 39, no. 2, pp. 2157–2165, 2012 (DOI: 10.1016/j.eswa.2011.07.065).
- [38] T. Kronvall, M. Juhlin, J. Swärd, S.I. Adalbjörnsson, and A. Jakobsson, “Sparse modeling of chroma features”, *Signal Processing*, vol. 130, pp. 105–117, 2017 (DOI: 10.1016/j.sigpro.2016.06.020).
- [39] M. Kavitha, R. Gayathri, K. Polat, A. Alhudhaif, and F. Alenezi, “Performance evaluation of deep e-CNN with integrated spatial-spectral features in hyperspectral image classification”, *Measurement*, vol. 191, 2022 (DOI: 10.1016/j.measurement.2022.110760).
- [40] L. An, *et al.*, “Multi-Level Canonical Correlation Analysis for Standard-Dose PET Image Estimation”, *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3303–3315, 2016 (DOI: 10.1109/TIP.2016.2567072).
- [41] X. Zhou, J. Lin, Z. Zhang, Z. Shao, and H. Liu, “Improved itracker combined with bidirectional long short-term memory for 3D gaze estimation using appearance cues”, *Neurocomputing In Press*, vol. 390, pp. 217–25, 2019 (DOI: 10.1016/j.neucom.2019.04.099).
- [42] D. Zhao, J. Wang, and Y. Zhang, “Extracting drug–drug interactions with hybrid bidirectional gated recurrent unit and graph convolutional network”, *Journal of Biomedical Informatics*, vol. 99, 2019 (DOI: 10.1016/j.jbi.2019.103295).
- [43] L. Abualigah, *et al.*, “Aquila Optimizer: A novel meta-heuristic optimization algorithm”, *Computers & Industrial Engineering*, vol. 157, 2021 (DOI: 10.1016/j.cie.2021.107250).
- [44] B.R. Rajakumar, “Impact of Static and Adaptive Mutation Techniques on Genetic Algorithm”, *International Journal of Hybrid Intelligent Systems*, vol. 10, no. 1, pp. 11–22, 2013 (DOI: 10.3233/HIS-120161).
- [45] B.R. Rajakumar, “Static and Adaptive Mutation Techniques for Genetic algorithm: A Systematic Comparative Analysis”, *International Journal of Computational Science and Engineering*, vol. 8, no. 2, pp. 180–193, 2013 (DOI: 10.1504/IJCSSE.2013.053087).
- [46] S.M. Swamy, B.R. Rajakumar, and I.R. Valarmathi, “Design of Hybrid Wind and Photovoltaic Power System using Opposition-based Genetic Algorithm with Cauchy Mutation”, *IET Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2013)*, 2013 (DOI: 10.1049/ic.2013.0361).
- [47] A. George and B.R. Rajakumar, “APOGA: An Adaptive Population Pool Size based Genetic Algorithm”, *AASRI Procedia*, vol. 4, pp. 288–296, 2013 (DOI: 10.1016/j.aasri.2013.10.043).
- [48] B.R. Rajakumar and A. George, “A New Adaptive Mutation Technique for Genetic Algorithm”, *In proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICCIIC)*, pp. 1–7, 2012, (DOI: 10.1109/ICCIIC.2012.6510293).
- [49] F. Chakraborty, P.K. Roy, and D. Nandi, “Oppositional elephant herding optimization with dynamic Cauchy mutation for multi-level image thresholding”, *Evol. Intel. 12*, pp. 445–467, 2019 (DOI: 10.1007/s12065-019-00238-1).
- [50] S.H.S. Moosavi and V.K. Bardsiri, “Poor and rich optimization algorithm: A new human-based and multi populations algorithm”, *Engineering Applications of Artificial Intelligence*, vol. 86, pp. 165–181, 2019 (DOI: 10.1016/j.engappai.2019.08.025).
- [51] F. Ahmed, “Social Spider Optimization Algorithm”, 2015 (DOI: 10.13140/RG.2.1.4314.5361).
- [52] M. Dehghani, Š. Hubálovský, and P. Trojovský, “Cat and Mouse Based Optimizer: A New Nature-Inspired Optimization Algorithm”, *Sensors*, vol. 21, no. 15, 2021 (DOI: 10.3390/s21155214).
- [53] M.O. Okwu and L.K. Tartibu, “Ant Lion Optimization Algorithm”, *Metaheuristic Optimization: Nature-Inspired Algorithms Swarm and Computational Intelligence, Theory and Applications. Studies in Computational Intelligence*, vol. 929, 2020 (DOI: 10.1007/978-3-030-61111-8_9).
- [54] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision”, *Circuits and Systems, International Symposium on*, pp. 253–256, 2010 (DOI: 10.1109/ISCAS.2010.5537907).
- [55] K. Ling-Jing and C.C. Chiu, “Application of integrated recurrent neural network with multivariate adaptive regression splines on SPC-EPC process”, *Journal of Manufacturing Systems*, vol. 57, pp. 109–118, 2020 (DOI: 10.1016/j.jmsy.2020.07.020).
- [56] Z. Masetic and A. Subasi, “Congestive heart failure detection using random forest classifier”, *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54–64, July 2016 (DOI: 10.1016/j.cmpb.2016.03.020).
- [57] P.T. Ilija, “Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size”, *Catena*, vol. 145, pp. 164–179, 2016 (DOI: 10.1016/j.catena.2016.06.004).
- [58] –, <https://github.com/soujanyaaporja/MUStARD>.



Dnyaneshwar Madhukar Bavkar is a Research Scholar under the guidance of Dr Ramgopal Kashyap in the Department of Computer Science and Engineering (CSE) at Amity University, Raipur, Chhattisgarh. He received a Bachelor of Engineering (B.E.) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad and a Master

of Engineering (M.E.) from the University of Mumbai, Maharashtra, India in Computer Science & Engineering (CSE). His research area is Machine learning, Natural Language Processing & Data Analysis.

 <https://orcid.org/0000-0003-4746-0429>

E-mail: dnyaneshwarbavkar@ternaengg.ac.in

Department of Computer Science and Engineering, Amity University, Raipur, Chhattisgarh, India



Ramgopal Kashyap has more than 15 years of teaching experience; his research area is Digital Image Processing, Pattern Recognition, and Machine Learning. He has done B.E. and M.Tech. in Computer Science with Honors. He has filed two patents and authored one international book. He has published more than 40 quality research papers in international

journals indexed in Science Citation Index (SCI) and Scopus (Elsevier). He serves as an Associate Editor and Editorial board member for more than 100 Science Citation Index, SCI-E, Scopus indexed Journals. He has also written more than 30 book chapters.

 <https://orcid.org/0000-0002-5352-1286>

E-mail: ram1kashyap@gmail.com

Department of Computer Science and Engineering, Amity University, Raipur, Chhattisgarh, India



Vaishali Khairnar is a Professor and Head of Department of Information Technology at Terna Engineering College, Navi-Mumbai. She has total 21 years of teaching experience. She is Board of Studies Member in Information Technology, University of Mumbai. She has guided many Ph.D.

and P.G. studies. Her areas of interest are wireless communication, connected vehicles, VANET, Storage etc. She has published more than 50 plus papers in Scopus journal, springer IEEE etc. She has published 3 patents. She has written and published more than five books under Wiley publication. She has completed one consultancy project and currently working on Research funded project in area of connected vehicles under Department of Science and Technology. She has received Best Research Award in 2021.

 <https://orcid.org/0000-0002-4867-1263>

E-mail: vaishalikhairnar@ternaengg.ac.in

Department of Information Technology, Terna Engineering College, Nerul, Navi Mumbai, India.

Information for Authors

Journal of Telecommunications and Information Technology (JTIT) is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review.** Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:

<https://www.il-pib.pl/pl/submission>

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@il-pib.pl

editorial.office@il-pib.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format.

Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

- [1] Y. Namiyama, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.
- [2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.
- [3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.

For copyright form see:

<https://www.il-pib.pl/pl/submission>

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the worlds most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishings databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(Contents Continued from Front Cover)

**Semantic Knowledge Management and Blockchain-based Privacy
for Internet of Things Applications**

M. Lamri and L. Sabri

Paper

75

**Multi-operator Differential Evolution with MOEA/D for
Solving Multi-objective Optimization Problems**

S. Aggarwal and K. K. Mishra

Paper

85

**Multimodal Sarcasm Detection via Hybrid Classifier with
Optimistic Logic**

D. M. Bavkar, R. Kashyap, and V. Khairnar

Paper

97



National Institute
of Telecommunications

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@il-pib.pl
editorial.office@il-pib.pl
<http://www.nit.eu>