JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2026

vol. 103

(Contents continued on back cover)

eISSN 1899-8852

Outage Probability in RIS-assisted LoRa Networks with Hardware Impairments and Asymmetric Channels

Lam-Thanh Tu[1], Bui Vu Minh[2], and Tan N. Nguyen[1]

[1]*Ton Duc Thang University, Ho Chi Minh City, Vietnam,*
[2]*Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam*

https://doi.org/10.26636/jtit.2026.1.2315

Abstract — This paper investigates the performance of reconfigurable intelligent surface (RIS)-assisted LoRa networks. Specifically, we consider a LoRa system enhanced by RIS under the influence of hardware impairments and asymmetric channel conditions. A closed-form expression for the outage probability at end devices is derived using the method of moments. The accuracy of the proposed analytical framework is extensively validated through Monte Carlo simulations. Several important insights are drawn from both the theoretical analysis and simulation results. In particular, the system's performance is significantly enhanced by an increase in the number of RIS elements and the transmission power of the gateway. Furthermore, comparisons with related works described in the literature are made to show that the proposed system outperforms these existing approaches simply by increasing the number of RIS elements. Additionally, we reveal that a higher spreading factor (SF) does not necessarily lead to worse performance than a lower SF, and the impact of hardware impairments is found to be minor under typical operating conditions.

Keywords — asymmetric channels, hardware impairments, long range, outage probability, reconfigurable intelligent surfaces

1. Introduction

LoRa networks have significantly improved the Internet of Things (IoT) by allowing long-range communication between devices while maintaining low energy consumption [1]. Using a distinctive modulation technique, LoRa ensures robust signal quality and reduced interference, supporting communication distances of up to 15 km in rural areas and 2 to 5 km in urban environments. Its low data rate makes it ideal for applications such as smart agriculture, environmental monitoring, and asset tracking.

One of LoRa's key advantages is its high energy efficiency (EE), allowing devices to operate for up to ten years on a single battery, making it suitable for large-scale IoT deployments. The architecture of LoRa networks is simple and scalable, with end devices transmitting data to gateways which then forward the information to a centralized server. This design eases the coordination of numerous devices, while the LoRa WAN protocol ensures secure and reliable communication through mechanisms such as adaptive data rate control and data encryption.

In comparison, other low-power wide area network (LPWAN) technologies such as NB-IoT [2] and Sigfox also aim to deliver energy-efficient wide area coverage, but LoRa distinguishes itself through its license-free spectrum use and flexible network deployment. As the IoT landscape continues to expand, LoRa remains a critical enabler of smart city initiatives and industrial automation, by offering a compelling blend of long-range connectivity and ultra-low power operation.

Reconfigurable intelligent surfaces (RIS) represent an emerging innovation in wireless communication, capable of actively shaping electromagnetic waves to improve signal transmission [3]. Composed of inexpensive components, RIS can manipulate signal properties such as phase, amplitude, and polarization to address issues such as signal degradation and interference. This leads to stronger and more widespread coverage without the need for additional transmitters. RIS has great potential for future networks, IoT systems, and smart city infrastructure, where connectivity is paramount.

By integrating RIS with LoRa networks, the performance of low-power IoT systems can be improved. RIS can help guide LoRa signals around physical barriers, enhancing signal integrity, and reducing data loss. This results in more reliable communication over extended distances. Since RIS operates passively, it complements LoRa's low-energy design, increasing network capacity without increasing power demands. The combination of LoRa and RIS supports more densely populated IoT environments and strengthens connectivity for applications such as precision agriculture, environmental sensing, and industrial automation.

Although LoRa networks operate at sub-GHz frequencies that inherently exhibit favorable propagation and penetration characteristics, their performance can still degrade in practical scenarios involving dense urban environments, shadowing, or deep fading. Furthermore, the low transmit power and the simple modulation scheme limit LoRa's ability to maintain reliable communication with distant or indoor-installed end devices. To overcome these limitations, integrating RIS into LoRa networks offers an efficient and cost-effective enhancement. Owing to their ability to reconfigure the wireless propagation environment by intelligently adjusting the phase of the reflected signals, RIS is capable of strengthening the

received power and mitigating the effects of non-linear sight (NLOS) conditions without introducing additional active components or power consumption [3]. Consequently, the joint use of LoRa and RIS combines the long-range, low-power benefits of LoRa with the passive beamforming and coverage-improving capabilities of RIS, resulting in more reliable connectivity, improved outage performance, and extended communication range for large-scale IoT deployments.

2. Related Work

Several problems involving LoRa networks were studied in [4]–[14]. More precisely, the work presented in [4] proposed a novel modulation scheme based on a single bit space-time block code. Based on the proposed modulation scheme, the bit error rate (BER) is significantly reduced. On the other hand, the authors of [5] minimized power consumption while satisfying the quality of service by considering hybrid systems that include both grid-based and renewable energy sources. The problem is solved by deploying the reinforcement learning approach.

The work described in [6] also relied upon deep learning to address coverage probability (Pcov) performance in a scenario in which many power beacons are used to wirelessly charge the battery of end devices (EDs), while the double training approach was proposed in [7] to estimate the energy efficiency (EE) of LoRa networks. The authors of [8] proposed an approach to allow EDs to estimate and compensate for the Doppler shift before transmitting information to low-Earth orbit (LEO) satellites.

On the other hand, the EE of LoRa networks that took into account all states and their associated power consumption levels of the EDs were investigated in [10]. The slotted ALOHA access protocol and carrier-sense multiple access protocol (CSMA) were proposed and studied in [11]. The authors showed that the slotted ALOHA protocol outperformed its pure ALOHA and CSMA counterparts.

The Pcov of the LoRa networks enabled by energy harvesting (EH) was derived in [12] based on nonlinear EH modelling. The authors in [13] derived the Pcov of LoRa networks employing either antenna diversity (using multiple antennas at the gateway) or time diversity. They demonstrated that utilizing diversity techniques is beneficial for system performance. However, their analysis focused solely on uplink LoRa networks and assumed the absence of hardware impairments at both the transmitter and receiver. In practice, this assumption is difficult to satisfy, especially when many antennas are deployed.

On the other hand, the authors in [14] investigated the performance of LoRa dual hop uplink relay networks. Their results showed that shortening the transmission distance can improve Pcov. However, this work did not address how synchronization is achieved among the gateway, EDs, and relays. Furthermore, the impact of asymmetric channels was not considered.

Regarding RIS-assisted wireless systems, the authors of [15] simultaneously optimized the non-orthogonal multiple access (NOMA) transmit power of a LEO satellite and the passive beamforming of RIS, assuming imperfect successive interference cancelation for the NOMA-satellite communications. The work presented in [16] derived the outage probability (OP) of NOMA systems with asymmetric channels comprising $\eta - \mu$ and Rayleigh fading channels. In [17] the latency to perform RIS communications systems aided by tasks of all users in the mobile edge computing (UAV). A combination of localization and communications in integrated sensing and communication (ISAC) systems assisted by RIS was proposed in [18]. The authors minimized the squared position error bound taking into consideration the transmit beamforming vector, phase shift of the RIS elements, and subcarrier assignments.

The trade-off between the security and reliability of RIS-assisted multi-hop cooperative communications with fountain codes was investigated in [19]. The derivations of OP, ergodic capacity, and EE of the cognitive radio NOMA network were provided in [20]. On the other hand, the performance of the two-way RIS-aided systems was studied in [21], under both optimal and random phase shift designs.

In [22], the authors investigated the impact of hardware impairment (HI) on the spectral efficiency (SE) of massive multiple-input multiple-output (MIMO) over Rician fading distributions. Meanwhile, paper [23] examined the performance of cognitive NOMA systems under hardware impairments and two user selection strategies, namely path-loss-based and channel gain-based selection. Their findings show that neither scheme consistently outperforms the other in all operating conditions. This suggests that an adaptive user selection mechanism, capable of switching between strategies based on network state, would be more effective in optimizing overall system performance.

The work described in [24] studied the location of RIS-aided systems under the influence of HI, while the performance analysis of rateless codes-based communication with transmit antenna selection, the harvest-to-jam technique, and hardware impairments was investigated in [25]. A similar work is presented in [26], where the authors analyze a LoRa system employing a single RIS to enhance the gateway-to-end-device link under favorable line of sight conditions. Their findings show that properly configured RIS phase shifts can substantially extend LoRa coverage.

This work advances the analysis in several important aspects: we explicitly incorporate transmitter and receiver hardware impairments, account for phase noise at the RIS, consider asymmetric fading conditions (Rician for the gateway-RIS link and Rayleigh for both the RIS-ED and direct links), and derive closed-form OP expressions under these practical impairments. While [26] assumes ideal RIS control and perfect hardware, our model includes both finite phase-shift resolution and RIS hardware imperfections. Consequently, the proposed framework significantly broadens the scope of [26] and offers a performance analysis that better reflects real-world IoT/LPWAN deployments.

The main contributions and novelties of the present work are primarily summarized as follows:

- Asymmetric systems in RIS-aided LoRa networks are considered, where the fading subject to the first-hop gateway to RIS is different from the second hop. Furthermore, we also consider the direct link from the gateway to the emergency department and study the impact of hardware impairments at both the transmitter and the receiver.
- A closed-form expression for OP is derived using the moment method (MM), which remains accurate for an arbitrary number of RIS elements, unlike approaches based on the central limit theorem (CLT) which typically require many RIS elements to be valid.
- We highlight several key insights from the simulation results. Specifically, we observe that an increase in the number of RIS elements and the transmit power of the gateway significantly improves OP. Additionally, we find that a larger SF does not necessarily lead to worse OP performance and that this relationship is more nuanced. Finally, the impact of hardware impairments on the overall performance of the system in LoRa networks that are considered RIS-assisted is shown to be minor.

The structure of the paper is organized as follows: Section 3 describes the system model, while its performance is analyzed in Section 4. Section 5 presents numerical calculations and discussions. Finally, Section 6 concludes the article. Table 1 summarizes the abbreviations used.

Tab. 1. List of acronyms used throughout the paper.

Acronym	Definition
AN	Artificial noise
AWGN	Additive white Gaussian noise
CLT	Central limit theorem
CSI	Channel state information
EE	Energy efficiency
ED	End device
HI	Hardware impairment
IoT	Internet of Things
LPWAN	Low power wide area network
LoRa	Long range
LOS / NLOS	Line-of-sight / non-line-of-sight
MM	Moment method
NOMA	Non-orthogonal multiple access
OP	Outage probability
RIS	Reconfigurable intelligent surface
SF	Spreading factor
SNR	Signal-to-noise ratio
STAR-RIS	Simultaneous transmitting and reflecting RIS
Tx / Rx	Transmitter / receiver
UAV	Unmanned aerial vehicle

3. System Model

Let us consider a downlink of single gateway LoRa networks as shown in Fig. 1. The gateway is located at the center of the disk with radius P, while EDs are randomly distributed around the gateway. The considered networks also comprise a RIS with K elements, in order to help the gateway send downlink information to the EDs. We assume that all nodes in the networks are equipped with a single antenna. The entire transmission takes place in one time slot.

3.1. Channel Modeling

All transmission links in the systems are subjected to both small-scale fading and large-scale path loss and are modeled in the sequel.

Small-scale fading is modeled by Rician and Rayleigh distributions. In particular, the channel coefficient from the gateway to the RIS will be modeled as a Rician distribution, and the link from the RIS to the EDs will follow a Rayleigh distribution. On the other hand, the direct link from the gateway to the ED always follows the Rayleigh distribution.

Path-loss is modeled based on the popular unbounded path-loss model and is formulated as [27]:

$$O_{x,y} = O_0 \, d_{x,y}^{\chi_a} \, , \tag{1}$$

where $d_{x,y}$ denotes the transmission distance from the transmitter to the receiver and χ_a with $a \in \{L, N\}$ represents the path-loss exponent.

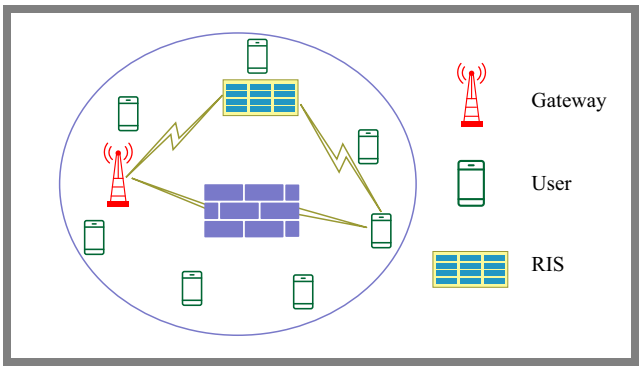

Fig. 1. RIS-aided downlink LoRa networks under consideration.

In this work, we consider different path loss exponents for different links. Specifically, the links from the gateway to the ED and from the RIS to the EDs are subject to Rayleigh fading and therefore experience a higher path loss exponent, denoted by χ_N. On the contrary, the link from the gateway to the RIS follows a Rician distribution and enjoys a lower path loss exponent χ_L. The term O_0 is the path loss constant measured at a reference distance of 1 m and depends on the carrier frequency given by:

$$O_0 = \left(\frac{4\pi f_c}{c}\right)^2 \, ,$$

where f_c is the carrier frequency and c is the speed of light.

3.2. Characteristics of Rayleigh and Rician Distributions

Since Rayleigh fading is a special case of the Rician distribution, we focus on the Rician case and highlight the Rayleigh case as a specific instance. Let C denote the ratio of the power in the LOS component to the power in all NLOS components, and let L represent the total power, including both LOS and NLOS paths. The mean and variance of a random variable (RV) X following the Rician distribution are calculated as follows [28]:

$$\begin{aligned} \mathbb{E}\{X\} &= \sqrt{\frac{\pi L}{4(C+1)}} e^{\frac{-C}{2}} \left[(1+C) I_0 \frac{C}{2} + C I_1 \frac{C}{2}\right], \\ \mathrm{Var}\{X\} &= L - \mu_X^2 \end{aligned} \quad (2)$$

where $I_n(\cdot)$ denotes the modified Bessel function of the first kind of order n. It is obvious that when $C = 0$, the distribution reduces to the Rayleigh distribution. The operators $\mathbb{E}\{\cdot\}$ and $\mathrm{Var}\{\cdot\}$ represent the expectation and the variance, respectively.

3.3. Spreading Factor Allocation

In this paper, the conventional distance-based SF assignment scheme is adopted to assign SFs to EDs. Specifically, an ED is assigned the o-th SF, $o \in \{7, \ldots, 12\}$, if its transmission distance to the gateway falls within the following range:

$$\left[\frac{(o-7)P}{6}, \frac{(o-6)P}{6}\right) .$$

3.4. Asymmetric Channels

In many wireless communication systems, especially in multi-hop scenarios, assuming a uniform fading distribution across all hops often oversimplifies the complex nature of real-world propagation environments. A more realistic and insightful approach is to consider scenarios where each hop experiences a distinct fading distribution.

This is motivated by the fact that the physical path and surrounding environment of each link can vary significantly. For example, the first hop may benefit from a clear LOS path in an open suburban area, resulting in Rician fading, while a subsequent hop might traverse a densely cluttered urban environment, leading to severe Rayleigh fading.

Accordingly, this paper adopts an asymmetric fading model. Wireless links in the LoRa network are modeled using a combination of Rician and Rayleigh fading. Specifically, the gateway-to-RIS link follows a Rician distribution to represent a scenario with a strong LOS component, as the RIS is typically placed at elevated or strategically selected positions with clear visibility of the gateway.

On the contrary, both the RIS-to-ED and the direct gateway-to-ED links experience Rayleigh fading, representing NLOS conditions caused by scattering and shadowing in practical IoT deployments [29]. It is worth emphasizing that the proposed analytical framework is general and can be directly applied to the reverse scenario (i.e., Rician fading for the RIS-to-ED link and Rayleigh fading for the gateway-to-RIS link) without modification.

3.5. Hardware Impairments

Although much of the existing research assumes ideal conditions for transmitter and receiver hardware, this study adopts a more practical perspective by considering the inevitable effects of hardware imperfections at both ends of the communication system. This consideration is especially important for setups involving affordable technologies such as LoRa, where flawless hardware is an unrealistic assumption.

More precisely, both the gateway and EDs are assumed to employ low-cost transceiver hardware, consistent with LoRa technology. The residual distortions introduced by non-ideal mixers, amplifiers, and oscillators are modeled as additive Gaussian hardware impairment noise [30], denoted as Ξ_X and Ξ_Y , parameterized by zero mean as well as variances $\delta^2_{\Xi_G}$ and $\delta^2_{\Xi_o}$. This modeling approach is widely used and accurately reflects low-cost IoT transceivers [31].

3.6. RIS Phase Modeling

The proposed model also accounts for non-idealities introduced by the RIS itself. Unlike many existing studies that assume perfectly tunable, continuous phase adjustments, we adopt a more practical representation in which each RIS element operates with quantized phase control.

Let $\mathcal{Q}$ be the number of quantization bits assigned to each element. Under this assumption, the resulting phase perturbation θ is modeled as a uniformly distributed random variable throughout the interval:

$$\left[-2^{-\mathcal{Q}}\pi,\, 2^{-\mathcal{Q}}\pi\right] ,$$

reflecting the inherent discretization error.

It is worth noting that the adopted phase-shift model is well supported by recent studies on programmable metasurfaces, showing that even low-cost RIS implementations with only a few quantization bits can achieve near-continuous phase control at sub-GHz and microwave frequencies [32], [33]. Furthermore, as illustrated in Tab. 2, RIS configurations employing only 2–3 bits of phase resolution can maintain more than 90% of the performance of an ideal continuous phase RIS, particularly in the low-frequency bands characteristic of LoRa systems.

These observations indicate that the aggregate reflection gain of the RIS effectively compensates for quantization-induced errors, thereby confirming that the adopted phase-shift model is both practically realistic and analytically tractable for modern metasurface hardware.

3.7. Signal-to-Noise Ratios at EDs

The received signal of a typical ED utilizing o-th SF is formulated as [34]:

$$\begin{aligned} y_o =& \sqrt{B_G}\left[\sqrt{O_{G,o}^{-1}} h_{G,o} + \sqrt{O_{G,R}^{-1} O_{R,o}^{-1}} \sum_{k=1}^{K} h_{G,k} \Phi_k h_{k,o}\right] \\ & \times (t_G + \Xi_G) + \Xi_o + n_o , \end{aligned} \quad (3)$$

where BG denotes the transmit power of the gateway, $O_{G,o}$, $O_{R,G}$, and $O_{R,o}$ represent the large-scale path-loss from the gateway to the ED using the o-th SF, from the gateway to the RIS, and from the RIS to the ED with o-th SF, respectively.

The t_G is the transmit signal of the gateway, while Ξ_G and Ξ_o denote the HI at the gateway and the ED, respectively. n_o is the additive white Gaussian noise (AWGN) at the receiver. The channel coefficients are represented by $h_{G,o}$ (gateway to the ED using SF o), $h_{G,k}$ (gateway to the k-th element of the RIS), and $h_{k,o}$ (from the k-th RIS element to the ED using SF o).

The phase-shift matrix of the RIS is given by $\Phi = \text{diag}\left(e^{j\iota_1}, \ldots, e^{j\iota_K}\right)$, where $\iota_k \in [-\pi, \pi]$ is the phase shift of the k-th RIS element, and $\text{diag}(\cdot)$ denotes a diagonal matrix.

From Eq. (3), the SINR of the ED using the o-th SF, denoted by ω_o, is given by Eq. (4):

$$\omega_o = \frac{\left|\sqrt{B_G}\left[\sqrt{O_{G,o}^{-1}}h_{G,o} + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\sum_{k=1}^{K} h_{G,k}\Phi_k h_{k,o}\right]\right|^2}{\left|\sqrt{B_G}\left[\sqrt{O_{G,o}^{-1}}h_{G,o} + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\sum_{k=1}^{K} h_{G,k}\Phi_k h_{k,o}\right]\right|^2 \left(\delta_{\Xi_G}^2 + \delta_{\Xi_o}^2\right) + \sigma_o^2} = \frac{B_G Z^2}{B_G\left(\delta_{\Xi_G}^2 + \delta_{\Xi_o}^2\right) Z^2 + \sigma_o^2}\,. \tag{4}$$

The composite signal term Z is defined as:

$$Z = \left|\sqrt{O_{G,o}^{-1}}\, h_{G,o} + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\sum_{k=1}^{K} h_{G,k}\Phi_k h_{k,o}\right|\,.$$

$$Z = \sqrt{O_{G,o}^{-1}}\,|h_{G,o}| + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\sum_{k=1}^{K}|h_{G,k}|\,|h_{k,o}|\,e^{j\theta_k}\,,$$

where $e^{j\theta_k}$ is the phase error of the k-th element, σ_o^2 is the variance of the AWGN in the ED using o-th SF.

3.8. Performance Metrics

In this paper, OP is considered the main metric to measure the performance of the considered networks. OP of the o-th SF is defined as:

$$\text{OP}_o = \Pr\{\omega_o < \beta_o\} \tag{5}$$

where $\Pr\{\cdot\}$ is the probability operator, β_o denotes the target SINR threshold for the o-th SF and is treated as a constant. The values are given by $\beta = \{-6, -9, -12, -15, -17, -20\}$ dBm, corresponding to SF7 through SF12, respectively. For example, $\beta_7 = -6$ dBm for SF7, and $\beta_{12} = -20$ dBm for SF12.

4. Performance Analysis

The OP of ED used o-th SF is computed as follows:

$$\begin{aligned}\text{OP}_o =& F_Z\sqrt{\frac{\beta_o}{\Psi\left(1-\beta_o\delta^2\right)}}\,U\left(1-\beta_o\delta^2\right)\\ &+ U\left(\beta_o\delta^2 - 1\right)\,,\end{aligned} \tag{6}$$

where $U(\cdot)$ denotes the unit step function, $\delta^2 = \delta_{\Xi_G}^2 + \delta_{\Xi_o}^2$ represents the total HI variance from both the gateway and the ED, $\Psi = \frac{B_G}{\sigma_o^2}$ denotes the average transmit power-to-noise ratio. Additionally, $F_X(x)$ is the cumulative distribution function (CDF) of the random variable X.

Proof. Let us begin the proof by rewriting the definition of the OP in Eq. (5) as follows:

$$\begin{aligned}\text{OP}_o &= \Pr\left\{\omega_o = \frac{B_G Z^2}{B_G\left(\delta_{\Xi_G}^2 + \delta_{\Xi_o}^2\right) Z^2 + \sigma_o^2} < \beta_o\right\}\\ &\overset{(a)}{=} \Pr\left\{Z^2 < \frac{\beta_o}{\Psi\left(1-\beta_o\delta^2\right)}\right\} U\left(1-\beta_o\delta^2\right) + U\left(\beta_o\delta^2 - 1\right)\,.\end{aligned} \tag{7}$$

Here (a) is held by applying the definition of the CDF of a RV and we close the proof. □

By directly inspecting Eq. (6), we observe that if $\beta_o\delta^2 > 1$, the system will always experience an outage. Further analysis reveals that if the negative impact of HI, whether at the transmitter or receiver, is too severe (i.e., when δ^2 is large), the system remains in outage regardless of other system parameters. This leads to the first and most important insight – hardware impairments must be minimized as much as possible to ensure system performance.

Equation (6) also reveals that increasing Ψ is always beneficial for reducing OP, provided that HI remain within acceptable limits-specifically, the condition $\beta_o\delta^2 < 1$ must be satisfied. This indicates that enhanced system performance can be achieved by increasing the transmit power of the gateway or by reducing the noise power, for example, through bandwidth reduction.

Equation (6) also indicates that the closed-form expression of the OP in both considered cases reduces to the CDF of the RV Z.

In the following, we derive the CDF of:

$$Z = \sqrt{O_{G,o}^{-1}}\,|h_{G,o}| + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\sum_{k=1}^{K}|h_{G,k}|\,|h_{k,o}|\,e^{j\theta_k}\,,$$

by applying the method of moments. It should be emphasized that a closed-form expression for the distribution of Z is not available, even under simple fading conditions, such as Rayleigh fading, due to the complexity introduced by the sum of products of multiple random variables. Consequently, prior work in the literature often resorted to using the central

limit theorem to approximate Z as a Gaussian random variable. However, this approach requires a large number of RIS elements to achieve an accurate approximation.

In contrast, the method of moments offers a more accurate and flexible alternative, as it does not rely on the assumption of a large number of RIS elements. Therefore, in this work, we adopt the MM approach and the resulting expression is provided in Theorem 1.

Theorem 1. Let us denote $\widehat{Z}$ is the equivalent RV of Z which follows a gamma distribution with shape and scale parameters denoted as λ and ξ as given below:

$$\lambda = \frac{\left(\mathbb{E}\left\{Z\right\}\right)^2}{\mathrm{Var}\left\{Z\right\}}, \quad \xi = \frac{\mathrm{Var}\left\{Z\right\}}{\mathbb{E}\left\{Z\right\}}, \tag{8}$$

The mean and variance of Z are computed as follows:

$$\begin{aligned}
\mathbb{E}\left\{Z\right\} &= \sqrt{O_{G,o}^{-1}}\mathbb{E}\left\{|h_{G,o}|\right\} \\
&\quad + K\sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\mathbb{E}\left\{|h_{G,R}|\right\} \\
&\quad \times \mathbb{E}\left\{|h_{R,o}|\right\}\mathbb{E}\left\{e^{j\theta_k}\right\} \\
\mathrm{Var}\left\{Z\right\} &= O_{G,o}^{-1}\mathrm{Var}\left\{|h_{G,o}|\right\} \\
&\quad + KO_{G,R}^{-1}O_{R,o}^{-1}\mathrm{Var}\left\{\mathcal{A}e^{j\theta_k}\right\} \\
\mathrm{Var}\{\mathcal{A} = |\mathrm{h_{G,R}}||h_{R,o}|\} &= (\mathbb{E}\{|h_{G,R}|\})^2\mathrm{Var}\{|\mathrm{h_{R,o}}|\} \\
&\quad + \left(\mathbb{E}\left\{|h_{R,o}|\right\}\right)^2\mathrm{Var}\left\{|h_{G,R}|\right\} \\
&\quad + \mathrm{Var}\left\{|h_{G,R}|\right\}\mathrm{Var}\left\{|h_{R,o}|\right\} \\
\mathrm{Var}\left\{\mathcal{A}e^{j\theta_k}\right\} &= (\mathbb{E}\{\mathcal{A}\})^2\mathrm{Var}\left\{e^{j\theta_k}\right\} \\
&\quad + \mathbb{E}\left(\left\{e^{j\theta_k}\right\}\right)^2\mathrm{Var}\left\{\mathcal{A}\right\} \\
&\quad + \mathrm{Var}\left\{\mathcal{A}\right\}\mathrm{Var}\left\{e^{j\theta_k}\right\}.
\end{aligned} \tag{9}$$

Proof. The proof of (9) is given as follows:

$$\begin{aligned}
&\mathbb{E}\left\{Z\right\} = \sqrt{O_{G,o}^{-1}}\mathbb{E}\left\{|h_{G,o}|\right\} + \sqrt{O_{G,R}^{-1}O_{R,o}^{-1}} \\
&\times \mathbb{E}\left\{\sum_{k=1}^{K}|h_{G,R}|\,|h_{R,o}|\,e^{j\theta_k}\right\} \\
&\overset{(a)}{=} \sqrt{O_{G,o}^{-1}}\mathbb{E}\left\{|h_{G,o}|\right\} \\
&+ K\sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\mathbb{E}\left\{|h_{G,R}|\,|h_{R,o}|e^{j\theta_k}\right\} \\
&\overset{(b)}{=} \sqrt{O_{G,o}^{-1}}\mathbb{E}\left\{|h_{G,o}|\right\} + K\sqrt{O_{G,R}^{-1}O_{R,o}^{-1}}\mathbb{E}\left\{|h_{G,R}|\right\} \\
&\times \mathbb{E}\left\{|h_{R,o}|\right\}\mathbb{E}\left\{e^{j\theta_k}\right\}, \\
&\mathrm{Var}\left\{Z\right\} = O_{G,o}^{-1}\mathrm{Var}\left\{|h_{G,o}|\right\} + O_{G,R}^{-1}O_{R,o}^{-1} \\
&\times \mathrm{Var}\left\{\sum_{k=1}^{K}|h_{G,k}|\,|h_{k,o}|\,e^{j\theta_k}\right\} \\
&\overset{(c)}{=} O_{G,o}^{-1}\mathrm{Var}\left\{|h_{G,o}|\right\} + KO_{G,R}^{-1}O_{R,o}^{-1}\mathrm{Var}\left\{\mathcal{A}e^{j\theta_k}\right\}.
\end{aligned} \tag{10}$$

where (a), (b) and (c) are attained owing to the independence between direct link and indirect links. Between the first and second hops of the indirect link, we finish the proof.

Tab. 2. Simulation parameters used in the performance evaluation.

Parameter	Symbol/value	Description
Carrier frequency	$f_c = 433$ MHz	LoRa operating frequency
Bandwidth	BW $= 125$ kHz	Transmission bandwidth
Transmit power	$B_G = 4$ dBm	Gateway transmit power
Network radius	$P = 2 - 6$ km	LoRa coverage area
RIS elements	$K = 24 - 80$	Number of reflecting units
Rician factor	$C_{G,R} = 0 - 10$	Gateway to RIS channel condition
Path-loss exponent	$\chi_{\mathrm{LOS}} = 2.2$, $\chi_{\mathrm{NLOS}} = 3.8$	Large-scale fading
HI variance	$\delta_{\Xi_G} = \delta_{\Xi_o} \in [0, 0.5]$	Hardware impairment level
Noise variance	σ_o^2	AWGN at ED
Quantization bits	$\mathcal{Q} = 6$ bits	RIS phase resolution

The mean and variance of the phase error can be found in [35], where the l-th moment is given by $\mathbb{E}\left\{e^{jl\theta_k}\right\} = \mathrm{sinc}\left(2^{-\mathcal{Q}+l-1}\right),\ \forall k.$ □

Finally from Theorem 1 and Eq. (2), we obtain the closed-form expression of the OP:

$$\begin{aligned}
\mathrm{OP}_o = &\frac{\gamma\left(\lambda, \frac{1}{\xi}\sqrt{\frac{\beta_o}{\Psi\left(1-\beta_o\delta^2\right)}}\right)}{\Gamma\left(\lambda\right)}U\left(1-\beta_o\delta^2\right) \\
&+ U\left(\beta_o\delta^2 - 1\right).
\end{aligned} \tag{11}$$

Here $\Gamma(\cdot)$ and $\gamma(\cdot)$ represent Gamma and lower incomplete Gamma function.

5. Numerical Results

In this section, simulation results based on the Monte Carlo method are presented to validate the accuracy of the derived mathematical framework. Without loss of generality, set of simulation parameters are provided in Tab. 2. Here, the system parameters, i.e. carrier frequency $f_c = 433$ MHz, bandwidth BW $= 125$ kHz, and network radius $P = 2 - 6$ km, follow the standard LoRa physical layer settings recommended by the LoRa Alliance. The Rician factor $C_{G,R}$ is varied from 0 to 10 to capture both weak and strong LOS conditions, while the HI variance $\delta_{\Xi} \in [0, 0.5]$ is chosen to cover practical non-idealities observed in commercial IoT devices.

Figure 2 illustrates the OP performance versus the number of quantization bits $\mathcal{Q}$ per RIS element. The results confirm an excellent agreement between the derived analytical framework and the Monte Carlo simulations, validating the accuracy of

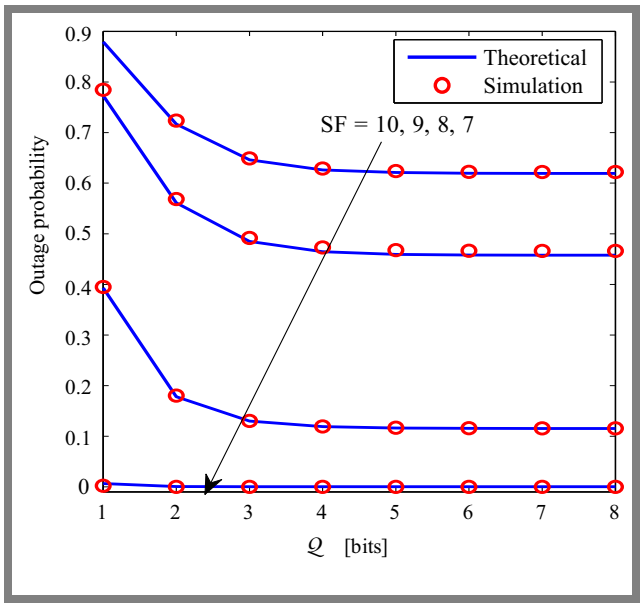

Fig. 2. OP vs. the number of quantization bits $\mathcal{Q}$. Lines are plotted based on Eq. (11), while markers originate from the Monte Carlo simulation.

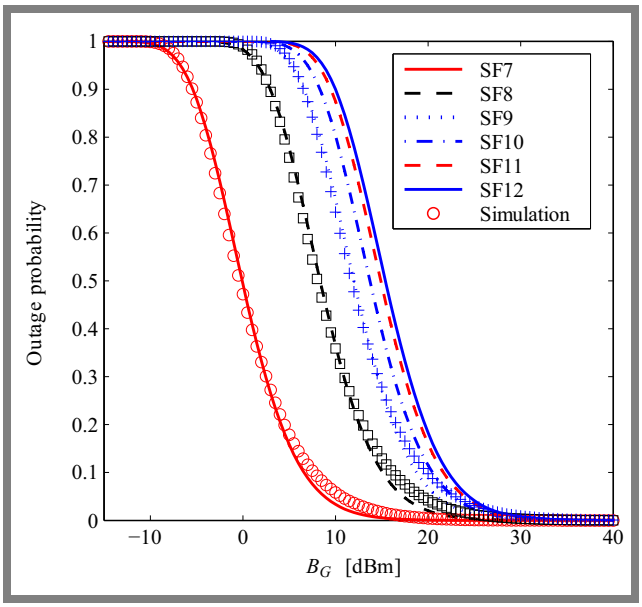

Fig. 3. OP vs. transmit power of gateway B_G. Lines are plotted based on Eq. (11), while markers originate from the Monte Carlo simulation.

the proposed model. It can be observed that increasing the phase resolution improves the OP performance. In particular, the OP decreases significantly when $\mathcal{Q}$ increases from 1 to 2 bits, and then gradually approaches a saturation point for $\mathcal{Q} \geqslant 4$. This indicates that further increasing the RIS phase resolution beyond four bits yields negligible performance improvement. Therefore, without loss of generality, a resolution of $\mathcal{Q} = 6$ bits is adopted in the subsequent analysis.

Figure 3 illustrates the outage performance as a function of the gateway transmit power B_G. It can be observed that the OP decreases monotonically with increasing B_G. Furthermore, smaller SFs generally result in better outage performance, except for SF12, which outperforms SF11. This is because, although SF12 suffers from longer transmission durations, it requires a lower target data rate, leading to a reduced outage probability. Nonetheless, the performance gap among the higher SFs (e.g., SF11 and SF12) is relatively small compared to the gap between these and SF7. The figure also validates the accuracy of the proposed MM-based analytical framework, as the curves generated from the analysis closely match those obtained from Monte Carlo simulations.

The impact of the HI level at both the gateway and EDs on the OP is illustrated in Fig. 4. It is observed that an increase in the HI level δ_Ξ leads to a degradation in OP performance. However, this adverse effect is relatively minor. For instance, with SF8, increasing δ_Ξ from 0 to 0.5 results in a rise in OP from 0.1 to 0.16, which is the largest increase observed among all SFs. Once again, we observe a strong agreement between the simulation results and the proposed analytical framework.

Figure 5 illustrates the OP performance with respect to the number of RIS elements N. It is evident that increasing N significantly enhances the OP performance. Specifically, under the current setup, when $N = 60$, the OP for all SFs approaches 10^{-5} level, and for SF7, this level is achieved even with $N < 30$.

This figure also compares the performance of the proposed network with existing works in the literature. In particular, we compare it with the dual-hop relay-assisted LoRa network in [14] (represented by red curves) and the multiple-antenna gateway scheme in [13] (represented by green – 2 antennas and black curves – 10 antennas). The results show that the proposed framework outperforms these existing approaches by simply increasing the number of RIS elements. It is emphasized that these comparisons are conducted under fair settings. All schemes operate with the same transmit power, experience the same hardware impairments, and are subject to asymmetric channels with identical channel coefficients.

For example, with SF7, the dual hop scheme achieves an OP of 4.2×10^{-6}, while the OPs for the multiple antenna gateway scheme are 0.05 for two antennas and 8×10^{-7} for

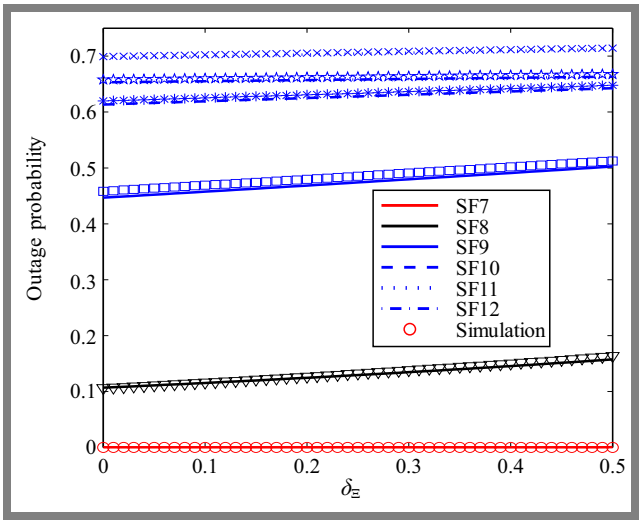

Fig. 4. OP vs. HI level at both the transmitter and receiver $\delta_{\Xi_G} = \delta_{\Xi_D} = \delta_\Xi$. Lines are plotted according to Eq. (11) while markers originate from the Monte Carlo simulation.

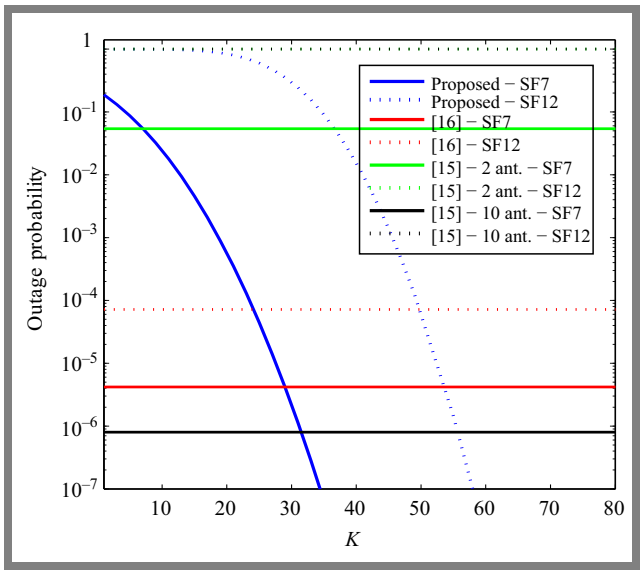

Fig. 5. OP vs. number of RIS elements K. Blue curves are plotted on Eq. (11), red curves originate from [14], while green and black curves originate from [13].

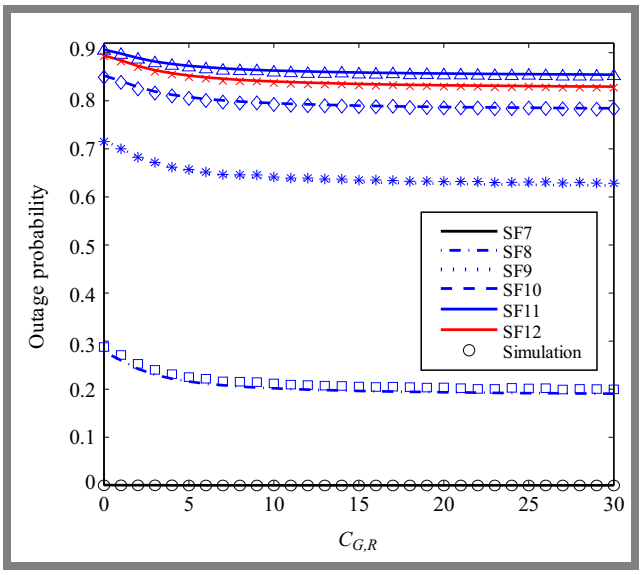

Fig. 6. OP vs. Rician factor of the link from the gateway to the RIS, $C_{G,R}$. Lines are plotted according to Eq. (11), while markers originate from the Monte Carlo simulation.

ten antennas. To match the performance of the 10-antenna configuration, the proposed scheme requires only 31 RIS elements. For SF12, an interesting observation arises: OP performance of both the 2- and 10-antenna configurations in the multiple-antenna gateway scheme is approximately 0.998, indicating poor reliability.

In contrast, the dual-hop scheme achieves an OP of 7.18×10^{-5}, while the proposed RIS-assisted scheme requires 50 elements to achieve comparable performance. In general, these comparisons demonstrate that the proposed scheme not only outperforms existing works described in the literature, but also effectively mitigates the adverse impact of hardware impairments by simply increasing the number of RIS elements.

Figure 6 presents the impact of the Rician factor of the link from the gateway to the RIS, denoted as $C_{G,D}$, on the OP. It is observed that increasing $C_{G,D}$ improves the OP performance. However, this improvement saturates across all SFs. Specifically, the OP decreases rapidly as $C_{G,D}$ increases from 0 to 5, but beyond this point, the rate of improvement diminishes and the OP becomes nearly constant.

The impact of the Rician factor on all links is investigated in Fig. 7, where $C_{G,R} = C_{G,D} = C_{R,D} = C$, indicating a symmetric system setup. Overall, the OP decreases as C increases, which aligns with practical expectations-systems perform better under more favorable channel conditions (i.e., stronger LOS components). The most significant improvement is observed for SF9, where the OP drops from 0.7 to approximately 0.15. Interestingly, the OP for SF12 increases slightly with higher C, which suggests that in this scenario, the lower target rate of SF12 may be offset by other limiting factors as the channel improves.

The OP performance with respect to the network radius P is investigated in Fig. 8. As expected, increasing the network radius degrades system performance. Under the current setup, the OP approaches one for all SFs when $P > 5$ km. To mitigate this performance degradation, one could consider increasing the number of RIS elements and/or the transmit power of the gateway.

It is worth noting that the present work focuses on theoretical modeling and analytical performance evaluation of RIS-assisted LoRa networks under hardware impairments and asymmetric fading conditions.

The objective is to establish a fundamental understanding of the outage behavior and to provide closed-form expressions that can serve as benchmarks for future experimental verification. Although hardware validation and field trials are not included in this study, the analytical and simulation results have been carefully cross-verified and are consistent with re-

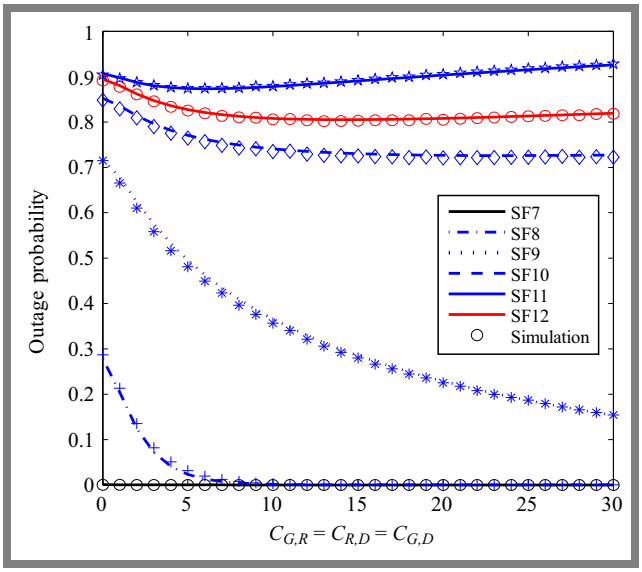

Fig. 7. OP vs. Rician factor of all links, $C_{G,R} = C_{G,D} = C_{R,D}$. Lines are plotted according to Eq. (11), while markers originate from the Monte Carlo simulation.

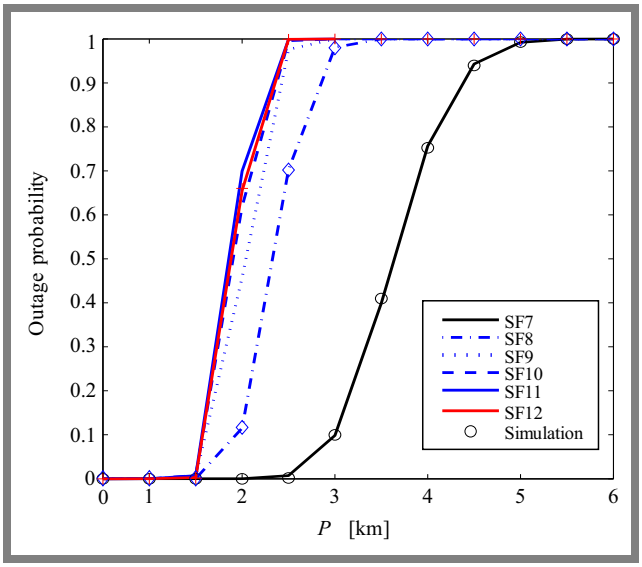

Fig. 8. OP vs. network radius, P. Lines are plotted according to Eq. (11), while markers originate from the Monte Carlo simulation.

cent measurement-based studies of LoRa and RIS systems reported in the literature [32], [33].

Furthermore, parameter values (e.g., carrier frequency, bandwidth, and Rician factors) were selected according to the practical LoRa system specifications and verified through Monte Carlo simulations for accuracy. In future work, we plan to implement a small-scale RIS-aided LoRa prototype using a programmable metasurface and commercial LoRa transceivers (e.g., SX1276) to experimentally validate the analytical framework and further quantify the impact of practical hardware imperfections.

6. Conclusions

This paper investigates the performance of LoRa networks under the impact of hardware impairments. A closed-form expression for the OP is derived using the method of moments. Numerical results reveal several key insights: the OP decreases monotonically with an increase in gateway transmit power, the number of RIS elements, and the Rician factor of the gateway-to-RIS link, among others.

It was emphasized that the analytical framework developed in this paper provides important insights into the performance of RIS-assisted LoRa networks under hardware impairments, phase noise, and asymmetric fading conditions, and several limitations should be acknowledged. The current system model considers a single RIS and a single gateway operating in the downlink direction, assuming quasi-static channels and perfect synchronization. In practical deployments, time-varying channels, interference from multiple gateways, and imperfect synchronization may introduce additional performance degradation. The RIS is modeled as a planar reflective surface with ideal or quantized phase control. Extensions to other reconfigurable architectures such as STAR-RIS or hybrid active-passive RIS could capture more realistic propagation scenarios. The proposed framework focuses on the probability of an outage as the primary metric. Incorporating additional metrics such as energy efficiency, latency, or spectral efficiency would provide a more comprehensive evaluation of the performance of the system.

Furthermore, the adopted mathematical framework is general enough to be extended beyond LoRa networks. By adapting the physical-layer parameters and channel characteristics, the same analytical methodology can be applied to other LPWAN technologies such as NB-IoT and Sigfox, or to short-range IoT protocols such as IEEE 802.15.4. This model can also be integrated with stochastic geometry to evaluate large-scale random deployments or with rate-adaptive mechanisms to capture dynamic data-rate selection in heterogeneous IoT environments. These extensions would further enhance the applicability of the proposed framework and bridge the gap between theoretical modeling and practical IoT network design.

Acknowledgments

This research is funded by Ton Duc Thang University under grant number FOSTECT.2024.21.

References

[1] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview", *IEEE Communications Surveys & Tutorials*, vol. 19, pp. 85–873, 2017 (https://doi.org/10.1109/COMST.2017.2652320).

[2] B. Doori and A. Zurfi, "Decreasing the RA Collision Impact for Massive NB-IoT in 5G Wireless Networks", *Jordanian Journal of Computers and Information Technology*, vol. 7, pp. 268–277, 2021 (https://doi.org/10.5455/jjcit.71-1620292048).

[3] A. Reda, T. Mekkawy, and A. Mahran, "Towards Optimizing the Downlink Transmit Power in UAV-integrated IRS Wireless Systems", *Jordanian Journal of Computers and Information Technology*, vol. 11, pp. 184–196, 2025 (https://doi.org/10.5455/jjcit.71-1727208249).

[4] H. Zhang *et al.*, "1-bit STBC-LoRa Modulation for Extreme Environment Communications", *IEEE Wireless Communications Letters*, vol. 14, pp. 2094–2098, 2025 (https://doi.org/10.1109/LWC.2025.3563205).

[5] R. Hamdi *et al.*, "LoRa-RL: Deep Reinforcement Learning for Resource Management in Hybrid Energy LoRa Wireless Networks", *IEEE Internet of Things Journal*, vol. 9, pp. 6458–6476, 2022 (https://doi.org/10.1109/JIOT.2021.3110996).

[6] T.T.H. Nguyen *et al.*, "Coverage Probability of EH-enabled LoRa Networks – A Deep Learning Approach", *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 12, art. no. 6780, 2024 (https://doi.org/10.4108/eetinis.v12i2.6780).

[7] L.T. Tu *et al.*, "Energy Efficiency Optimization in LoRa Networks – A Deep Learning Approach", *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 15435–15447, 2022 (https://doi.org/10.1109/TITS.2022.3183073).

[8] M.A. Ullah *et al.*, "Extending the LoRa Direct-to-satellite Limits: Doppler Shift Pre-compensation", *IEEE Open Journal of the Communications Society*, vol. 6, pp. 2256–2273, 2025 (https://doi.org/10.1109/OJCOMS.2025.3554077).

[9] H.R. Barua and I.A. Chowdhury, "Design and Simulation of a High Performance 5G mm-Wave MIMO Antenna Array for Mobile Applications", *J. of Advanced Engineering and Computation*, vol. 8, art. no. 437, 2024 (https://doi.org/10.55579/jaec.202481.437).

[10] L.T. Tu, A. Bradai, Y. Pousset, and A.I. Aravanis, "Energy Efficiency Analysis of LoRa Networks", *IEEE Wireless Communications Letters*,

vol. 10, pp. 1881–1885, 2021 (https://doi.org/10.1109/LWC.2021.3084996).
[11] L. Beltramelli, A. Mahmood, P. Osterberg, and M. Gidlund, "LoRa Beyond ALOHA: An Investigation of Alternative Random Access Protocols", *IEEE Tran. on Industrial Informatics*, vol. 17, pp. 3544–3554, 2021 (https://doi.org/10.1109/TII.2020.2977046).
[12] T.T. Duy *et al.*, "On the Performance of the Coverage Probability of LoRa Networks with Non-linear Energy Harvesting", *2024 International Conference on Advanced Technologies for Communications (ATC)*, Ho Chi Minh City, Vietnam, 2024 (https://doi.org/10.1109/ATC62344.2024.10908157).
[13] A. Hoeller *et al.*, "Analysis and Performance Optimization of LoRa Networks with Time and Antenna Diversity", *IEEE Access*, vol. 6, pp. 32820–32829, 2018 (https://doi.org/10.1109/ACCESS.2018.2839064).
[14] T.H. Nguyen *et al.*, "Performance Analysis and Optimization of the Coverage Probability in Dual Hop LoRa Networks with Different Fading Channels", *IEEE Access*, vol. 8, pp. 107087–107102, 2020 (https://doi.org/10.1109/ACCESS.2020.3000600).
[15] W.U. Khan *et al.*, "RIS-assisted Energy-efficient LEO Satellite Communications with NOMA", *IEEE Transactions on Green Communications and Networking*, vol. 8, pp. 780–790, 2024 (https://doi.org/10.1109/TGCN.2023.3344102).
[16] I. Khirwar, A. Gupta, B. Verma, and P. Garg, "Asymmetric Dual-hop NOMA Based Communication System with Practical Constraints: Imperfect CSI, HWI and ISIC", *2023 9th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2023 (https://doi.org/10.1109/ICSC60394.2023.10441547).
[17] P.Q. Truong *et al.*, "Computation Offloading and Resource Allocation Optimization for Mobile Edge Computing-aided UAV-RIS Communications", *IEEE Access*, vol. 12, pp. 107971–107983, 2024 (https://doi.org/10.1109/ACCESS.2024.3435483).
[18] P. Saikia, K. Singh, W.J. Huang, and T.Q. Duong, "Hybrid Deep Reinforcement Learning for Enhancing Localization and Communication Efficiency in RIS-aided Cooperative ISAC Systems", *IEEE Internet of Things Journal*, vol. 11, pp. 29494–29510, 2024 (https://doi.org/10.1109/JIOT.2024.3411158).
[19] V.T. Ty *et al.*, "Security-reliability Tradeoff of Multi-hop Secure Communication Networks Using Fountain Codes and RIS-aided Cooperative Communication", *2023 International Conference on Advanced Technologies for Communications (ATC)*, Da Nang, Vietnam, 2023 (https://doi.org/10.1109/ATC58710.2023.10318517).
[20] X. Li *et al.*, "Performance Analysis of Star-RIS-CR-NOMA-based Consumer IoT Networks for Resilient Industry 5.0", *IEEE Transactions on Consumer Electronics*, vol. 70, pp. 1380–1391, 2024 (https://doi.org/10.1109/TCE.2023.3319402).
[21] L.T. Tu *et al.*, "Performance of RIS-assisted Two-way Communications with Phase Noise", AETA 2022 – Recent Advances in Electrical Engineering and Related Sciences: Theory and Application, pp. 707–718, 2022 (https://doi.org/10.1007/978-981-99-8703-0_59).
[22] X. Chen *et al.*, "Performance of Cell-free Massive MIMO Systems with Channel Aging and Phase Noise Over Rician Fading Channels", *IEEE Transactions on Vehicular Technology*, vol. 74, pp. 17765–17778, 2025 (https://doi.org/10.1109/TVT.2025.3581908).
[23] L.T. Tu, V.N.Q. Bao, and B. An, "On the Performance of Outage Probability in Underlay Cognitive Radio with Imperfect CSI", *2013 International Conference on Advanced Technologies for Communications (ATC 2013)*, Ho Chi Minh City, Vietnam, 2013 (https://doi.org/10.1109/ATC.2013.6698091).
[24] X. Zhang, N.P. Le, and M.S. Alouini, "RIS-based DOA Estimation for Communication-assisted Sensing Systems Under Hardware Impairments", *IEEE Open Journal of Vehicular Technology*, vol. 6, pp. 1736–1748, 2025 (https://doi.org/10.1109/OJVT.2025.3580041).
[25] P.T. Tin *et al.*, "Rateless Codes-based Secure Communication Employing Transmit Antenna Selection and Harvest-to-jam Under Joint Effect of Interference and Hardware Impairments", *Entropy*, vol. 21, art. no. 700, 2019 (https://doi.org/10.3390/e21070700).
[26] X. Zhang *et al.*, "A New Reconfigurable Intelligent-surface-assisted LoRa System", *IEEE Transactions on Vehicular Technology*, vol. 71, pp. 9055–9060, 2022 (https://doi.org/10.1109/TVT.2022.3173675).
[27] S.P. Le *et al.*, "On the Secrecy Performance of Reconfigurable Intelligent Surfaces-assisted Satellite Networks Under Shadow-Rician Channels", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 61, pp. 6794–6808, 2025 (https://doi.org/10.1109/TAES.2025.3532227).
[28] M.K. Simon and M.S. Alouini, *Digital Communication over Fading Channels*, Wiley, Newark, 900 p., 2005 (https://doi.org/10.1002/0471715220).
[29] J. Breitegger *et al.*, "Long-term LoRa Experiments in a Chemical Plant", *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, Valencia, Spain, 2021 (https://doi.org/10.1109/ICIT46573.2021.9453474).
[30] E. Bjornson, M. Matthaiou, and M. Debbah, "A New Look at Dual-hop Relaying: Performance Limits with Hardware Impairments", *IEEE Transactions on Communications*, vol. 61, pp. 4512–4525, 2013 (https://doi.org/10.1109/TCOMM.2013.100913.130282).
[31] N.T. Anh *et al.*, "Reliability-security Analysis for Harvest-to-jam Based Multi-hop Cluster MIMO Networks Using Cooperative Jamming Methods Under Impact of Hardware Impairments", *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 8, pp. 1–11, 2021 (https://doi.org/10.4108/eai.17-9-2021.170963).
[32] C. Dhote, A. Singh, and P.K. Sharma, "RIS Assisted Near Field Communication in Sub-6 GHz Band: Experimental Perspective", *IEEE Communications Letters*, vol. 29, pp. 1814–1818, 2025 (https://doi.org/10.1109/LCOMM.2025.3576503).
[33] L. Dai *et al.*, "Reconfigurable Intelligent Surface-based Wireless Communications: Antenna Design, Prototyping, and Experimental Results", *IEEE Access*, vol. 8, pp. 45913–45923, 2020 (https://doi.org/10.1109/ACCESS.2020.2977772).
[34] P.T. Tran, T.N. Nguyen, and L.T. Tu, *Developing MATLAB Comprehensive Simulation for 5G/6G Wireless Communications*, Springer Singapore, 2026 (ISBN 9789819511167).
[35] M.A. Badiu and J.P. Coon, "Communication Through a Large Reflecting Surface with Phase Errors", *IEEE Wireless Communications Letters*, vol. 9, pp. 184–188, 2020 (https://doi.org/10.1109/LWC.2019.2947445).

Lam-Thanh Tu, Ph.D.
Communication and Signal Processing Research Group
Faculty of Electrical and Electronics Engineering
https://orcid.org/0000-0001-5735-4641
E-mail: tulamthanh@tdtu.edu.vn
Ton Duc Thang University, Ho Chi Minh City, Vietnam
https://tdtu.edu.vn

Bui Vu Minh, M.Sc.
Faculty of Engineering and Technology
https://orcid.org/0000-0003-0222-166X
E-mail: bvminh@ntt.edu.vn
Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
https://ntt.edu.vn

Tan N. Nguyen, Ph.D.
Advanced Intelligent Technology Research Group
Faculty of Electrical and Electronics Engineering
https://orcid.org/0000-0002-2286-6652
E-mail: nguyennhattan@tdtu.edu.vn
Ton Duc Thang University, Ho Chi Minh City, Vietnam
https://tdtu.edu.vn

Optimizing Circular Arrays with Concentric Subarray Rings for Wireless Power Transmission Applications

Jafar Ramadhan Mohammed

Ninevah University, Mosul, Iraq

https://doi.org/10.26636/jtit.2026.1.2387

Abstract — The emerging wireless power transmission technology creates new opportunities in numerous real-world applications such as wireless charging systems, robots, and aerospace solutions. This paper introduces an optimized method for designing transmit antenna arrays which may be used for long-distance wireless power transmission with narrow focusing of RF power on remote receivers. The novelty of this paper consists in using an effective clustered subarray rings configuration with a transmit circular array instead of its conventional full aperture array, based on the configuration of individual elements. The final goal is to obtain simpler, cheaper, and lighter arrays. Amplitude and phase excitation weighting as well as the number of elements in each clustered subarray are optimized jointly to maximize the efficiency of transferring power to a target region while minimizing sidelobe powers outside the intended region. The simulation results show that the beam collection efficiency of the proposed system with 21 subarray rings was 98.99%, while that of the conventional circular array with individual dense elements of size 21×21 equaled 99.68%.

Keywords — beam collection efficiency, circular array, concentric ring subarrays, wireless power transmission

1. Introduction

Wireless power transmission (WPT) is a system that inverts electrical power into microwave power and transmits it through a wireless channel. Array antennas are used to transmit microwave powers to the intended regions [1]. At the receiving site, rectifier antennas are used to capture microwave powers and reconvert them back into electrical power in order to feed the target receiver devices [2], [3].

Real-life applications of long-distance WPT include charging systems for electric vehicles and electronic devices such as smartphones, smartwatches, and biomedical devices, including those that can be implanted in the human body. The technology may be also used in drone-based RFID systems for monitoring living organisms and may be relied upon in the aerospace industry for transferring power between spacecraft. Industrial robots and automated systems are common use cases of this technology as well [4], [5].

Phased arrays are deployed in WPT to highly efficiently concentrate RF power delivered to remote receivers while minimizing sidelobe powers present outside the intended receivers [6]–[10]. High efficiency, narrow focused arrays require numerous radiating elements, and the resulting arrays with a highly dense arrangement are capable of achieving excellent power transmission efficiency and range. However, these large-scale arrays often face critical challenges such as high cost, high power consumption, and increased hardware complexity [11]–[15]. These drawbacks are proportional to the number of elements in the array or the size of the entire array. This renders the hardware impractical.

In contrast, clustered subarray designs offer significant advantages, such as cost reduction, lower power consumption, and reduced hardware complexity, as their array elements are grouped into specific categories [16]–[19]. Each cluster shares a common amplitude and phase excitation weighting mechanism. Therefore, its array feeding network is much simpler, lighter, and consumes less power – with all those design parameters being of high significance when deploying these transmit arrays in limited space systems, such as drones [20]–[22]. Although clustered arrays provide near-optimal solutions with a smaller number of design parameters, it meets better hardware constraints.

The most important performance metric used to evaluate WPT systems is beam collection efficiency (BCE) – a parameter which may be defined as the ratio of the received power from the rectifying antenna to the total transmitted power [23]. Small size arrays with fewer radiating elements are not capable of achieving the required BCE values. Therefore, larger arrays with dense element arrangements are generally used in these applications.

One of the effective approaches to implement such large-scale arrays relies on clustered subarray configurations [24]–[26]. Clustered subarrays are practically more efficient than fully dense element arrangements. Thus, only a small number of clusters is needed to meet the required BCE values. The works described in [26], [27] presented an optimization algorithm used to generate clustered subarrays, while in [28]–[30], clustered rings were used to design rectangular planar arrays. The design of circular planar arrays with concentric subarray rings is a new idea and can offer a great advantage when it comes to modern wireless communication system applications. The idea is to effectively approximate amplitude excitation weights of the fully circular array elements by using a set of concentric subarray rings starting from the largest (or outer) ring at the array's perimeter and ending at

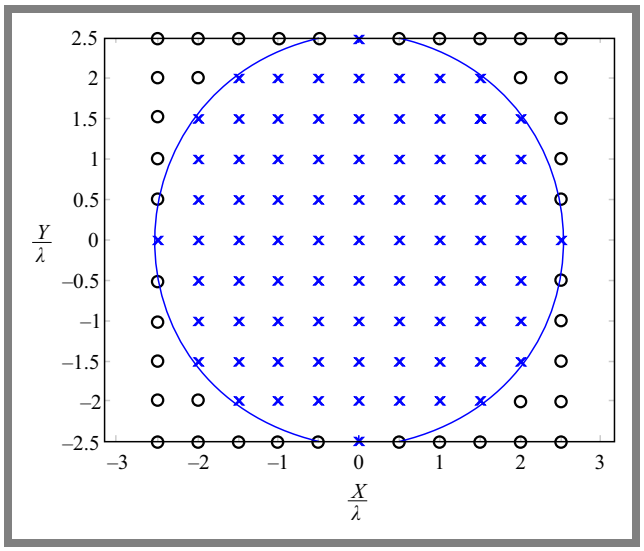

Fig. 1. Geometry of the circular planar array.

the center of the array – in a manner resembling the shape of a circular, ascending ladder.

This new configuration of clustered rings is presented in this paper for the very first time. Once the concentric subarray rings have been created, their amplitude and phase distributions are optimized to maximize the efficiency of transferring power to a target region while minimizing sidelobe powers outside the intended region.

To improve the BCE of the wireless power transmission system, a hybrid optimization algorithm is used in the clustered rings configuration, where the genetic algorithm and the k-means clustering approach are used.

Implementation of the proposed clustered level array instead of its conventional element-based counterpart offers many advantages, such as simplified feeding network, efficient taper efficiency, lower side lobe level, and higher directivity.

2. Circular Array Based on Individual Elements

A circular array with N individual elements is formed by considering a regular square-grid array element with a circular boundary at the array ends, as shown in Fig. 1.

The corner elements that are located outside the circular boundary are turned off, while all the remaining elements that are located inside the circular boundary are active and are assumed to be isotropic. The element pattern and the mutual coupling effect between the active elements are neglected, and thus the array factor with such a configuration can be given by:

$$AF(u,v) = \sum_{n=1}^{N} w_n e^{j[k(ux_n+vy_n)]} , \tag{1}$$

where $w_n = Amp_n\, e^{jPh_n}$ is the element excitation weight in terms of amplitude Amp_n and phase Ph_n of the n-th element, x_n, y_n is the element's location along the x and y axes, $k = \frac{2\pi}{\lambda}$, λ is the wavelength, $u = \sin\theta\cos\varphi$, $v = \sin\theta\sin\varphi$ are the angular coordinates, θ and φ are the elevation and azimuth angles, respectively.

The beam collection efficiency (BCE) metric is defined as the ratio of the received power at the rectifying antenna to the total transmitted power:

$$BCE = \frac{P_R}{P_T} = \frac{\int_{\psi} |AF(u,v)|^2 \,\mathrm{d}u\,\mathrm{d}v}{\int_{\phi} |AF(u,v)|^2 \,\mathrm{d}u\,\mathrm{d}v} , \tag{2}$$

where ψ and ϕ are the fields of view for the received and transmitted powers, respectively.

For a rectangular grid array, the range of either ψ or ϕ can be represented by:

$$\psi = \left\{(u,v) : -u_0 \leqslant u \leqslant u_0,\ -v_0 \leqslant v \leqslant v_0,\right\} . \tag{3}$$

Thus, the denominator of Eq. (2) can be expressed as:

$$P_R = \int_{-u_0}^{u_0}\int_{-v_0}^{v_0} e^{jk(x_m u+vy_m)} e^{-jk(x_n u+vy_n)} \mathrm{d}u\,\mathrm{d}v . \tag{4}$$

$AF(u,v)$ can be found in terms of the element excitation weights w_n^e of the array elements and the steering vector $\boldsymbol{V}(u,v)$, in the following manner:

$$AF(u,v) = \boldsymbol{w}^H \boldsymbol{V}(u,v) , \tag{5}$$

where:

$$\boldsymbol{w} = [w_1^e\, w_2^e \ldots w_N^e]^H,\ n = 1,2,\ldots,N ,$$

$$\boldsymbol{V}(u,v) = \left[e^{-jk(x_1u+vy_1)}\, e^{-jk(x_2u+vy_2)} \ldots e^{-jk(x_Nu+vy_N)}\right]^H ,$$

and H is the conjugate transpose. Now, substituting Eq. (3) in Eq. (2), BCE for the circular planar array based on the individual element configuration can be rewritten as:

$$BCE = \frac{\boldsymbol{w}^H\{\int_{\psi} \boldsymbol{V}(u,v)\boldsymbol{V}^H(u,v)\,\mathrm{d}u\,\mathrm{d}v\}\boldsymbol{w}}{\boldsymbol{w}^H\{\int_{\phi} \boldsymbol{V}(u,v)\boldsymbol{V}^H(u,v)\,\mathrm{d}u\,\mathrm{d}v\}\boldsymbol{w}} . \tag{6}$$

3. Circular Array Based on Concentric Subarray Rings

The circular array based on the dense element arrangements that was shown in the previous section is now divided into several concentric subarray rings or clusters equal to C, as shown in Fig. 2. The number of cluster rings is always less than the total number of the individual elements, i.e. $C < N$. The individual elements and the clustered subarray rings of the circular array are related by a matrix with dimension $N \times C$, as follows:

$$M = \begin{bmatrix} M_{11} & \ldots & M_{1C} \\ \vdots & \vdots & \vdots \\ M_{N1} & \ldots & M_{NC} \end{bmatrix} , \tag{7}$$

where

$$M_{nc} = \begin{cases} 1 \text{ if } n\text{-th element belongs to } c\text{-th cluster} \\ 0 \text{ if } n\text{-th element does not belong to } c\text{-th cluster.} \end{cases} \tag{8}$$

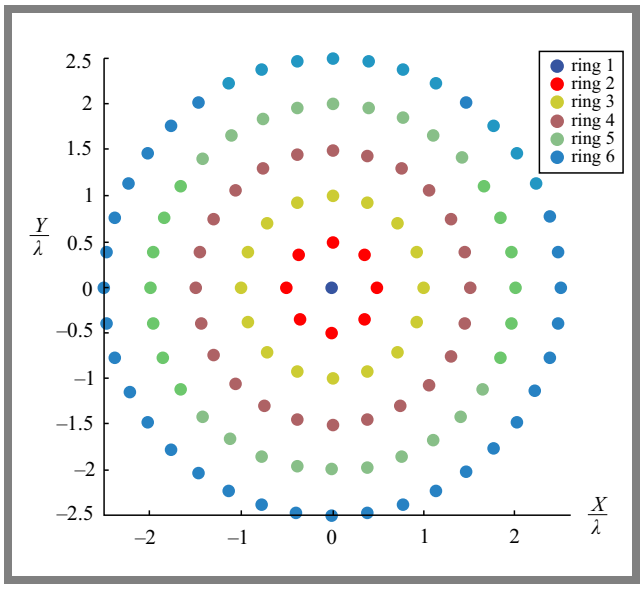

Fig. 2. Geometry of the proposed circular array with concentric subarray rings.

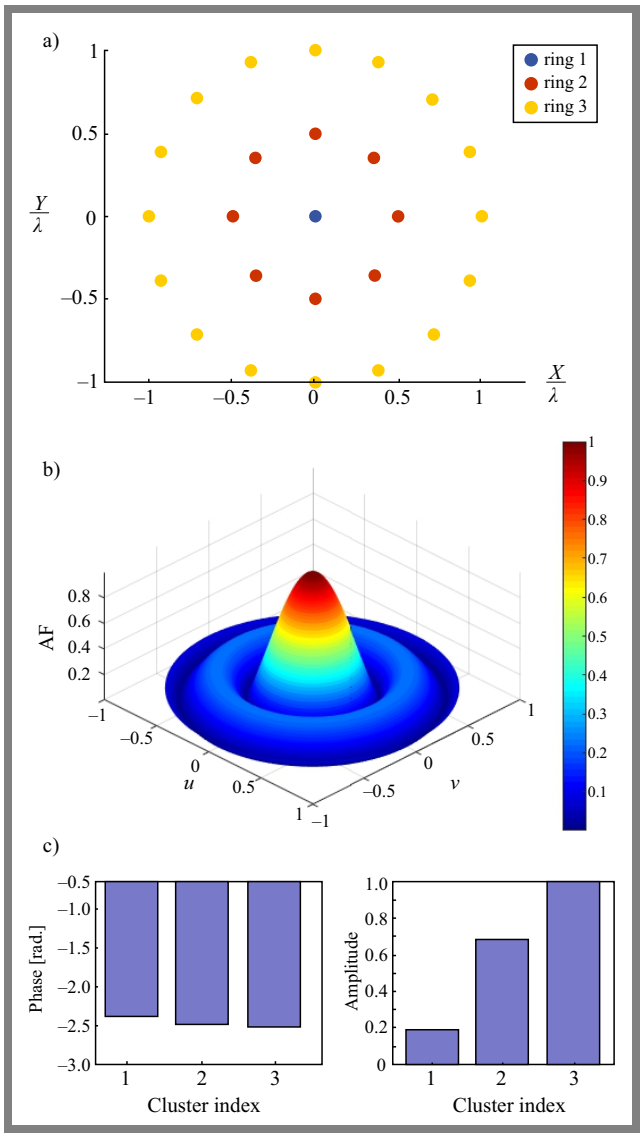

Fig. 3. Results for concentric subarray rings for $C = 3$.

The excitation weights of the resulting cluster rings can be presented by a vector of dimension $C \times 1$ as $\boldsymbol{w^c} = [w_1^c\ w_2^c\ \ldots\ w_C^c]^H$. Then, the excitation weights of each element in the clustered array can be written as:

$$\boldsymbol{w}_{new} = \boldsymbol{M}\,\boldsymbol{w^c}\,. \tag{9}$$

The beam-collecting efficiency of the circular array based on concentric subarray rings can be given by:

$$BCE_{rings} = \frac{\boldsymbol{w}_{new}^H \left\{ \int_\psi \boldsymbol{V}(u,v)\boldsymbol{V}^H(u,v)\,\mathrm{d}u\,\mathrm{d}v \right\} \boldsymbol{w}_{new}}{\boldsymbol{w}_{new}^H \left\{ \int_\phi \boldsymbol{V}(u,v)\boldsymbol{V}^H(u,v)\,\mathrm{d}u\,\mathrm{d}v \right\} \boldsymbol{w}_{new}}\,. \tag{10}$$

The fields of view of the transmitting ϕ, and receiving ψ arrays were of the rectangular geometry, as given by Eq. (3). Here, the fields of view are extended to descript the circular target region as follows:

$$\psi = \left\{ (u,v) : \sqrt{u^2+v^2} \leqslant r_0 \leqslant \sin\theta_o \right\}. \tag{11}$$

Where r_0 represents the radius of the circular aperture and θ_0 is the main beam direction of the pattern. The denominator of Eq. (10) can be expressed as:

$$P_R = \int_0^{\theta_0} \int_0^{2\pi} e^{jk(x_m+y_m)\sin\theta\cos\phi} e^{-jk(x_n+y_n)\sin\theta\sin\phi} \mathrm{d}\phi \sin\theta \mathrm{d}\theta\,. \tag{12}$$

From Eqs. (4) and (12), it was shown that beam collection efficiency is a function of the design variables of the excitation weights at the element level, i.e. w^e for the ordinary circular array without subarray rings, or of the excitation weights at the subarray level, i.e. w^c for the circular array based on clustered subarray rings. A circular planar array with concentric subarray rings is considered here and each ring represents a single cluster. The following steps are used to find the excitation weights of the clustered subarray rings w^c [31].

Starting from the central element of the array, the first clustered weight w_1^c is calculated as $w_1^c = \frac{w_i^e + w_j^e}{2}$, where $i, j = 1, 2, \ldots, N$, $i \neq j$, w_i^e and w_j^e are the element excitation weights of two elements that have the smallest distance equal to $|w_i^e - w_j^e|$. Then, the second clustered weight w_2^c is computed from the largest distance between remaining element excitation weights and the first cluster weight, i.e., $w_2^c = \max\{d(w_i^e, w_1^c)\}$. The third clustered weight w_3^c, is derived from the smallest distance from w_1^c and w_2^c as:

$$w_3^c = \max\left\{ \min\{d(w_i^e, w_1^c), d(w_i^e, w_2^c)\} \right\}.$$

The remaining clustered excitation weights are computed as follows:

$$w_k^c = \max\left\{ \min\{d(w_i^e, w_1^c), d(w_i^e, w_2^c), \ldots, d(w_i^e, w_C^c)\} \right\}$$
$$k = 1, 2, \ldots, C\,, \tag{13}$$

where C is the total number of cluster rings.

To obtain better results, these clustered excitation weights are considered initial values for the optimized algorithm. A genetic algorithm with the following cost function was used

to maximize beam collection efficiency:

$$F\{w_1^c, w_2^c, \ldots, w_C^c, N_1^c, N_2^c, \ldots, N_C^c\} = -BCE$$
$$\text{s.t. } 0 < w_k^c \leqslant 1 \text{ and } 1 \leqslant N_k^c \leqslant N - C+ \tag{14}$$

where N_k^c is the number of array elements in each clustered subarray ring $k = 1, 2, \ldots, C$.

4. Simulation Results

In order to illustrate the effectiveness of the proposed circular array with concentric subarray rings for wireless power transmission, extensive simulation results are presented below:

In all considered examples, the distance between transmitter and receiver sites is assumed to be 100 m and the target region is limited by the angular separation of $\pm 3^\circ$ on both sides of the broadside direction, along both elevation and azimuth planes. The transmitting power is assumed to be 1 W. The radial spacing between any two successive subarray rings is 0.5λ. For genetic optimization, we used the built-in function called `fmincon` with the maximum number of iterations equal to 400.

In the first example, the circular radius of the transmitting circular array is $R_t = 1.0\lambda$, i.e., the number of the circular subarray rings is $C = 3$, including the central element of the array. For this case, the normalized transmitted and received power values are 1 and 0.8486, respectively, while the beam collection efficiency is BCE $= 84.59\%$. The geometrical configuration of the resulting circular array with 3 concentric cluster rings, its corresponding far-field radiation pattern as well as their subarray amplitude and phase excitation weights are shown in Fig. 3.

The normalized received power variations in both RF and DC values, as a function of the target angular region bounded by $\pm 3^\circ$, is shown in Fig. 4. For comparison purposes, the results of the circular array based on individual elements are depicted in Fig. 5.

From the figures presenting the array geometries (compare Fig. 3a with Fig. 5a), it is clear that the array feeding network in the proposed arrangement with circular subarray rings is less complex. Also, from figures of radiation patterns (Fig. 3b with Fig. 5b), it becomes clear that the pattern of the proposed circular subarray rings is almost like that of the conventional array with dense element arrangements and feeding RF components.

In the second example, the design of circular clustered subarrays with different numbers of rings was studied. The results are shown in Figs. 6–8. It is obvious from those results that the concentration of radiation patterns improves along with an increase in the number of concentric subarray rings. On the other hand, array geometries are becoming more complicated.

In the next example, the design of the array feeding structure is considered, where the number of concentric subarray rings equaled 6, including the central element. Figure 9 shows the geometric layout of the designed feeding network for the circular array with concentric subarray rings.

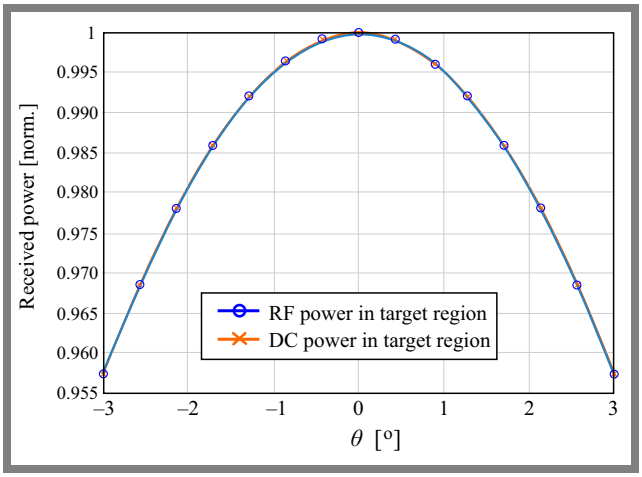

Fig. 4. Variations in received power versus target region for $C = 3$.

Tab. 1. Beam collection efficiencies for various subarray rings.

Method	Normalized P_T	Normalized P_R	BCE [%]
Circular array based on individual elements	1	0.996	99.68
Circular array with 3 subarray rings	1	0.848	84.86
Circular array with 6 subarray rings	1	0.870	87.01
Circular array with 10 subarray rings	1	0.945	94.54
Circular array with 21 subarray rings	1	0.989	98.99

The number of transmitting elements in the first cluster ring amounted to 8, while the number of other rings equaled 16, 24, 32, and 40. Since all transmitting elements that belong to a certain ring share a common amplitude and phase, the required number of amplitude and phase weightings (i.e., the number of RF attenuators and phase shifters) is equal to the number of subarray clustered rings $C = 6$.

From Fig. 9, the central individual element is left without being connected to any attenuators or phase shifters, since its normalized amplitude weighting is one while its phase is zero. The proposed circular subarray rings require only 5 RF attenuators and 5 phase shifters, whereas a conventional circular array based on individual element arrangements needs 140 RF attenuators and 140 phase shifters. This great reduction makes the proposed circular array lighter, lowers its power consumption, renders simpler, and cheaper.

Table 1 shows the variations of the beam collection efficiencies versus the number of clustered subarray rings. BCE improves as the number of subarray rings increases.

From this table, one may notice that BCE of the ordinary circular matrix with individual elements of size 21×21 is better than the results from the proposed matrix with a certain number of sub-array rings. However, the difference between these two values vanishes for a larger number of subarray rings.

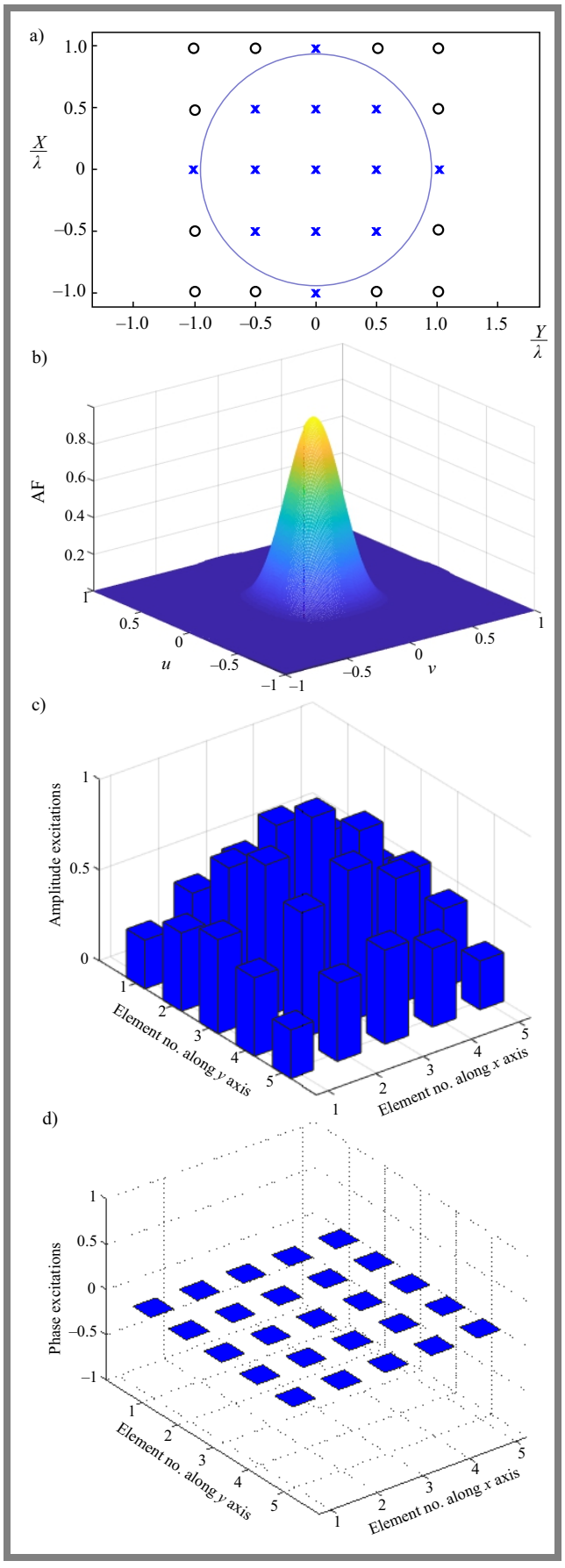

Fig. 5. Results for the conventional circular array with individual elements of size 5×5.

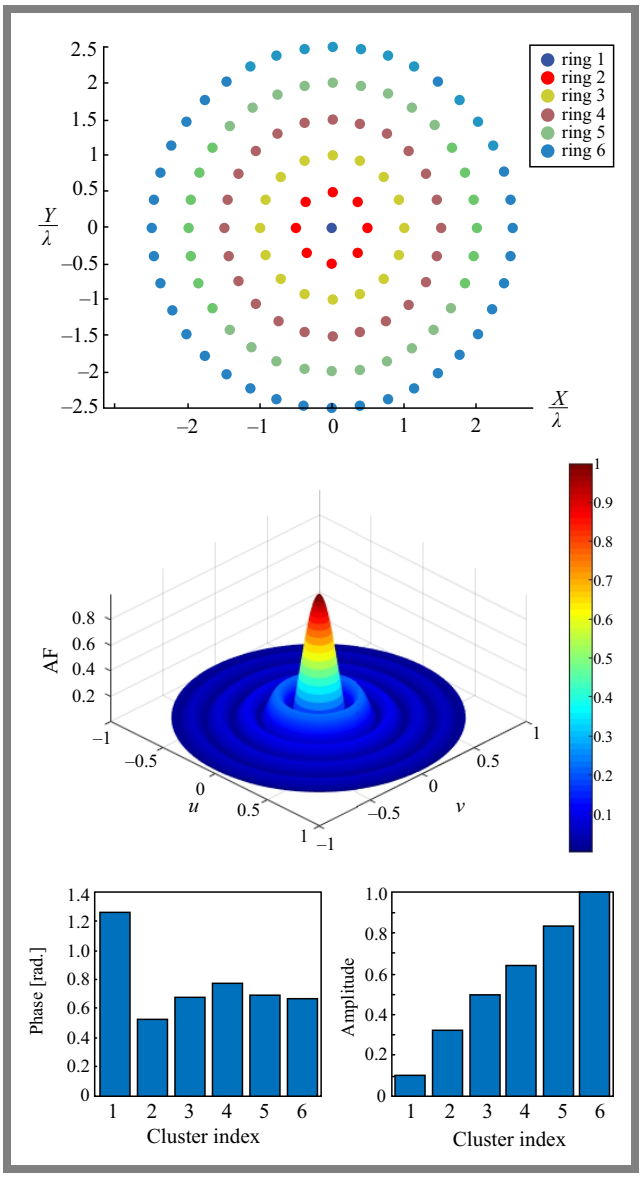

Fig. 6. Results for concentric subarray rings for $C = 6$.

5. Conclusions

A circular planar array with concentric subarray rings was proposed as a simpler alternative to conventional fully dense array elements in order to improve beam collection efficiency while simultaneously significantly reducing the number of hardware components making up the array feeding network Thus, the designed array becomes lighter and consumes less power, which is very important in applications that require long-term operation with limited resources and space.

A k-means clustering approach was adopted and optimization of the genetic algorithm was performed to optimize the design parameters of the proposed concentric subarray rings (i.e., amplitude and phase excitation weights of the subarray rings, number of rings, and the total number of elements in each separate ring), in order to maximize beam collection efficiency.

From the results, it was shown that BCE, radiation pattern, array structure, and feeding network complexity of the proposed

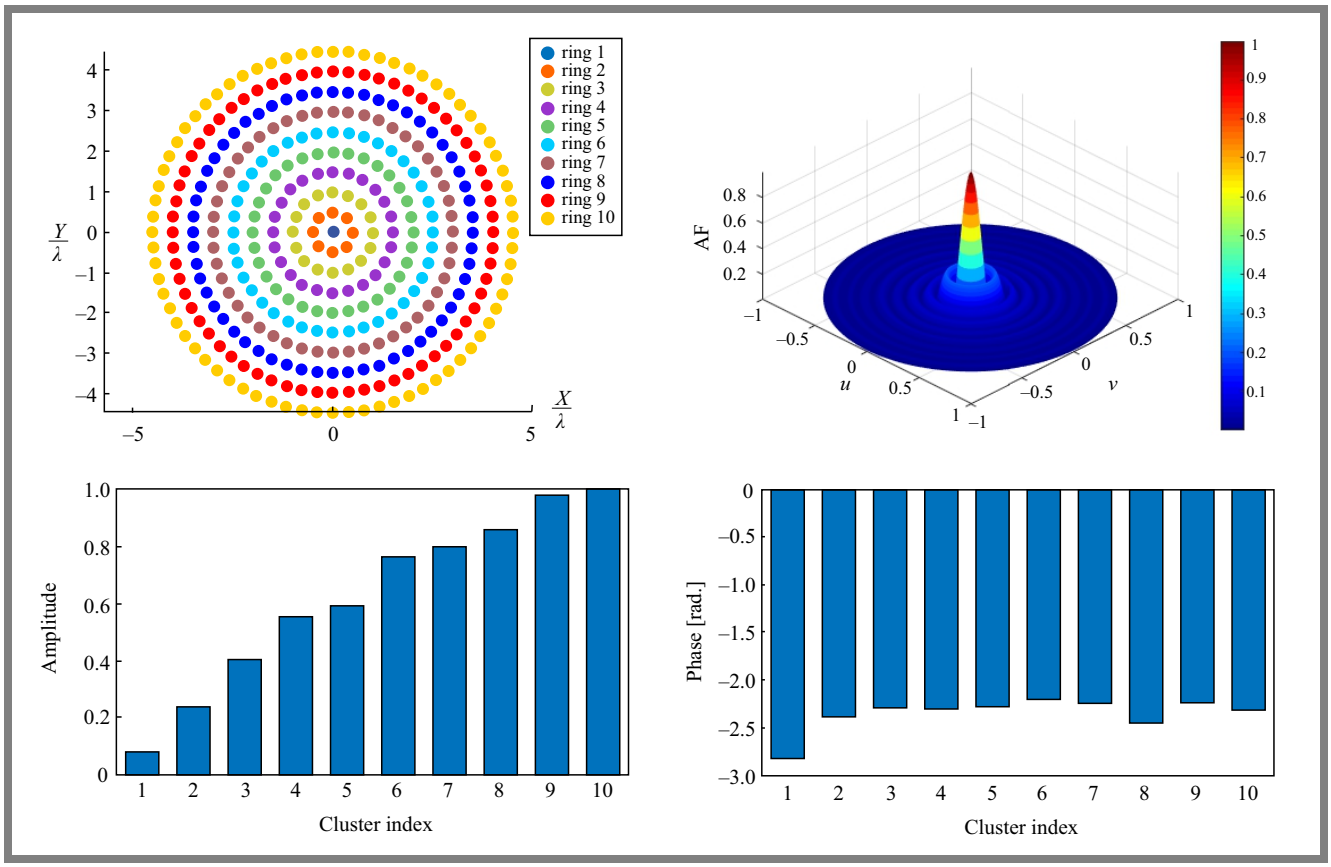

Fig. 7. Results for the proposed antenna for $C = 10$.

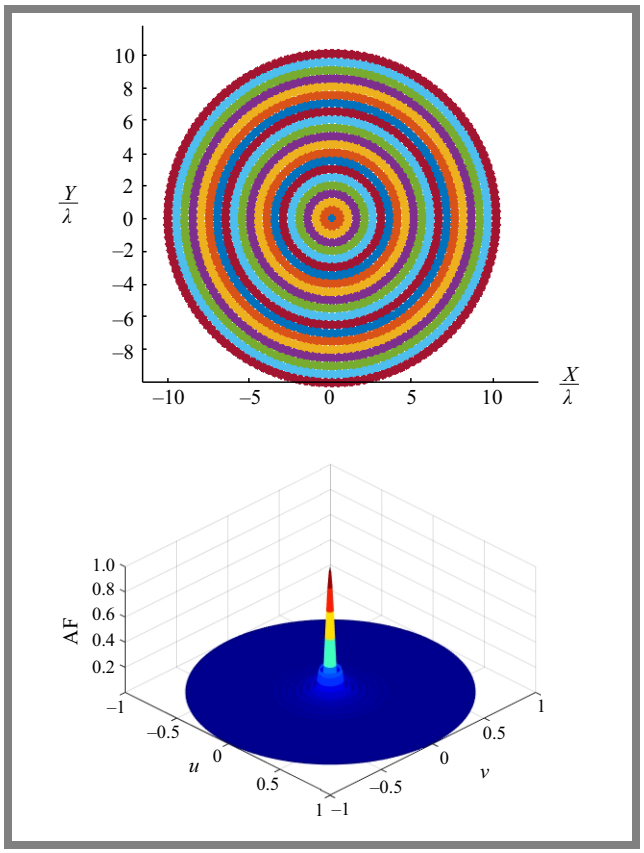

Fig. 8. Results for the proposed antenna for $C = 12$.

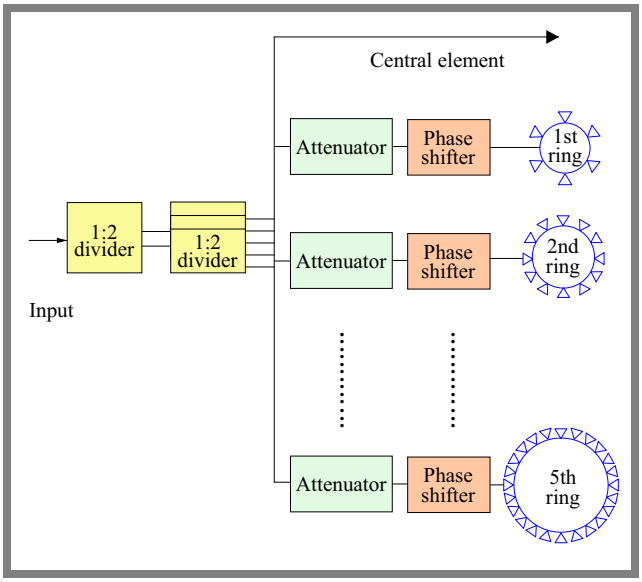

Fig. 9. Implementation of the proposed circular array with concentric subarray rings.

array are much better than in the case of the conventional circular array based on individual elements. This proves the effectiveness and practical applicability of the proposed array. The described idea may be expanded in future research to include multiple wireless power transmission receivers, with each receiver being at a different location, so that the optimizer maximizes the received powers over all of them.

References

[1] A. Massa, G. Oliveri, F. Viani, and P. Rocca, "Array Designs for Long-distance Wireless Power Transmission: State-of-the-art and Innovative Solutions", *Proceedings of the IEEE*, vol. 101, pp. 1464–1481, 2013 (https://doi.org/10.1109/JPROC.2013.2245491).

[2] R.H. Nansen, "Wireless Power Transmission: The Key to Solar Power Satellites", *IEEE Aerospace and Electronic Systems Magazine*, vol. 11, pp. 33–39, 1996 (https://doi.org/10.1109/62.484148).

[3] N. Shinohara, "Beam Control Technologies with a High-efficiency Phased Array for Microwave Power Transmission in Japan", *Proceedings of the IEEE*, vol. 101, pp. 1448–1463, 2013 (https://doi.org/10.1109/JPROC.2013.2253062).

[4] Y. Wang *et al.*, "Design of a Microwave Power Transmission Demonstration System for Space Solar Power Station", *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 32, art. no. e23523, 2022 (https://doi.org/10.1002/mmce.23523).

[5] Y. Song *et al.*, "Research on the Multiobjective Optimization of Microwave Wireless Power Receiving in an Unmanned Aerial Vehicle Network", *Complexity*, vol. 2020, art. no. 8882528, 2020 (https://doi.org/10.1155/2020/8882528).

[6] G. Oliveri, L. Poli, and A. Massa, "Maximum Efficiency Beam Synthesis of Radiating Planar Arrays for Wireless Power Transmission", *IEEE Transactions on Antennas and Propagation*, vol. 61, pp. 2490–2499, 2013 (https://doi.org/10.1109/TAP.2013.2241714).

[7] J.R. Mohammed, "Synthesizing Non-uniformly Excited Antenna Arrays Using Tiled Subarray Blocks", *Journal of Telecommunications and Information Technology*, vol. 2023, pp. 25–29, 2023 (https://doi.org/10.26636/jtit.2023.4.1417).

[8] J.R. Mohammed and K.H. Sayidmarie, "Sensitivity of the Adaptive Nulling to Random Errors in Amplitude and Phase Excitations in Array Elements", *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, pp. 51–56, 2018 (https://jtec.utem.edu.my/jtec/article/view/2023).

[9] J.R. Mohammed, "Beam-pattern Control via Thinned Elements Strategy in Linear and Planar Phased Arrays", *Progress In Electromagnetics Research Letters*, vol. 111, pp. 79–84, 2023 (https://doi.org/10.2528/PIERL23022403).

[10] A.J. Abdulqader, J.R. Mohammed, and A.A. Yaser, "A T-shaped Polyomino Subarray Design Method for Controlling Sidelobe Level", *Progress In Electromagnetics Research C*, vol. 126, pp. 243–251, 2022 (https://doi.org/10.2528/PIERC22080803).

[11] T. Isernia, M. D'Urso, and O.M. Bucci, "A Simple Idea for an Effective Sub-arraying of Large Planar Sources", *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 169–172, 2008 (https://doi.org/10.1109/LAWP.2008.2000943).

[12] J.R. Mohammed, "A Method for Thinning Useless Elements in the Planar Antenna Arrays", *Progress In Electromagnetics Research Letters*, vol. 97, pp. 105–113, 2021 (https://doi.org/10.2528/PIERL21022104).

[13] G. Oliveri and L. Poli, "Optimal Sub-arraying of Compromise Planar Arrays Through an Innovative ACO-weighted Procedure", *Progress In Electromagnetics Research*, vol. 109, pp. 279–299, 2010 (https://doi.org/10.2528/PIER10092008).

[14] J.R. Mohammed, "An Optimum Side-lobe Reduction Method with Weight Perturbation", *Journal of Computational Electronics*, vol. 18, pp. 705–711, 2019 (https://doi.org/10.1007/s10825-019-01323-5).

[15] J.R. Mohammed, R.H. Thaher, and A.J. Abdulkadeer, "Linear and Planar Array Pattern Nulling via Compressed Sensing", *Journal of Telecommunications and Information Technology*, vol. 2021, pp. 50–55, 2021 (https://doi.org/10.26636/jtit.2021.152921).

[16] L. Xun, D. Baoyan, and S. Liwei, "Design of Clustered Planar Arrays for Microwave Wireless Power Transmission", *IEEE Transactions on Antennas and Propagation*, vol. 67, pp. 606–611, 2019 (https://doi.org/10.1109/TAP.2018.2874670).

[17] P. Rocca, N. Anselmi, A. Polo, and A. Massa, "An Irregular Two-sizes Square Tiling Method for the Design of Isophoric Phased Arrays", *IEEE Transactions on Antennas and Propagation*, vol. 68, pp. 4437–4449, 2020 (https://doi.org/10.1109/TAP.2020.2970088).

[18] M. Salucci, G. Gottardi, N. Anselmi, and G. Oliveri, "Planar Thinned Array Design by Hybrid Analytical-stochastic Optimization", *IET Microwaves, Antennas and Propagation*, vol. 11, pp. 1841–1845, 2017 (https://doi.org/10.1049/iet-map.2017.0349).

[19] J.R. Mohammed, "Unconventional Method for Antenna Array Synthesizing Based on Ascending Clustered Rings", *Progress In Electromagnetics Research Letters*, vol. 117, pp. 69–73, 2024 (https://doi.org/10.2528/PIERL23122201).

[20] R. Vescovo and L. Pajewski, "Multiple-ring Circular Array for Ground-penetrating Radar Applications: Basic Ideas and Preliminary Results", *Journal of Telecommunications and Information Technology*, vol. 2017, pp. 25–29, 2017 (https://doi.org/10.26636/jtit.2017.120217).

[21] B. Yektakhah, A.M.H. Nasr, A.H. Mohamed, and K. Sarabandi, "Low-complexity Wideband Circularly Polarized Modular Scalable Phased Array for Vehicular Satellite Communication", *IEEE Open Journal of Antennas and Propagation*, vol. 6, pp. 854–863, 2025 (https://doi.org/10.1109/OJAP.2025.3551624).

[22] J.R. Mohammed, "Design of Printed Yagi Antenna with Additional Driven Element for WLAN Applications", Progress In Electromagnetics Research C, vol. 37, pp. 67–81, 2013 (https://doi.org/10.2528/PIERC12121201).

[23] M.M. Rahman *et al.*, "A Comprehensive Review of Wireless Power Transfer Methods, Applications, and Challenges", *Engineering Reports*, vol. 6, art. no. e12951, 2024 (https://doi.org/10.1002/eng2.12951).

[24] J.R. Mohammed, "Rectangular Grid Antennas with Various Boundary Square-rings Array", *Progress In Electromagnetics Research Letters*, vol. 96, pp. 27–36, 2021 (https://doi.org/10.2528/PIERL20112402).

[25] J.R. Mohammed, A.J. Abdulkadeer, and R.H. Thaher, "Array Pattern Recovery Under Amplitude Excitation Errors Using Clustered Elements", *Progress In Electromagnetics Research M*, vol. 98, pp. 183–192, 2020 (https://doi.org/10.2528/PIERM20101906).

[26] A.E.A. Blomberg, A. Austeng, and R.E. Hansen, "Adaptive Beamforming Applied to a Cylindrical Sonar Array Using an Interpolated Array Transformation", *IEEE Journal of Oceanic Engineering*, vol. 37, pp. 25–34, 2012 (https://doi.org/10.1109/JOE.2011.2169611).

[27] J.R. Mohammed, "Phased Sub-arrays Pattern Synthesis Method with Deep Sidelobe Reduction and Narrow Beamwidth", *IETE Journal of Research*, vol. 68, pp. 3130–3137, 2020 (https://doi.org/10.1080/03772063.2020.1853616).

[28] J.R. Mohammed, A.J. Abdulkadeer, and R.H. Thaher, "Antenna Pattern Optimization via Clustered Arrays", *Progress In Electromagnetics Research M*, vol. 95, pp. 177–187, 2020 (https://doi.org/10.2528/PIERM20042307).

[29] J.R. Mohammed, "Simplified Rectangular Planar Array with Circular Boundary for Side Lobe Suppression", *Progress In Electromagnetics Research M*, vol. 97, pp. 57–68, 2020 (https://doi.org/10.2528/PIERM20062906).

[30] A.J. Abdulqader, J.R. Mohammed, and Y.E.M. Ali, "Beam Pattern Optimization via Unequal Ascending Clusters", *Journal of Telecommunications and Information Technology*, vol. 2023, pp. 1–7, 2023 (https://doi.org/10.26636/jtit.2023.168523).

[31] Q. Shi, Z. Zheng, and Y. Sun, "Pattern Synthesis of Subarrayed Large Linear and Planar Arrays Using K-means Solution", *IEEE Antennas and Wireless Propagation Letters*, vol. 20, pp. 693–697, 2021 (https://doi.org/10.1109/LAWP.2021.3060190).

Jafar Ramadhan Mohammed, Professor
College of Electronics Engineering
https://orcid.org/0000-0002-8278-6013
E-mail: jafar.mohammed@uoninevah.edu.iq
Ninevah University, Mosul, Iraq
https://uoninevah.edu.iq

Comparative Analysis of Classifiers for Higher-order Statistics-based Modulation Recognition in Cooperative STBC-OFDM

Brahim Dehri[1], Hakima Moulay[2], Ahmed Amine Daikh[1,3], and Mokhtar Besseghier[3]

[1]*University Centre of Naama, Naama, Algeria,*
[2]*University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria,*
[3]*University of Mustapha Stambouli, Mascara, Algeria*

https://doi.org/10.26636/jtit.2026.1.2299

Abstract — Precise classification of modulation in cooperative relaying networks remains challenging in the presence of carrier frequency offset (CFO) and imperfect channel state information (CSI). This paper conducts a comprehensive comparative analysis of automatic modulation classification (AMC) methods for distributed space-time block-coded orthogonal frequency division multiplexing (DSTBC-OFDM) systems under these impairments. A unified simulation framework is developed that combines pilot-assisted CFO and CSI estimation with higher-order statistics (HOS)-based feature extraction. Four widely used machine learning classifiers, i.e. feedforward neural network, support vector machine, random forest classifier, and adaptive boosting, are benchmarked under identical channel and noise conditions. Monte Carlo simulations are performed across varying SNR levels and fading scenarios, enabling a fair assessment of classification accuracy, robustness to residual estimation errors, and relative computational complexity. The results provide practical insights into the strengths and limitations of each classifier in cooperative STBC-OFDM environments, offering valuable guidelines for selecting AMC techniques in future cooperative wireless systems.

Keywords — automatic modulation classification, CFO estimation, CSI, DSTBC-OFDM, higher-order statistics, machine learning

1. Introduction

The design of blind receivers for cooperative relaying (CR) networks remains a challenge due to the combined effects of carrier frequency offset (CFO) and imperfect channel state information (CSI). In distributed space-time block-coded orthogonal frequency division multiplexing (DSTBC-OFDM) systems, these impairments complicate frequency synchronization, channel estimation, and modulation recognition, especially in cooperative scenarios where multiple relay nodes introduce additional CFOs and diverse channel conditions.

Cooperative communications exploit the diversity by employing intermediate relay nodes to form virtual multiple-input multiple-output (MIMO) channels between a source and a destination [1]. In a typical half-duplex relaying set-up, transmission occurs in two phases: the source broadcasts to both destination and relay in the first phase, and in the second phase, the relay forwards the received signal to the destination [2]. Among the various relaying strategies, the amplify-and-forward (AF) technique remains attractive due to its low complexity, while the decode-and-forward (DF) method can provide improved reliability at the expense of higher processing demands [3], [4]. DSTBC is often applied in AF systems to preserve spectral efficiency and improve diversity gains.

Orthogonal frequency division multiplexing (OFDM) is widely adopted for wireless broadband communications due to its robustness against frequency-selective fading and its ability to simplify the equalization process [5]. The combination of DSTBC and OFDM creates a flexible and high-performance framework for cooperative wireless systems, capable of mitigating multipath fading while leveraging spatial diversity. However, achieving these benefits in practice requires an accurate estimation of the CFO and CSI [6], as errors in either parameter directly affect data detection and, consequently, precise classification of the modulation.

Automatic modulation classification (AMC) plays a central role in both civilian and military applications, enabling spectrum monitoring, interference detection, adaptive waveform selection, and electronic warfare countermeasures [7]. AMC techniques are generally classified into maximum likelihood (ML) and feature-based (FB) approaches [8]. Although ML methods can be optimal, they are computationally expensive and require detailed signal models, making FB approaches – especially those relying on higher-order statistics (HOS) – attractive due to their good balance between complexity and robustness [9].

In FB-based AMC, the classifier is as important as the feature set. Popular single-learner models include support vector machines (SVM) and neural networks, while ensemble learning methods such as random forest classifier (RFC) and adaptive boosting (AdaBoost) have shown promise in improving classification performance [10]. However, most AMC studies focus on non-cooperative transmission scenarios, and only limited work has examined cooperative DSTBC-OFDM systems, where CFO estimation, CSI acquisition, and classification must be performed jointly under more challenging conditions [11], [12].

Furthermore, existing cooperative AMC research rarely benchmarks multiple classifier types within the same feature extraction and channel-impairment framework, leaving practitioners with little guidance on optimal classifier selection for this context.

To bridge these gaps, this work introduces an integrated framework that unifies realistic channel estimation, synchronization, and intelligent classification in cooperative DSTBC-OFDM networks. Unlike previous studies that treated estimation and classification separately or relied on idealized assumptions, this approach combines pilot-assisted CFO and CSI estimation with higher-order statistical features to enable blind and robust modulation recognition.

The novelty lies in the systematic benchmarking of single and ensemble classifiers under identical channel and feature conditions. This contribution is not extensively addressed in the prior literature.

This paper fills this gap by conducting a comprehensive benchmarking study of single- and ensemble-learning classifiers for HOS-based AMC in cooperative DSTBC-OFDM systems using AF relaying. A unified simulation framework is developed that integrates pilot-assisted CFO estimation via virtual subcarriers (VSCs), pilot-based CSI estimation, and HOS feature extraction, followed by classification with feedforward neural networks (FFNN), SVM, RFC, and AdaBoost. The evaluation compares classification accuracy, robustness to residual estimation errors, and computational complexity under varying SNR and fading conditions.

Table 1 shows the abbreviations and notations used in this article.

Tab. 1. Used notation.

Upper case bold letters	Matrices
Lower case bold-italic letters	Column vectors
$(\cdot)^*$	Complex conjugate
$(\cdot)^T$	Transpose operator
$(\cdot)^\sharp$	Pseudo-inverse
$(\cdot)^H$	Hermitian transpose
$\otimes$	Kronecker products
$\Re[\cdot]$	Real parts
$\Im[\cdot]$	Imaginary parts
$E\{\cdot\}$	Expected value of a random variable
diag[$\mathbf{x}$]	Diagonal matrix ($\mathbf{x}$ is main diagonal)
$\mathbf{D}_N(\mathbf{h}) =$ diag[$\mathbf{h}$]	Diagonal matrix of size $N \times N$
$\lvert\cdot\rvert$	Modulus of a complex number
$\lVert\cdot\rVert$	Euclidean norm
$\mathrm{tr}(\mathbf{X})$	Trace of the matrix $\mathbf{X}$
$[\mathbf{x}]_m$	m-th entry of the column vector $\mathbf{x}$
$[\mathbf{A}]_{k,m}$	(k,m)-th entry of a matrix $\mathbf{A}$
$\mathbf{I}_N$	Identity matrix of size $N \times N$
$[\mathbf{F}_N]_{m,n}$	FFT matrix of size $N \times N$
$\mathbb{F}_p$	FFT of pilot symbols
$\mathbb{F}_d$	FFT of data symbols

2. System Model

The concept behind the cooperative DSTBC-OFDM technique is to implement a communication process between source S and destination D nodes in a distributed way with the help of communicating nodes R that act as relays [1]. In this paper, we consider that only one antenna occupies the S and R terminals, while one or two receive antennas can be used at the D terminal (Fig. 1).

The source and relay nodes cooperate using the AF protocol based on the architecture of Alamouti STBC-OFDM for data and pilot blocs are given by:

$$\begin{cases} \mathbb{S}_i = \begin{bmatrix} \mathbf{S}_{2i} & \mathbf{S}_{2i+1} \\ -\mathbf{S}^*_{2i+1} & \mathbf{S}^*_{2i} \end{bmatrix} \\ \mathbb{P}_i = \begin{bmatrix} \mathbf{P}_{2i} & \mathbf{P}_{2i+1} \\ -\mathbf{P}^*_{2i+1} & \mathbf{P}^*_{2i} \end{bmatrix} \end{cases}, \tag{1}$$

$$\begin{cases} \mathbf{S}_{2i+m} = \mathrm{diag}(\mathbf{s}_{2i+m}) \\ \mathbf{s}_{2i+m} = \left[s_{2i+m}(0), s_{2i+m}(1), \Phi, s_{2i+1}(N_d-1) \right] \\ \mathbf{P}_{2i+m} = \mathrm{diag}(\mathbf{p}_{2i+m}) \\ \mathbf{p}_{2i+m} = \left[p_{2i+m}(0), p_{2i+m}(1), \Phi, p_{2i+1}(N_d-1) \right] \end{cases}, \tag{2}$$

where S_{2i+m} and P_{2i+m} are the $(2i+m)$-th data and pilot symbols, respectively, and $m \in \{0,1\}$ is the index of two successive data and pilot vectors for the i-th STBC coding matrix.

We consider an OFDM modulation with N subcarriers. Only N_e from N subcarriers are used, the remaining $N_{vsc} = N - N_e$ subcarriers represent the virtual subcarriers (VSCs) [13]. The active subcarriers N_e are divided into two subsets of N_p subcarriers for pilot symbols and N_d subcarriers for information symbols. We define $\mathbf{i}_\beta = \{\beta_1, \ldots, \beta_{N_p}\}$ as a subset of indexes for pilot tones and $\mathbf{i}_\alpha = \{\alpha_1, \ldots, \alpha_{N_d}\}$ as a subset of indexes of subcarriers carrying data symbols.

The insertion of the pilot sub-block $\mathbf{p}_{2i+m}$ into the information sub-block $\mathbf{s}_{2i+m}$ is realized by:

$$\mathbf{z}_{2i+m} = \mathbf{Q}^{\beta}_{N_e}\mathbf{p}_{2i+m} + \mathbf{Q}^{\alpha}_{N_e}\mathbf{s}_{2i+m}\,, \tag{3}$$

where:

$$\begin{cases} \mathbf{Q}^{\alpha}_{N_e} = \mathbf{I}_{N_e}\{\boldsymbol{\alpha}\} = \left[\mathbf{e}^{(\alpha_1)}_{N_e} \mathbf{e}^{(\alpha_2)}_{N_e} \ldots \mathbf{e}^{(\alpha_{N_d})}_{N_e} \right] \\ \mathbf{Q}^{\beta}_{N_e} = \mathbf{I}_{N_e}\{\boldsymbol{\beta}\} = \left[\mathbf{e}^{(\beta_1)}_{N_e} \mathbf{e}^{(\beta_2)}_{N_e} \ldots \mathbf{e}^{(\beta_{N_p})}_{N_e} \right] \end{cases}, \tag{4}$$

are the permutation matrices [5].

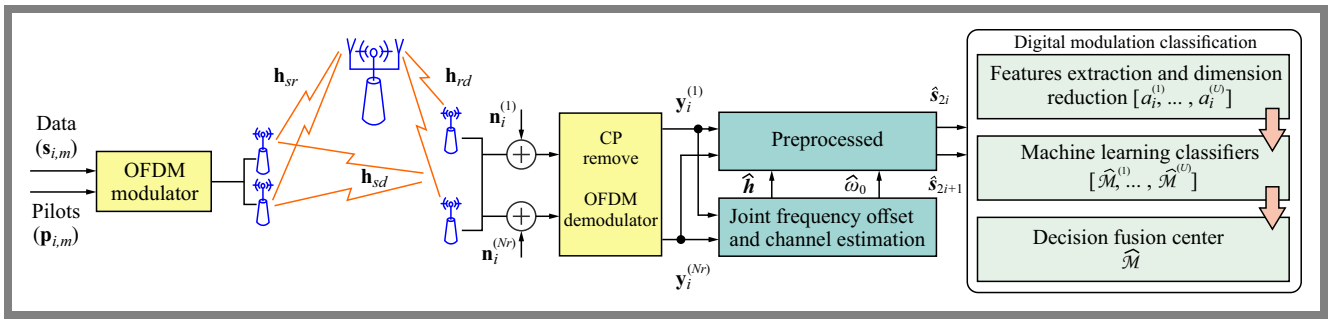

Fig. 1. Baseband equivalent model of the cooperative relaying DSTBC-OFDM system.

N_{vsc} virtual subcarriers are added to $\mathbf{z}_{2i+m}$ to form the vectors $\hat{z}_{2i+m}$. The multiplexed sub-blocks $\hat{\mathbf{z}}_{2i+m}$ are left multiplied by the inverse fast Fourier transform (IFFT) matrix $\mathbf{F}_N^H$ to form the vectors $\widetilde{\mathbf{z}}_{2i+m}$. Before emission, cyclic prefix (CP) of length N_{cp} must be inserted by multiplication with an appropriate matrix $\mathbf{T}_{cp} = [\mathbf{I}^T_{N_{cp}\times N}, \mathbf{I}_N^T]^T$. The i-th transmitted STBC matrix becomes:

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}_{2i}\mathbf{x}_{2i+1} & \\ -\mathbf{x}^*_{2i+1} & \mathbf{x}^*_{2i} \end{bmatrix} = (\mathbf{I}_2 \otimes \mathbf{T}_{cp}) \begin{bmatrix} \tilde{\mathbf{z}}_{2i} & \tilde{\mathbf{z}}_{2i+1} \\ -\tilde{\mathbf{z}}^*_{2i+1} & \tilde{\mathbf{z}}^*_{2i} \end{bmatrix} . \quad (5)$$

The two sub-blocks $[x_{2i}, -\mathbf{x}^*_{2i+1}]$ correspond to symbols conveyed by the source in the time slots 1 and 2, whereas $[\mathbf{x}_{2i+1}, x^*_{2i}]$ refers to the one forwarded by the relay in the time slots 3 and 4.

2.1. DSTBC Design

During the first broadcasting phase, the two sub-blocks $[\mathbf{x}_{2i}, -\mathbf{x}^*_{2i+1}]$ are broadcast to R and D nodes with a transmitted power p_s.

At reception, we remove the cyclic prefix by multiplying each received STBC block by $\mathbf{R}_{cp}$ where $\mathbf{R}_{cp} = [\mathbf{0}_{N\times N_{cp}}, \mathbf{I}_N]$. In this paper, we adopt the model developed in [7] where the CFO between the source and the relay is not considered. As we assume that the CFO is the same for all receive antennas, that is true for the case where CFO is generated by the carrier frequency mismatch between the transmit and the receive oscillators. Next, the signals received by the n_r-th antenna of the destination node and by the unique antenna at the relay node are given by Eqs. (6) and (7).

$$\mathbb{y}^{(n_r)}_{i,sd} = \begin{bmatrix} \mathbf{y}^{n_r}_{i,sd}[1] \\ \mathbf{y}^{n_r}_{i,sd}[2] \end{bmatrix} = \sqrt{p_s}\,\mathbb{D}_N(\omega_0) \cdot \left(\mathbb{F}_p^H \begin{bmatrix} \mathbf{P}_{2i} \\ -\mathbf{P}^*_{2i+1} \end{bmatrix} \mathbf{\Gamma}_p + \mathbb{F}_d^H \begin{bmatrix} \mathbf{S}_{2i} \\ -\mathbf{S}^*_{2i+1} \end{bmatrix} \mathbf{\Gamma}_d \right) \mathbf{h}^{(n_r)}_{i,sd} + \boldsymbol{\eta}^{(n_r)}_{i,sd} \quad (6)$$

$$\mathbb{y}_{i,sr} = \begin{bmatrix} \mathbf{y}_{i,sr}[1] \\ \mathbf{y}_{i,sr}[2] \end{bmatrix} = \sqrt{p_s} \left(\mathbb{F}_p^H \begin{bmatrix} \mathbf{P}_{2i} \\ -\mathbf{P}^*_{2i+1} \end{bmatrix} \mathbf{\Gamma}_p + \mathbb{F}_d^H \begin{bmatrix} \mathbf{S}_{2i} \\ -\mathbf{S}^*_{2i+1} \end{bmatrix} \mathbf{\Gamma}_d \right) \mathbf{h}_{i,sr} + \boldsymbol{\eta}_{i,sr} \quad (7)$$

where:

$$\begin{cases} \mathbb{F}_p^H = \mathbf{I}_2 \otimes \left(\mathbf{F}_N^H \mathbf{Q}_{2N}^{\beta}\right) \\ \mathbb{F}_d^H = \mathbf{I}_2 \otimes \left(\mathbf{F}_N^H \mathbf{Q}_{2N}^{\alpha}\right) \\ \mathbf{\Gamma}_p = [f^L_{\beta_1}, \ldots, f^L_{\beta_{N_p}}]^H \\ \mathbf{\Gamma}_d = [f^L_{\alpha_1}, \ldots, f^L_{\alpha_{N_d}}]^H \\ \mathbf{f}^L_n = [1, e^{j2\pi n/N}, \ldots, e^{j2\pi n(L-1)/N}]^T \end{cases}$$

$$\mathbb{D}_N(\omega_0) = e^{j\omega_0(2iN_q+N_{cp})} \begin{bmatrix} \mathbf{D}_N^{(0)}(\omega_0) & \mathbf{0}_{N\text{x}N} \\ \mathbf{0}_{N\text{x}N} & \mathbf{D}_N^{(1)}(\omega_0) \end{bmatrix} \quad (8)$$

where

$$\mathbf{D}_N^{(u)}(\omega_0) = \text{diag}\left(e^{j(uN_q)\omega_0}, \ldots, e^{j(uN_q+N-1)\omega_0}\right) .$$

$\omega_0 = \Delta\omega_0 T$ is the normalized CFO, $\Delta\omega_0$ is the CFO, and T is the sampling interval and $N_q = N + N_{cp}$.

$$\mathbf{h}^{(n_r)}_{i,sd} = \left[h^{(n_r)}_{i,sd}(0), \ldots, h^{(n_r)}_{i,sd}(L-1)\right]^T$$

and

$$\mathbf{h}_{i,sr} = \left[h_{i,sr}(0), \ldots, h_{i,sr}(L_{sr}-1)\right]^T$$

represent respectively L and L_{sr} taps of channels between source S and the n_r-th receive antenna of the destination node, and the channel between source S and the unique receive antenna at the relay node R.

Vectors $\mathbb{y}^{(n_r)}_{i,sd}$, $\boldsymbol{\eta}^{(n_r)}_{i,sd}$, $\mathbb{y}_{i,sr}$, and $\boldsymbol{\eta}_{i,sr}$ are the received signal and the noise at the n_r-th receive antenna at the destination node D and the received signal and the noise at the relay node R, respectively.

$$\boldsymbol{\eta}^{(n_r)}_{i,sd} = \left[\eta^{(n_r)}_{i,sd}(0), \ldots, \eta^{(n_r)}_{i,sd}(N-1)\right]^T ,$$

and

$$\boldsymbol{\eta}^{(n_r)}_{i,sr} = \left[\eta_{i,sr}(0), \ldots, \eta_{i,sr}(N-1)\right]^T ,$$

are complex additive white Gaussian noise (AWGN) vectors with zero mean and variances $\sigma^2_{sd}\mathbf{I}_N$ and $\sigma^2_{sr}\mathbf{I}_N$, respectively. In the relaying phase, R uses the factor

$$\mathcal{A}_f = \sqrt{p_s \left(p_s \left\|\hat{\mathbf{h}}_{i,sr}\right\|^2 + N_0\right)^{-1}}$$

to adjust the power of the signal received from S, where $\hat{\mathbf{h}}_{sr}$ is the estimated source-relay channel. Then, the second sub-block $[\mathbf{x}_{2i+1}, \mathbf{x}^*_{2i}]$ is generated from the processed versions of received signals at the relay in earlier time slots by adequately applying the conjugation and sign-reversion.

The resulting signals $-\mathcal{A}_f\left[\mathbf{y}_{i,sr}[2]\right]^*$ and $\mathcal{A}_f\left[\mathbf{y}_{i,sr}[1]\right]^*$ are then transmitted to destination node D in time slots 3 and 4, respectively with a power p_r as follows:

$$\begin{aligned}\mathbb{y}_{i,rd}^{(n_r)} &= \begin{bmatrix}\mathbf{y}_{i,rd}^{n_r}[3]\\ \mathbf{y}_{i,rd}^{n_r}[4]\end{bmatrix} = \mathcal{A}_f\sqrt{p_s p_r}\,\mathbb{D}_N(\omega_0)\\ &\cdot\left(\mathbb{F}_p^H\begin{bmatrix}\mathbf{P}_{2i+1}\\ \mathbf{P}_{2i}^*\end{bmatrix}\bar{\mathbf{\Gamma}}_p + \mathbb{F}_d^H\begin{bmatrix}\mathbf{S}_{2i+1}\\ \mathbf{S}_{2i}^*\end{bmatrix}\bar{\mathbf{\Gamma}}_d\right)\mathbf{h}_{i,srd}^{(n_r)} + \mathbb{n}_{i,sd}^{(n_r)}\end{aligned} \tag{9}$$

where

$$\mathbb{n}_{i,srd}^{(n_r)} = \left[(\boldsymbol{\eta}_{2i,srd}^{(n_r)})^T, (\boldsymbol{\eta}_{2i+1,srd}^{(n_r)})^T\right]^T$$

are complex AWGN with $\mathcal{CN}\left(0, (\mathcal{A}_f^2 p_r\sigma_{sr}^2 + \sigma_{rd}^2)\mathbf{I}_N\right)$.

We define the equivalent channel impulse response (CIR) of source-relay-destination ($S-R-D$) links by [12]:

$$\mathbf{h}_{i,srd}^{(n_r)} = \mathrm{conv}(\mathbf{h}_{i,sr}^*, \mathbf{h}_{i,rd}^{(n_r)})\,, \tag{10}$$

where conv means the convolution operation, and $\mathbf{h}_{i,srd}^{(n_r)}$ is $(L_{srd}\times 1)$ vector with $L_{srd} = 2\max(L_{sr}, L_{rd}) - 1$.

$$\bar{\mathbf{\Gamma}}_d = \begin{bmatrix}\bar{\mathbf{f}}_{\alpha_1}^{L_{srd}} & \bar{\mathbf{f}}_{\alpha_2}^{L_{srd}} & \cdots & \bar{\mathbf{f}}_{\alpha_{N_d}}^{L_{srd}}\end{bmatrix}^{\mathcal{H}},$$

$$\bar{\mathbf{\Gamma}}_p = \begin{bmatrix}\bar{\mathbf{f}}_{\beta_1}^{L_{srd}} & \bar{\mathbf{f}}_{\beta_2}^{L_{srd}} & \cdots & \bar{\mathbf{f}}_{\beta_{N_p}}^{L_{srd}}\end{bmatrix}^{\mathcal{H}},$$

are the FFT matrix applied to $\mathbf{h}_{i,srd}^{(n_r)}$, and

$$\begin{aligned}\bar{\mathbf{f}}_n^{L_{srd}} = \big[&e^{j2\pi n(N-\max(L_{sr},L_{rd})+1)/N}, \cdots,\\ &1, \cdots, e^{j2\pi n(\max(L_{sr},L_{rd})-1)/N}\big]^T.\end{aligned}$$

After receiving the signals released from the source and relay nodes at the destination node, we add the signals received at time slots 1 and 3, while applying conjugation before adding the signals received at time slots 2 and 4. The artificially produced signal can be expressed in matrix form as:

$$\begin{aligned}\mathbf{y}_{i,\mathcal{D}}^{(n_r)} &= \begin{bmatrix}\mathbf{y}_{2i,\mathcal{D}}^{(n_r)}\\ \mathbf{y}_{2i+1,\mathcal{D}}^{(n_r)}\end{bmatrix} =\\ \mathbb{D}_N(\omega_0)&\left(\mathbb{F}_p^{\mathcal{H}}\mathbb{P}_i\mathbb{J}_p + \mathbb{F}_d^{\mathcal{H}}\mathbb{S}_i\mathbb{J}_d\right)\begin{bmatrix}\sqrt{p_s}\,\mathbf{h}_{i,sd}^{(n_r)}\\ \mathcal{A}_f\sqrt{p_s p_r}\,\mathbf{h}_{i,srd}^{(n_r)}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\eta}_{2i,\mathcal{D}}^{(n_r)}\\ \boldsymbol{\eta}_{2i+1,\mathcal{D}}^{(n_r)}\end{bmatrix}\end{aligned} \tag{11}$$

$$\begin{cases}\mathbb{J}_p = \begin{bmatrix}\mathbf{\Gamma}_p & \mathbf{0}_{N_p\times L_{srd}}\\ \mathbf{0}_{N_p\times L} & \bar{\mathbf{\Gamma}}_p\end{bmatrix}\\ \mathbb{J}_d = \begin{bmatrix}\mathbf{\Gamma}_d & \mathbf{0}_{N_d\times L_{srd}}\\ \mathbf{0}_{N_d\times L} & \bar{\mathbf{\Gamma}}_d\end{bmatrix}\end{cases} \tag{12}$$

$$\boldsymbol{\eta}_{2i,D}^{(n_r)} = \boldsymbol{\eta}_{2i,sd}^{(n_r)} + \boldsymbol{\eta}_{2i,rd}^{(n_r)}$$

and

$$\boldsymbol{\eta}_{2i+1,D}^{(n_r)} = \boldsymbol{\eta}_{2i+1,sd}^{(n_r)} + \boldsymbol{\eta}_{2i+1,rd}^{(n_r)}$$

are complex AWGN vectors with zero-mean and variance $\sigma_D^2\mathbf{I}_N$ with $\sigma_D^2 = \sigma_{sd}^2 + \mathcal{A}_f^2 p_r\sigma_{sr}^2$.

3. CFO and Channel Estimation

Precise channel and carrier frequency offset (CFO) estimation is essential for reliable data recovery in cooperative DSTBC-OFDM systems, particularly under time-varying channel conditions. The receiver must obtain precise CSI to correctly decode the transmitted signal and apply CFO compensation to mitigate frequency shifts caused by oscillator mismatches or Doppler effects.

A variety of channel estimation methods have been reported for STBC-OFDM systems, broadly categorized into pilot-based and blind approaches. Pilot-based methods embed known training sequences into the transmitted frame, enabling straightforward channel estimation at the receiver. On the contrary, blind methods exploit the statistical properties of the received signal to infer the channel without explicit training [14]. Although blind estimation reduces overhead, it generally requires longer observation intervals and is less robust under severe fading or time-varying conditions.

For cooperative systems, joint estimation of CFO and CSI has been explored in prior works. For example, papers [15]-[19] present a semi-blind joint estimation method for MIMO STBC-OFDM systems. However, such schemes often assume multiple antennas at each node, which may be impractical in many wireless devices due to size and cost constraints. Moreover, propagation conditions may not consistently support full MIMO specifications in real deployments.

In Tab. 2 summary of selected literature on channel and CFO estimation in STBC-OFDM systems is provided.

In the simulation framework of this study, pilot-assisted CSI estimation and a CFO estimation scheme that combines pilot symbols and virtual subcarriers (VSCs) are implemented for AF-based cooperative DSTBC-OFDM. The AF protocol is adopted for its low implementation complexity, despite its inherent noise amplification effect, as it remains competitive in performance compared to decode and forward (DF) under the considered channel conditions.

This estimation approach provides the necessary preprocessing for the subsequent benchmarking of modulation classifiers. By standardizing the CFO and CSI estimation across all classifiers, the evaluation isolates the classification performance from the variability related to the estimation, ensuring a fair and consistent comparison.

By assuming $\hat{\omega}_0 = \omega_0$ in Eq. (8), we can recover $\hat{\mathfrak{h}}^{(n_r)}$ by solving a least-square (LS) problem, as:

$$\hat{\mathfrak{h}}_i^{(n_r)} = e^{-j\hat{\omega}_0(2iN_q+N_{cp})}\left(\mathbb{P}_i\mathbb{J}_p\right)\mathbb{F}_p\mathbb{D}_N^{\mathcal{H}}(\hat{\omega}_0)\mathbb{y}_{i,\mathcal{D}}^{(n_r)}\,. \tag{13}$$

Based on VSCs, the blind CFO estimate is obtained by minimizing the following quadratic cost function:

$$\mathcal{J}^{(n_r)}(\hat{\omega}_0) = \frac{1}{N_b}\sum_{i=0}^{N_b-1}\left\|e^{-j(2iN_q+N_{cp})\hat{\omega}_0}\mathbb{F}_z\mathbb{D}_N^H(\hat{\omega}_0)\mathbb{y}_{i,D}^{(n_r)}\right\|^2 \tag{14}$$

where

$$\begin{cases} \mathbb{F}_z = \begin{bmatrix} \mathbf{F}_z & \mathbf{0}_{N_{vsc}\times N} \\ \mathbf{0}_{N_{vsc}\times N} & \mathbf{F}_z \end{bmatrix} \\ \mathbf{F}_z^{\mathcal{H}} = \begin{bmatrix} \mathbf{f}_{N_e+1}^{\mathcal{N}} & \cdots & \mathbf{f}_N^{\mathcal{N}} \end{bmatrix} \end{cases} . \tag{15}$$

We compensate the received signal by the estimated CFO; then, we perform demodulation using $\mathbb{F}_d$. Next, we use estimated channels to recover the transmitted data employing the zero-forcing equalizer:

$$\mathbb{H}_i^{(n_r)} = \begin{bmatrix} \tilde{\mathrm{H}}_{i,sd}^{(n_r)} & \tilde{\mathrm{H}}_{i,srd}^{(n_r)} \\ \left(\tilde{\mathrm{H}}_{i,srd}^{(n_r)}\right)^* & -\left(\tilde{\mathrm{H}}_{i,sd}^{(n_r)}\right)^* \end{bmatrix} , \tag{16}$$

where

$$\tilde{\mathrm{H}}_{i,sd}^{(n_r)} = \sqrt{p_s}\mathrm{D}_N\left(\tilde{h}_{i,sd}^{(n_r)}\right)$$

and

$$\tilde{\mathrm{H}}_{i,srd}^{(n_r)} = \mathcal{A}_f\sqrt{p_s p_r}\mathrm{D}_N\left(\tilde{h}_{i,srd}^{(n_r)}\right)$$

are the frequency-response of the equivalent channel between the source node and the n_r-th receive antenna, in the direct and indirect link, respectively.

The recovered data symbols are given by:

$$\begin{cases} \hat{\mathbf{s}}_{2i} \\ \hat{\mathbf{s}}_{2i+1} \end{cases} = e^{-j\hat{\omega}_0(2iN_q+N_{cp})} \sum_{n_r=1}^{N_r} \left(\mathbf{H}_i^{(n_r)}\right)^H \begin{bmatrix} \mathbf{F}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_d^* \end{bmatrix} \cdot \begin{bmatrix} \mathbf{D}_N^{(0)}(\hat{\omega}_0) & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{D}_N^{(1)}(\hat{\omega}_0) \end{bmatrix} \begin{bmatrix} \mathbf{y}_{D,2i}^{(n_r)} \\ -\left(\mathbf{y}_{D,2i+1}^{(n_r)}\right)^* \end{bmatrix} . \tag{17}$$

4. Modulation Classification Algorithms

A standard digital modulation classification (DMC) system consists of two main subsystems: feature extraction and classification. The feature extraction subsystem typically includes a pre-processing stage followed by feature selection.

Effective pre-processing – such as CFO compensation and channel equalization – plays a crucial role in improving modulation identification performance. Proper pre-processing facilitates more reliable feature extraction, thus improving classifier accuracy. In contrast, inadequate pre-processing can significantly degrade classification results [14].

Previous studies [9], [13], [14] have demonstrated that higher-order cumulants (HOC) and higher-order moments (HOM) serve as effective features for modulation classification, offering a favorable trade-off between classification performance and computational complexity. In this work, we extract the features comprising HOM up to the eighth order and HOC up to the sixth order to form higher order T statistics (HOS) vectors [28]. These feature vectors are subsequently processed using principal component analysis (PCA) before being input into the classifiers.

The feasibility of deploying a model on real hardware depends largely on selecting trainable parameters from candidate features, as these directly impact computational cost. To address this, PCA is used to compress the original dataset vector D_g into a lower-dimensional dataset D_s, effectively removing redundant and irrelevant information. This reduction in dimensionality enables the selection of an optimal subset of HOS that improves the precision of modulation identification while minimizing computational complexity [29].

The optimized feature subset is then fed into the classifiers to accurately differentiate between various linear modulation schemes.

4.1. Classification Tools

Digital modulation classification (DMC) plays a crucial role in the definition of an intelligent receiver. Numerous studies have explored modulation recognition using either single or ensemble learning algorithms. A single learning algorithm employs one classifier model, whereas ensemble learning combines multiple weaker classifiers to form a more robust predictor.

In this work, we use feedforward neural networks (FFNN) and support vector machines (SVM) as single-learning algorithms. For ensemble learning, we employ random forest classifier (RFC) and AdaBoost, representing bagging and booster methods, respectively.

FFNNs are a popular choice for trainable pattern classification due to their straightforward architecture, where information flows sequentially from one layer to the next [15]. The network consists of three types of layers: input, hidden, and output. The input layer receives external data to initiate pattern recognition. The hidden layers process this information through weighted connections, and the output layer produces the classification result [30].

In this study, FFNN training begins by feeding the selected feature subset a_i into the input layer to predict the corresponding label N_i. After training, performance is evaluated during the testing phase by measuring the probability of correct identification. The final decision is made by selecting modulation type from the candidate pool Θ ={2PSK, 8PSK, 8PAM, 16QAM}, and presented by the label vector θ_i of length equal to cardinal of Θ.

Selection of Θ provides a comprehensive baseline for evaluating the performance of the communication system in different modulation domains. 2PSK, with its simple two-point phase constellation, offers robustness against noise and serves as a clear reference for low-complexity scenarios. 8PSK increases spectral efficiency by encoding three bits per symbol through phase variations, allowing the evaluation of system behavior under moderate complexity. 8PAM, varying only in amplitude, provides insight into amplitude-sensitive impairments and complements phase-based evaluations. Finally, 16QAM combines amplitude and phase to encode four bits per symbol and represents a standard high-throughput modulation widely used in modern systems, making it a realistic baseline for practical applications.

Collectively, these schemes cover phase-only, amplitude-only, and combined amplitude-phase modulations, providing

Tab. 2. Summary of selected literature on channel and CFO estimation in STBC-OFDM systems. The symbol "×" indicates that the survey did not review a classification method.

Ref.	Classification method	System	Contribution
[7]	×	Cooperative STBC OFDM	Analyzes the impact of CFO and channel estimation errors on LS receiver performance; derives closed-form expressions for output SNR evaluation
[20]	×	MIMO OFDM	Proposes a PARAFAC-based blind channel estimation method using recursive least squares tracking, improving estimation accuracy by minimizing uncertainty via weighted least squares cost function
[21]	×	MIMO OFDM	Proposes iterative joint ML CFO and channel estimators for asynchronous cooperative systems, reducing complexity via SAGE-based iterative estimation (SAGE-IE)
[22]	Maximum likelihood-based multi-cumulant classification	MIMO	Proposes a feature-based AMC framework using blind channel estimation and multi-cumulant vectors combining arbitrary orders and lags. Introduces natural ICA and fast ICA methods to improve channel estimation and classification accuracy
[23]	Support vector machines	SISO	Examines the impact of phase offset estimators, channel state information, and noise on classification accuracy
[24]	DNN and ML-based classifiers	SISO	Studies modulation classification by mapping signal samples to posterior probabilities; demonstrates robustness under uncertain channels and noise
[25]	×	SISO OFDM	Proposes a blind CFO estimation method using DCT/IDCT frequency-domain symmetry combined with equalization via banded-matrix approximation to improve system performance
[26]	×	STBC MIMO OFDM	Proposes a Kalman-filter-based channel estimation method combined with STBC and orthogonal pilot sequences to enhance diversity gains and mitigate antenna interference
[27]	×	SISO OFDM	Models joint estimation of CFO, sparse channel, and noise statistics, addressing challenges of sparse signal recovery with unknown CFO and noise variance

a scalable framework for baseline evaluation while enabling future extension to more complex constellations, such as higher-order QAM or adaptive hybrid schemes.

Similarly to the FFNN, SVM can be a global approximator of any multivariate function for an undefined level of accuracy. It is used in DMC due to the efficiency and capacity of treating high-dimensional data with a few needed parameters.

When the pattern recognition assignments are manageable, SVM uses hyperplanes to separate data linearly and with maximal boundaries. When a given class cannot be linearly separated in the input space, SVM converts this input space into a high-dimensional feature space.

Tab. 3. Theoretical values of selected HOM and HOC for various modulation types [30].

Type	M_{60}	M_{61}	M_{63}	C_{60}	C_{61}	C_{63}
2PSK	1	1	1	16	16	16
8PSK	0	0	1	0	0	4
8PAM	3.62	3.62	3.62	7.19	7.19	7.19
16QAM	0	0.38	2.08	0	1.8	1.8

We use the "one-against-all" (OAA) method for the multi-class classification of SVM. Furthermore, we engage cross-validation, a standard form of validation technique usually adopted in the training stage to serve as a performance metric to avoid overfitting [31].

The random forest classifier (RFC) is constructed from a set of learners, and it uses decision trees as the base learners. Each learner votes on the class labels to make a final prediction.

The property of being an aggregation of multiple learners enables RFC to be prominent over conventional classification trees [17]. AdaBoost is designed to transform weak learners – in this paper, decision trees – into strong ones, to reach the highest accuracy level [18]. Different learners are trained sequentially by including a new learner per cycle to compensate for inaccuracies created by former learners.

When the cycle ends, the low weight samples acquired are the ones that the weak learner H_t had accurately classified, but the misclassified examples are identified and emphasized

to be fed back into the beginning of the subsequent round, then the new learner H_t is trained.

The previous procedure extends for several rounds. Eventually, AdaBoost forms an ultimate hypothesis using weighted votes to combine all the weak learners [19].

5. Simulation Results

A set of experiments was carried out to assess the performance of the proposed method employing Monte Carlo simulations. We evaluate the performance of CFO and channel estimators by calculating the bit error rate (BER) of CFO, the mean square errors (MSE) of CFO and channels. The MSEs of CFO and channel estimators are defined respectively by:

$$\mathrm{MSE}_{\mathrm{CFO}} = \frac{1}{M_c}\sum_{j=1}^{M_c}\left|\hat{\omega}_0^j - \omega_0\right|^2, \tag{18}$$

$$\mathrm{MSE}_{\mathrm{Channel}} = \frac{1}{M_c N_r}\sum_{j=1}^{M_c}\sum_{n_r=1}^{N_r}\left\|\hat{\mathbf{h}}_j^{(n_r)} - \mathbf{h}^{(n_r)}\right\|^2. \tag{19}$$

The proposed approach was validated for the digital modulation pool Θ=2PSK, 8PSK, 8PAM, 16QAM. 1000 Monte Carlo trials per modulation were carried out for each SNR value.

We consider the SNR across source-relay link SNR_{sr} and the SNR across relay–destination links SNR_{rd} is $\mathrm{SNR} = \mathrm{SNR}_{rd} = \mathrm{SNR}_{sr}$.

We generated a random massage and $L = L_{sr} = L_{rd}$ Rayleigh fading taps channel from i.i.d. zero-mean independent complex Gaussian random variables with the same variance in each run.

We calculate the average classification accuracy for different values of SNR to analyze the performance of the proposed classifier.

Simulations are performed considering an OFDM system with $N = 64$ and 512 subcarriers, and with a cyclic prefix of length $N_{cp} = L + 1$. We consider a single relay for all the experiments and the STBC encoder type Alamouti with (2×1) $(N_r = 1)$ and (2×2) $(N_r = 2)$ antenna configurations are inspected.

Regarding the MUSIC-based CFO estimator for the cooperative DSTBC-OFDM system, we note that it has a complexity order of $O(12N^2N_bN_rG)$, where G is the number of CFO candidates, $\hat{\omega}_0$ in the search grid.

We employ the parameters specified in Tab. 4 to confirm the highest integrity and consistency of the classifier comparison.

5.1. Performance of CFO and Channels Estimation

A set of experiments was conducted to compare the performance between the estimated CFO and channels and the exact CFO and channels for the cooperative DSTBC-OFDM system. Figure 2 shows the BER versus SNR. Pool Θ contains the modulation types for the comparison. The source and relay nodes are occupied by one antenna at each, while $N_r = 2$ receive antennas occupy the destination node.

Tab. 4. Simulation parameters.

Parameter	Value
Dataset	D_g
Modulation pool	4 digital modulations: Θ={2PSK, 8PSK, 8PAM, 16QAM}
Number of data subcarriers	$\begin{cases} N_d = 51 \text{ for } N = 64 \\ N_d = 443 \text{ for } N = 512 \end{cases}$
Number of pilots subcarriers	$\begin{cases} N_p = 8 \text{ for } N = 64 \\ N_p = 64 \text{ for } N = 512) \end{cases}$
Number of virtual subcarriers	$N_{vsc} = 5$
Signal format	In-phase and quadrature (IQ)
Signal dimension	$2 \times 64/2 \times 512$ per time-slot
SNR range	$[-5, 0, \ldots, 15, 20]$ dB
CFO range	[0, 0.1]
CFO candidate	$0.02\,\pi$
Monte Carlo trials	$M_c = 1000$ for each modulation type

The performance of the system using the proposed estimated CFO and channel parameters closely matches that achieved with exact CFO and perfect channel knowledge, demonstrating the reliability of the estimators. For modulation schemes such as 2PSK, 8PSK, 8PAM, and 16QAM, the performance difference between the proposed estimation method and the ideal case is negligible. However, for 16QAM, a relatively larger performance gap is observed, likely due to the increased constellation size, which typically leads to greater system degradation.

5.2. MSE of CFO and Channels Estimators

To evaluate the reliability of the proposed CFO and channel estimation methods, Figs. 3a and 3b present the MSE of the CFO and channel estimates plotted against the SNR, respectively.

The results indicate that changing the modulation scheme does not significantly impact the CFO MSE. In contrast, the accuracy of the CFO estimation is consistent, as the SNR increases for all types of modulation types within the set Θ.

Figure 3b confirms that the proposed channel estimator performs better with increasing SNR.

5.3. Robustness Tests

This subsection evaluates the robustness of the proposed estimation technique by examining the classification performance of various classifiers combined with principal component analysis (PCA).

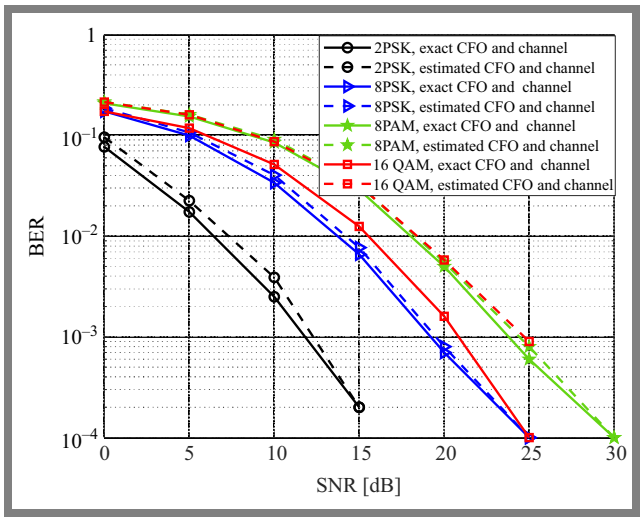

Fig. 2. BER versus SNR of the cooperative DSTBC-OFDM system in the presence of CFO and channel estimation for $N_r = 2$ and $N = 512$.

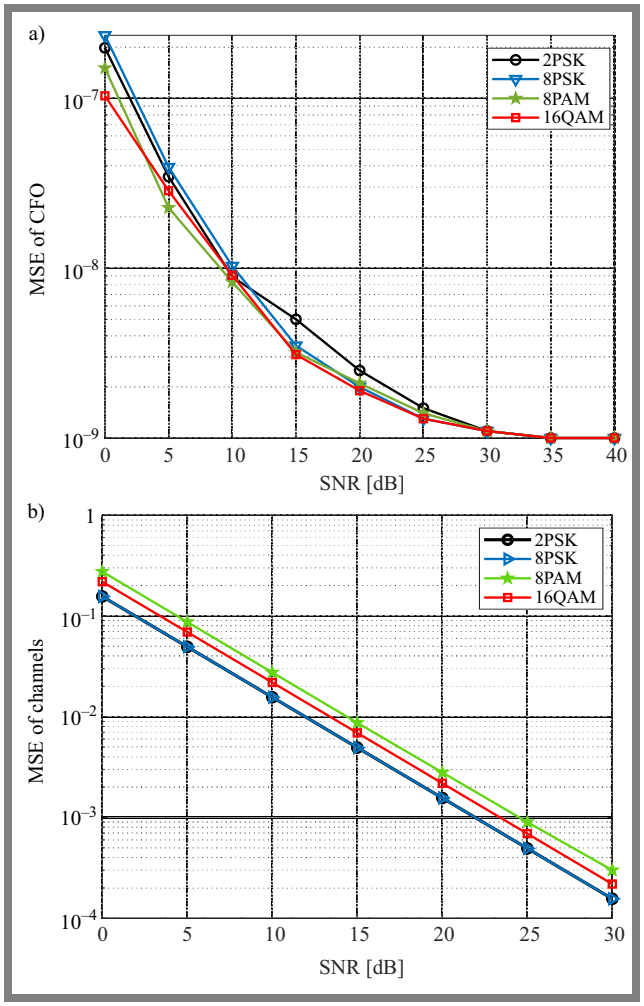

Fig. 3. MSE of the CFO estimator a) MSE of channel estimator b) versus SNR of cooperative DSTBC-OFDM system for $N = 512$ and $N_r = 2$.

The study investigates the impact of the size of the OFDM symbol N and the number of receiving antennas N_r on classification accuracy. The parameters used are summarized in Tab. 4. Specifically, OFDM symbol sizes of N= 64 and 512 are considered, alongside two antenna configurations: 2×1 and 2×2.

Figure 4a presents the average classification accuracy of different classifiers using estimated CFO and channel parameters, with the number of receiving antennas fixed at N_r= 2. An increase in classification accuracy is observed as N increases from 64 to 512, attributed to improved time diversity of the received signals that benefits all classifiers. At N= 64, all classifiers perform similarly, achieving an average accuracy of 92% at SNR = 10 dB. For SNR 15 dB, accuracy rises to approximately 98%.

When N= 512, FFNN and RFC outperform other classifiers, reaching 66% classification accuracy at SNR = 0 dB. SVM and AdaBoost follow closely, with average accuracies of 60% and 58%, respectively. At SNR = 5 dB, all classifiers have 94% accuracy, and achieve 100% accuracy for SNR $\geqslant$ 10 dB.

Figure 4b illustrates the probability of correct identification versus SNR considering 2×1 and 2×2 antenna configurations with N=512. All classifiers show stable performance, with classification accuracy improving as N_r increases from 1 to 2. For example, at SNR = 5 dB and N_r=2, all classifiers reach 94% accuracy. In contrast, at the same SNR with N_r= 1, RFC and SVM achieve 62% accuracy, while AdaBoost and FFNN outperform them with 84% and 78% accuracy, respectively. For N_r= 2 and SNR $\geqslant$ 10 dB, all classifiers achieve 100% accuracy.

At N_r= 1 and SNR = 10 dB, AdaBoost and FFNN perform similarly, achieving 100% accuracy, whereas SVM and RFC attain 90% and 88%, respectively. For SNR $\geqslant$ 15 dB, all classifiers reach 98% accuracy.

Such results confirm that high classification accuracy can be achieved even at low SNR levels. These results reflect the effectiveness of the proposed CFO and channel estimation techniques. Among the classifiers tested, FFNN and AdaBoost consistently deliver superior performance at SNR $\geqslant$ 10 dB, reaching 100% accuracy regardless of antenna configuration and OFDM symbol length.

5.4. Performance Comparison and Evaluation

Figure 5 compares the average classification precision achieved using the proposed CFO and channel estimation method against the ideal case with exact CFO and perfect channel knowledge. The comparison is performed for classifiers under the parameters listed in Tab. 4, with an OFDM symbol size $N = 512$ and the number of receive antennas $N_r = 2$.

The results demonstrate that the proposed estimation method achieves a classification accuracy that closely matches the ideal case at SNR = 5 dB and attains near-perfect accuracy for SNR values greater than 5 dB. This confirms the robustness across all classifiers under low SNR conditions.

At SNR = 5 dB, the benchmarking shows that among FFNN, RFC, SVM, and AdaBoost, FFNN provides the highest accuracy and robustness. At SNR = 0 dB, FFNN and RFC

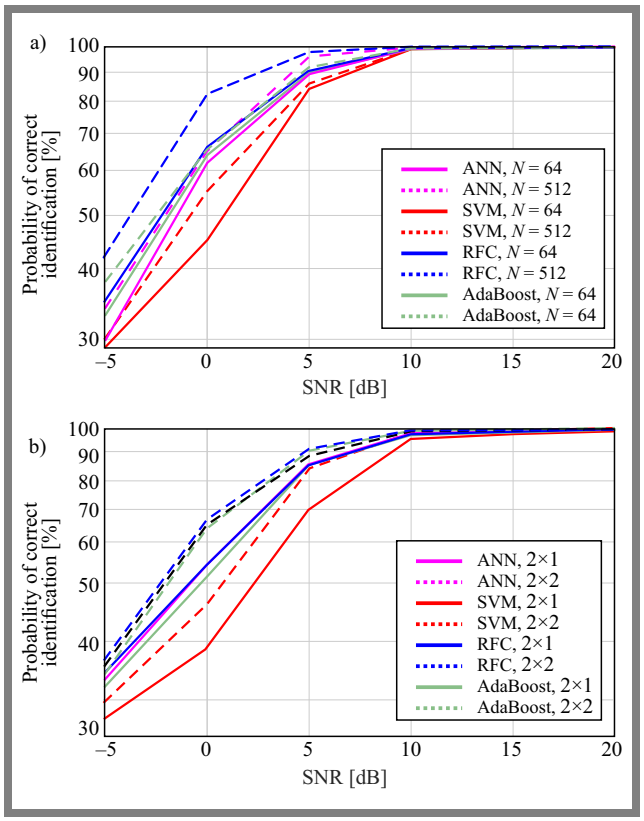

Fig. 4. Classification performance vs. SNR with estimated CFO and channels for: $N_r = 2$, OFDM symbol length $N = 512, 64$ a) and $N_r = 1, 2$, $N = 512$ b).

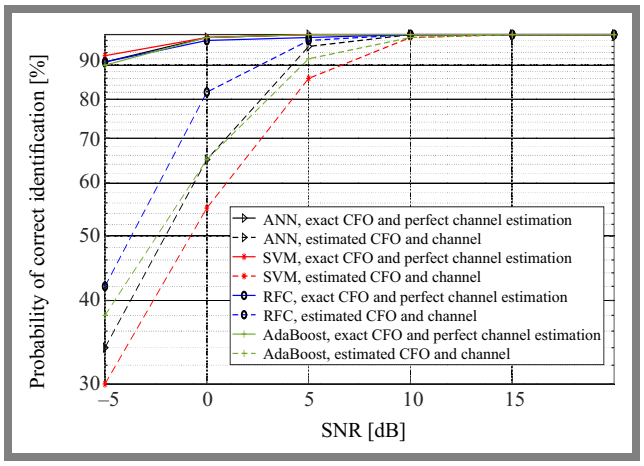

Fig. 5. Classification performance vs. SNR system with estimated CFO and channel and with exact CFO and perfect channel estimated for $N = 512$ and $N_r = 2$.

outperform the other classifiers in distinguishing between different linear modulation schemes.

In general, this benchmarking study reveals that the proposed CFO and channel estimation technique significantly improves the performance of the system in cooperative DSTBC-OFDM environments, reaching near-ideal classification accuracy at SNR 5 dB. Furthermore, FFNN consistently provides superior classification performance in low SNR regimes (below 0 dB), highlighting its suitability for poor channel conditions.

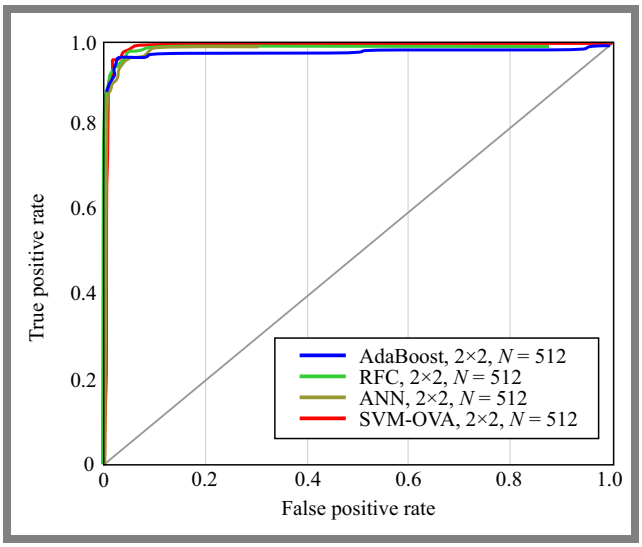

Fig. 6. ROC curves of the ANN, SVM, RFC and AdaBoost classifiers using DSTBC-OFDM with exact CFO and perfect channel estimated for $N = 512$ and $N_r = 2$.

In Fig. 6, the receiver operating characteristic (ROC) curves for AdaBoost, RFC, ANN, and support vector machine using one-versus-all (SVM-OVA) are presented for the 2×2 classification scenario with N= 512 samples. ROC curves are generated using the average posterior probabilities computed in all SNR values ranging from –5 to 20 dB, thereby reflecting the aggregated detection performance under varying noise conditions.

All four classifiers demonstrate strong discriminative capability, as their ROC curves remain close to the upper left region of the plot. AdaBoost, RFC, and ANN exhibit nearly overlapping curves, indicating comparable and consistently high classification performance over the full SNR range. The SVM-OVA classifier achieves a similar overall trend, although minor deviations appear in mid-range false-positive regions.

The close proximity of the curves to the ideal operating point (0,1) suggests that averaging over SNR does not significantly degrade model performance, and all classifiers maintain high true-positive rates with minimal false-positive rates across noise levels. These results confirm that the ensemble-based methods and the neural network provide robust performance under varying SNR conditions, while SVM-OVA remains competitive with slightly higher variability.

6. Conclusions

This paper presented a comprehensive benchmarking study of blind modulation classification in cooperative DSTBC-OFDM systems under realistic impairments, including carrier frequency offset (CFO) and imperfect channel state information (CSI). A hybrid estimation approach combining pilot symbols and virtual subcarriers (VSCs) for CFO estimation with pilot-assisted CSI estimation was developed, leading to improved synchronization accuracy and system reliability.

Using higher-order statistics (HOS) as features, several machine learning classifiers – feedforward neural networks (FFNN), support vector machines (SVM), random forest classifier (RFC) and adaptive boosting (AdaBoost) – were evaluated under varying SNR and fading conditions. Results show that while the proposed estimation scheme effectively mitigates CFO and channel distortions, classifier performance differs across scenarios, with FFNN and AdaBoost achieving the highest accuracy, particularly under low SNR conditions.

References

[1] H. Tayakout, I. Dayoub, K. Ghanem, and H. Bousbia-Salah, "Automatic Modulation Classification for D-STBC Cooperative Relaying Networks", *IEEE Wireless Communications Letters*, vol. 7, pp. 780–783, 2018 (https://doi.org/10.1109/LWC.2018.2824813).

[2] G. Ryu, D. Jang, U. Jeong, and K. Ko, "BER Performance Analysis of Orthogonal Space-time Block Codes in Cooperative MIMO DF Relaying Networks", *IEEE International Conference on Communications (ICC)*, Kansas City, USA, 2018 (https://doi.org/10.1109/ICC.2018.8423038).

[3] W. Swasdio, C. Pirak, S. Jitapunkul, and G. Ascheid, "Alamouti-coded Decode-and-forward Protocol with Optimum Relay Selection and Power Allocation for Cooperative Communications", *Journal of Wireless Communications and Networking*, vol. 2014, art. no. 112, 2014 (https://doi.org/10.1186/1687-1499-2014-112).

[4] A. Abdaoui, S.S. Ikki, and M.H. Ahmed, "Performance Analysis of MIMO Cooperative Relaying System Based on Alamouti STBC and Amplify-and-forward Schemes", *IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, 2010 (https://doi.org/10.1109/ICC.2010.5501917).

[5] S. Yiu, D. Calin, O. Kaya, and K. Yang, "Distributed STBC-OFDM and Distributed SFBC-OFDM for Frequency-selective and Time-varying Channels", *IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, 2012 (https://doi.org/10.1109/WCNC.2012.6214222).

[6] E. Chenguttuvan, L.P. Karuppiah, and K. Sakthisudhan, "Estimating Time and Frequency Under Imperfect Channel Knowledge Using ECM and SAGE Algorithms in Multi-relay Cooperative Networks", *Journal of Wireless Communications and Networking*, vol. 2025, art. no. 1, 2025 (https://doi.org/10.1186/s13638-024-02418-9).

[7] T. Lin and F. Hwang, "Analysis and Design of Joint CFO/Channel Estimate Techniques for a Cooperative STBC-OFDM System", *International Journal of Communication Systems*, vol. 32, art. no. e3845, 2018 (https://doi.org/10.1002/dac.3845).

[8] M. Besseghier *et al.*, "Enhanced Estimation of Channel and CFO in FBMC/OQAM via ZFBMC-based Preamble", *Wireless Personal Communications*, vol. 139, pp. 1815–1836, 2024 (https://doi.org/10.1007/s11277-024-11701-3).

[9] K. Hassan *et al.*, "Blind Digital Modulation Identification for Spatially Correlated MIMO Systems", *IEEE Transactions on Wireless Communications*, vol. 11, pp. 683–693, 2012 (https://doi.org/10.1109/TWC.2011.122211.110236).

[10] B. Xu *et al.*, "Towards Explainability for AI-based Edge Wireless Signal Automatic Modulation Classification", *Journal of Cloud Computing*, vol. 13, art. no. 10, 2024 (https://doi.org/10.1186/s13677-024-00590-3).

[11] K. Akhilesh and K. Vinay, "A Review of Diverse MIMO Antennas Design for Cognitive Radio Applications", *AEU – International Journal of Electronics and Communications*, vol. 200, art. no. 155930, 2025 (https://doi.org/10.1016/j.aeue.2025.155930).

[12] M. Besseghier, A.B. Djebbar, A. Zouggaret, and I. Dayoub, "Joint Channel Estimation and Data Detection for OFDM Based Cooperative System", *Telecommunication Systems*, vol. 73, pp. 545–556, 2019 (https://doi.org/10.1007/s11235-019-00622-3).

[13] Q. Xiao *et al.*, "Research on OFDM Modulation Recognition Method Based on High-order Cyclic Cumulants and Neural Networks", *IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, pp. 712–716, 2024 (https://doi.org/10.1109/ICPICS62053.2024.10796030).

[14] M. Ghogho and A. Swami, "Semi-blind Frequency Offset Synchronization for OFDM", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, USA, 2002 (https://doi.org/10.1109/ICASSP.2002.5745113).

[15] X. Ma, M.K. Oh, G.B. Giannakis, and D.J. Park, "Hopping Pilots for Estimation of Frequency-offset and Multiantenna Channels in MIMO-OFDM", *IEEE Transactions on Communications*, vol. 53, pp. 162–172, 2005 (https://doi.org/10.1109/TCOMM.2004.840663).

[16] R.N. Yang, W.T. Zhang, and S.T. Lou, "Joint Adaptive Blind Channel Estimation and Data Detection for MIMO-OFDM Systems", *Wireless Communications and Mobile Computing*, vol. 2020, art. no. 2508130, 2020 (https://doi.org/10.1155/2020/2508130).

[17] B. Dehri, M. Besseghier, A.B. Djebbar, and I. Dayoub, "Blind Digital Modulation Classification for STBC-OFDM System in Presence of CFO and Channels Estimation Errors", *IET Communications*, vol. 13, pp. 2827–2831, 2019 (https://doi.org/10.1049/iet-com.2019.0362).

[18] T. Liu and S. Zhu, "Joint CFO and Channel Estimation for Asynchronous Cooperative Communication Systems", *IEEE Signal Processing Letters*, vol. 19, pp. 643–646, 2012 (https://doi.org/10.1109/LSP.2012.2210039).

[19] S. Huang *et al.*, "Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants", *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 6089–6101, 2017 (https://doi.org/10.1109/TVT.2016.2636324).

[20] K. Ramadan, M.I. Dessouky, and F.E. El-Samie, "Joint Equalization and CFO Compensation for Performance Enhancement of MIMO-OFDM Communication Systems Using Different Transforms with Banded-matrix Approximation", *AEU – International Journal of Electronics and Communications*, vol. 119, art. no. 153157, 2020 (https://doi.org/10.1016/j.aeue.2020.153157).

[21] R. Tang, X. Zhou, and C. Wang, "Kalman Filter Channel Estimation in 2x2 and 4x4 STBC MIMO-OFDM Systems", *IEEE Access*, vol. 8, pp. 189089–189105, 2020 (https://doi.org/10.1109/ACCESS.2020.3027377).

[22] S. Salari and F. Chan, "Joint CFO and Channel Estimation in OFDM Systems Using Sparse Bayesian Learning", *IEEE Communications Letters*, vol. 25, pp. 166–170, 2021 (https://doi.org/10.1109/LCOMM.2020.3024817).

[23] J. Jagannath, N. Polosky, and D. O'Connor, "Artificial Neural Network Based Automatic Modulation Classification over a Software Defined Radio Testbed", *IEEE International Conference on Communications (ICC)*, Kansas City, USA, 2018 (https://doi.org/10.1109/ICC.2018.8422346).

[24] S. Huang *et al.*, "Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants", *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 6089–6101, 2017 (https://doi.org/10.1109/TVT.2016.2636324).

[25] A. Swami and B. Sadler, "Hierarchical Digital Modulation Classification Using Cumulants", *IEEE Transactions on Communications*, vol. 48, pp. 416–429, 2000 (https://doi.org/10.1109/26.837045).

[26] K.A. Ahmed and E. Ergun, "Automatic Modulation Classification Using Different Neural Network and PCA Combinations", *Expert Systems with Applications*, vol. 175, art. no. 114931, 2021 (https://doi.org/10.1016/j.eswa.2021.114931).

[27] I. Klyueva, "Improving Quality of the Multiclass SVM Classification Based on the Feature Engineering", *1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA)*, Lipetsk, Russia, pp. 491–494, 2019 (https://doi.org/10.1109/SUMMA48161.2019.8947599).

[28] O.P. Awe, A. Deligiannis, and S. Lambotharа, "Spatio-temporal Spectrum Sensing in Cognitive Radio Networks Using Beamformer-

aided SVM Algorithms", *IEEE Access*, vol. 6, pp. 25377–25388, 2018 (https://doi.org/10.1109/ACCESS.2018.2825603).

[29] K. Triantafyllakis, M. Surligas, and G. Vardakis, "Phasma: An Automatic Modulation Classification System Based on Random Forest", *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Piscataway, USA, 2017 (https://doi.org/10.1109/DySPAN.2017.7920749).

[30] S. Yuan *et al.*, "Efficient and Privacy-preserving Outsourcing of Gradient Boosting Decision Tree Inference", *IEEE Transactions on Services Computing*, vol. 17, pp. 2334–2348, 2024 (https://doi.org/10.1109/TSC.2024.3395928).

[31] Y. Zhou *et al.*, "A Modulation Recognition Method Based on Bispectrum and Ensemble Learning", *2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Nanjing, China, pp. 124–128, 2022 (https://doi.org/10.1109/CEI57409.2022.9950111).

Brahim Dehri, Ph.D.
Telecommunications and Digital Signal Processing Laboratory
https://orcid.org/0000-0002-9106-6931
E-mail: dehri.brahim@cuniv-naama.dz
University Centre of Naama, Naama, Algeria
https://www.cuniv-naama.dz

Hakima Moulay, Ph.D.
Telecommunications and Digital Signal Processing Laboratory
https://orcid.org/0000-0002-5637-3944
E-mail: moulay.hakima@cuniv-naama.dz
University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria
https://www.univ-sba.dz/en

Ahmed Amine Daikh, Ph.D.
LIS2T Laboratory
https://orcid.org/0000-0002-4666-2750
E-mail: aadaikh@cuniv-naama.dz
University Centre of Naama, Naama, Algeria
https://www.cuniv-naama.dz
University of Mustapha Stambouli, Mascara, Algeria
https://www.univ-mascara.dz

Mokhtar Besseghier, Ph.D.
LIS2T Laboratory
https://orcid.org/0000-0003-3522-1332
E-mail: m.besseghier@univ-mascara.dz
University of Mustapha Stambouli, Mascara, Algeria
https://www.univ-mascara.dz

Priority-aware Radio Resource Scheduling for mMTC in 5G Networks – Balancing Efficiency and Fairness

Prashant Kumar Baheti and Ajay Khunteta

Poornima University, Jaipur, Rajasthan, India

https://doi.org/10.26636/jtit.2026.1.2336

Abstract — Efficient and fair resource allocation for massive machine-type communication remains a significant challenge in 5G New Radio networks due to the diverse quality of service requirements and dynamic traffic patterns. This paper proposes a priority-aware uplink scheduling (PAUS) algorithm that jointly considers channel quality, 5G QoS identifier, packet aging, and fairness in physical resource block allocation, while simultaneously mitigating starvation of low-priority user equipment. The algorithm utilizes a composite fitness function to implement binary integer optimization for uplink scheduling, supported by heuristic resource assignment to ensure scalability. Simulation results demonstrate that the PAUS algorithm achieves an improved balance between throughput, resource utilization, delay, priority satisfaction, and fairness compared to baseline schedulers with polynomial-time complexity.

Keywords — 5G networks, efficiency, fairness, priority-aware scheduling, quality of service, radio resource scheduling

1. Introduction

The advent of 5G New Radio networks has fundamentally transformed wireless communications [1]. The International Telecommunication Union radiocommunication sector (ITU-R) introduced its 5G vision in 2015 through ITU-R recommendation M.2083, emphasizing that 5G NR would expand beyond enhanced mobile broadband (eMBB) to include massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC), each with stringent and diverse requirements [2].

According to [3], mMTC is characterized by low cost and low complexity of user equipment (UE), offering extended battery life, sporadic small data transmissions and relaxed latency constraints compared to URLLC applications. Different mMTC applications have different priority and delay tolerance characteristics, creating heterogeneous service requirements that pose unique challenges to radio resource allocation. 5G NR networks must ensure high throughput, massive connectivity, and optimal support for latency-tolerant and mission-critical applications [4].

Machine-type communication (MTC) represents a paradigm shift from traditional wireless communications. The integration of MTC into 5G NR networks introduces fundamental changes in traffic patterns [5], resource utilization, and scheduling requirements compared to previous generations. Most MTC traffic is uplink driven, with UEs periodically or sporadically transmitting data, status updates, or alarm notifications.

The 5G NR uplink framework introduces several key enhancements capable of addressing mMTC requirements. Flexible system-supporting subcarrier spacings equaling from 15 to 240 kHz may enable the allocation of resources for distinct mMTC service classes. Additionally, the bandwidth part (BWP) configuration allows for dynamic spectrum allocation to meet specific mMTC application requirements [6].

Despite these improvements, mMTC uplink scheduling in 5G NR faces several challenges, for example extreme heterogeneity in priority requirements, bursty traffic patterns with varying delay sensitivities, energy constrained UEs, massive connectivity scenarios that require scalable scheduling algorithms for high number of UEs, and the necessity to maintain long-term fairness to balance fair resource distribution with adherence to application priorities [7].

Traditional uplink scheduling algorithms such as proportional fair (PF) best channel quality (BestCQI) were optimized primarily for LTE communication, with relatively uniform traffic patterns and QoS requirements. In contrast, mMTC scenarios introduce diverse QoS requirements and varying UE capabilities across service classes, necessitating priority-aware scheduling mechanisms with dynamic physical resource block (PRB) assignment [8].

The scheduling problem also intensifies under high UE densities, due to resource shortages and the need to simultaneously support grant-based and grant-free transmission modes. Existing scheduling algorithms often fail to optimally integrate all relevant QoS indicators into effective PRB allocation decisions, and their performance is degraded under massive connectivity and heterogeneous service requirements [9].

Therefore, 5G NR networks introduce complexity to uplink PRB scheduling, especially since they converge diverse service requirements, flexible frame structures, and massive UE quality. Efficient and fair scheduling solutions are essential to realize the potential of 5G NR for mMTC applications – from low-priority environmental sensors to mission-critical automation systems.

Tab. 1. Comparison of key papers on priority-aware uplink scheduling for mMTC in 5G.

Ref.	Methodology	Traffic types	Contribution	Results
[24]	Heuristic and DRL	Heterogeneous mMTC	Adaptive GF resource partitioning	Improved fairness, lower packet drop rate, better resource utilization
[25]	Simulation-based	Smart city mMTC	Fairness and priority-based scheduling	Outperforms PF and BestCQI in terms of fairness and utilization
[26]	CTMC queuing model	eMBB, URLLC, mMTC	Priority-based allocation in C-RAN slicing	Higher resource utilization, reduced forced termination for mMTC
[27]	Dynamic priority assignment	Multi-RAT MTC	Two-stage dynamic priority scheduling	20 – 30% improvement in outage/success probability
[28]	Mathematical modeling	eMBB, URLLC, mMTC	Mixed reservation/priority RAN slicing	Up to 95% resource utilization, improved isolation

This work addresses these challenges by proposing a priority-aware uplink radio resource scheduling framework for mMTC scenarios that optimally balances efficiency and fairness, adheres to 5G NR standards, and addresses the unique QoS requirements of dense mMTC deployments.

1.1. Literature Review

The author of [10] proposes a 5G scheduler to minimize the long-term average age of information (AoI) for dense urban mMTC scenarios in which UEs frequently connect and disconnect, and adapts per-slot scheduling decisions with a priority algorithm to meet delay needs and reduce the average AoI by 10%, the deadline violation rate by 40%, and the consecutive violation rate by 20%. Similarly, [11] proposes a service-based M2M scheduling approach that integrates UE priority, service type, and channel quality to optimize throughput and resource utilization.

The studies also focus on deep reinforcement learning (DRL)-based research. Dynamic DRL-based slicing selects optimal slices under fluctuating demands while minimizing violations of service level agreements and allowing virtual network function migration for resource contention scenarios [12]. DRL-based resource slicing for multiservice coexistence provides near-optimal decoding and sum-rate performance according to [13]. In vehicular networks, slicing with QoS support prioritizes emergency traffic and maintains optimal throughput and reliability under road-specific conditions [14].

Paper [15] focused on traffic modeling for mMTC and revealed that despite growing UE numbers, aggregated inter-arrival times converge to a Poisson process, suggesting predictable scheduling opportunities even in massive deployments. For mMTC based on DECT-2020 NR, the MAC-layer design for power control and relay selection affects the packet delivery success, improving it by up to 10% [16].

The authors [17] propose a priority-enabled grant-free (GF) transmission access scheme that dynamically allocates slots within each 5G NR subframe, giving first access while assigning remaining slots to low-priority traffic, and develop a two-dimensional Markov chain model that accounts for both types of traffic. Furthermore, paper [18] proposes a scheduling mechanism to optimize resource allocation by addressing the trade-off between throughput, energy efficiency, and fairness through dynamic allocation of radio resources.

In [19], a fairness-aware uplink resource allocation scheme and an optimization problem are presented to ensure QoS satisfaction, using a demand-oriented greedy algorithm for PRB assignment and a bisection method for power allocation in UE, achieving performance improvements of up to 44.3% in fairness and 19.17% in QoS satisfaction. Furthermore, [20] proposes the min–max rate (MMR) algorithm to improve throughput and fairness compared to standard scheduling algorithms. In the same context, article [21] proposes an uplink scheduler to address spatial and frequency domain coupling, while achieving superior performance with polynomial-time complexity.

The authors of [22] address the PRB allocation problem as a mixed integer linear program (MILP), which utilizes mixed numerology, latency, and throughput requirements, multiple slices per UE, and internumerology interference, and propose a solution based on DRL. Also, article [23] proposes two schedulers based on the shortest job first (SJF) principle for minimum latency, and another for the joint optimization of energy efficiency (EE) and latency for complex multicellular scenarios.

Table 1 presents a comparison of key studies focusing on priority-based uplink scheduling for mMTC in 5G networks. The review demonstrates that priority-aware uplink scheduling is essential for supporting diverse mMTC requirements. Although these schemes are capable of reducing latency as well as improving resource utilization and fairness, the trade-off between prioritizing critical traffic and maintaining fairness for all UEs remains a challenge, particularly in congested scenarios with mixed traffic types and delay constraints.

1.2. Research Motivation and Contribution

The increase in the number of mMTC UEs requires optimal resource scheduling solutions capable of addressing diverse QoS constraints, latency budgets, and dense deployment challenges. In 5G NR, differentiated services enabled by 5QI require scheduling frameworks that ensure fast transmission

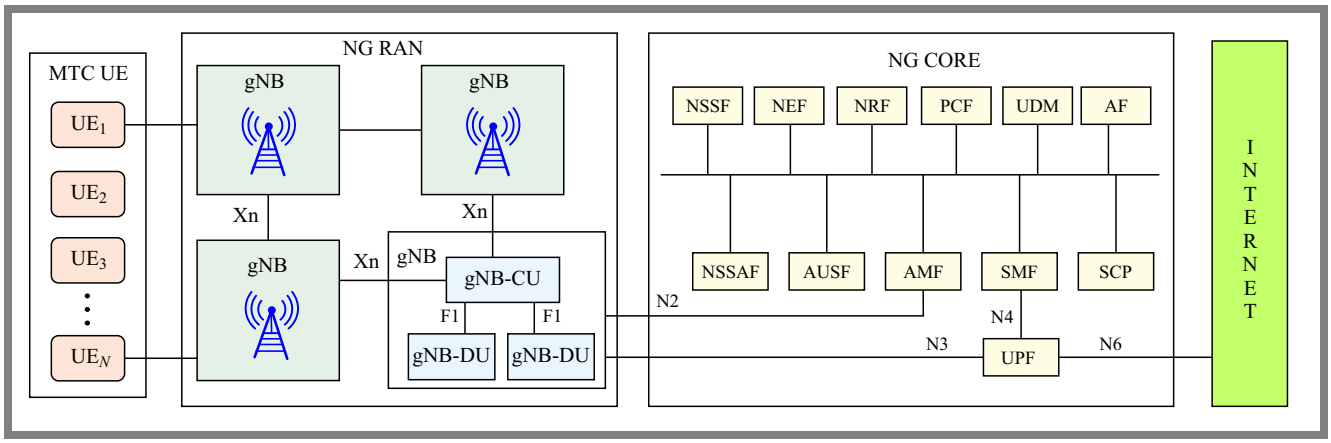

Fig. 1. 5G NR architecture components and interfaces [29].

for priority traffic, while maintaining fair resource sharing among heterogeneous UEs. Overcoming traditional scheduling limitations, such as the starvation of low-priority UEs and inefficient PRB utilization, requires optimal scheduling algorithms that explicitly balance efficiency with fairness.

The contributions of this work are as follows:

- It proposes a dynamic fitness function that integrates instantaneous channel quality, 5QI priorities, packet aging, and fairness adjustment in uplink scheduling decisions.
- It formulates the scheduling challenge as a constrained binary integer program, solved with greedy heuristics, suitable for real-time implementation with polynomial complexity.
- It implements comprehensive simulation experiments using the 5G NR system simulation library, benchmarking against BestCQI, priority-only, packet scheduling, and proportional fair algorithms. The outcomes are demonstrated with the use of comprehensive performance metrics such as throughput, resource utilization, fairness, priority satisfaction, and delay violation rate.

The remainder of this paper is organized as follows. Section 2 provides a background of the 5G network. Section 3 describes the proposed priority-aware scheduling algorithm. Section 4 presents the performance evaluation methodology. Section 5 discusses the results, evaluation, and analysis. Finally, Section 6 concludes the article.

2. 5G Network Background

This section provides an overview of the 5G NR architecture, the QoS framework, and the fundamentals of the radio resource grid. 5G NR networks can be deployed based on non-standalone (NSA) or standalone (SA) architectures. The SA architecture uses the complete 5G core (5GC), while the NSA extends the existing LTE infrastructure.

As shown in Fig. 1, the essential components are the UE, the radio access network (RAN), and the service-based 5GC. The main element of RAN is a set of gNB (5G node B) or (next generation node B) units, i.e. the base station and is interconnected by the Xn interface. A gNB can be subdivided into a central gNB unit (gNB-CU) and gNB distributed units (gNB-DU), connected by the F1 interface, supporting FDD and TDD modes. 5GC comprises modular network functions (NF) such as:

- user plane function (UPF) for data transport and data network (DN) for external connectivity,
- control plane for access and mobility management function (AMF), session management function (SMF), application function (AF), unified data management (UDM), policy control function (PCF), network repository function (NRF) and network slice selection function (NSSF).

The separation of control and user planes, which supports flexible user/data anchoring and virtualization of network functions, distinguishes 5G from LTE [30].

2.1. 5G QoS Framework

The 5G NR QoS framework supports diverse service requirements using a flow-based model, where each service flow is mapped to a data radio bearer (DRB) and is identified by a 5QI. The 5QI ensures differentiation by defined values ranging from 1 to 86 [31]. Each 5QI value corresponds to a predefined set of QoS parameters, such as:

- resource type: guaranteed bit rate (GBR), delay critical GBR, and non-GBR,
- priority: small priority value indicates higher priority,
- packet delay budget (PDB): maximum allowable delay D_{max} [ms],
- packet error rate (PER): maximum error rate,
- default maximum data burst volume: for delayed critical GBR services.

2.2. Priority and Delay Management

Let $r_{i,p}$ and $r_{j,q}$ represent PRB allocations for UE i and j with priority p and q, respectively. A priority-based QoS violation occurs if:

$$\mathcal{V}_{p,i}(t) = \begin{cases} 1 & \text{if } r_{i,p} = \emptyset \text{ and } r_{j,q} \neq \emptyset \ \ (p < q) \\ 0 & \text{otherwise} \end{cases} . \quad (1)$$

Each flow delay is managed with its PDB, D_i^{max}. If the network access delay D_i^n for UE i exceeds D_i^{max}, a delay violation occurs, thus:

$$D_i^n \leqslant D_i^{\max}, \quad \forall i \in \mathcal{U}(t)\,. \tag{2}$$

A QoS violation for delay-bound MTC occurs if:

$$\mathcal{V}_{d,i}(t) = \begin{cases} 1, & \text{if } D_i^n > D_i^{\max} \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

It ensures that the scheduler satisfies the priority and delay requirements for each QoS flow [32].

2.3. 5G Frame Structure and Resource Grid

5G NR uses flexible numerology μ ($\mu = 0, 1, 2, 3, 4$) to define the subcarrier spacing (SCS) $\Delta f = 2^\mu \times 15$ kHz, impacting slot duration and radio frame structure. Each radio frame spans 10 ms, comprising 10 subframes (1 ms each). Unlike LTE, which uses a fixed 15 kHz SCS, NR supports 15, 30, 60, 120, and 240 kHz SCS, allowing slot durations as low as 0.0625 ms for low-latency services [33].

Numerologies enable scalable, low-latency, or high-bandwidth scheduling, with slot durations decreasing as SCS increases. Each slot contains 14 orthogonal frequency-division multiplexing (OFDM) symbols (normal cyclic prefix). Table 2 presents the supported flexible transmission numerologies in 5G. As shown in Fig. 2, the resource grid represents the allocation of time-frequency resources as a two-dimensional grid with the frequency axis representing subcarriers ($N_{grid,x}^{size,\mu} \times N_{sc}^{PRB}$) and the time axis representing OFDM symbols ($N_{symb}^{slot} \times 2^\mu$), where PRB is the collection of 12 subcarriers in the frequency domain for one slot duration and resource element (RE) is the smallest resource unit representing one subcarrier for one OFDM symbol duration.

2.4. 5G Radio Resource Scheduling Framework

The scheduler, a key function of the gNB, allocates time-frequency resources to UEs for uplink and downlink communication. It is located within the medium access control (MAC) layer, but its operation is coordinated with the radio link control (RLC) and the physical layer (PHY) to enable efficient resource mapping [36].

The scheduling framework is driven by a set of metrics and reports transmitted by the UE. The gNB is based on a combination of UE-reported and gNB-measured information to efficiently allocate resources [37].

Scheduling metrics include uplink channel state measurement, buffer status report (BSR) and scheduling request, traffic pattern, power headroom report, logical channel configuration, and hybrid automatic repeat request (HARQ) [38].

PRB allocations in NR can be either contiguous or non-contiguous for a given UE. NR defines multiple types of the resource allocation process [39]:

- type 0 (resource block group, RBG-based, allowing non-contiguous allocation across the BWP,

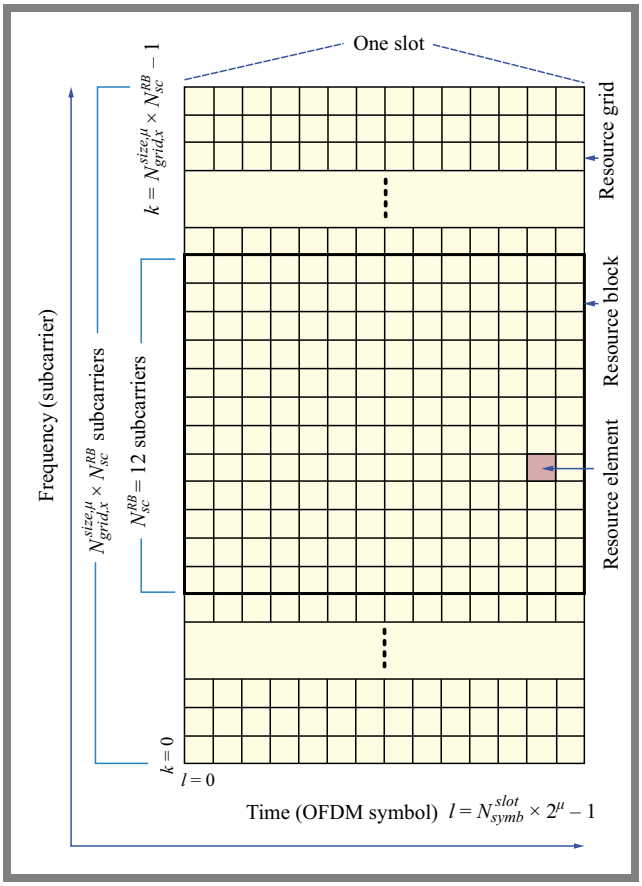

Fig. 2. Resource grid in 5G NR: subcarriers, OFDM symbols, PRB, and RE [35]

- type 1 (RB-based, enabling contiguous resource assignments for each UE).

The proposed scheduler considers service classes and dynamically assigns PRBs for sporadic, bursty mMTC traffic. The BSR, PDB, and throughput requirements are implemented in resource assignment.

3. Priority-aware Uplink Scheduling

3.1. Network Model

Consider a 5G NR cell with a single gNB serving a set $\mathcal{U} = \{1, 2, \ldots, N\}$ of mMTC UEs that are randomly distributed within the cell coverage area. At any scheduling interval t, let $\mathcal{X}_t \subseteq \mathcal{U}$ represent the subset of eligible UEs with buffered data. Each UE $i \in \mathcal{X}_t$ has a pending uplink grant request, with a resource requirement of r_i^t PRBs. Scheduling decisions are made at each transmission time interval (TTI) of the duration T_{TTI}. The total number of PRBs available in the cell is M [40].

3.2. Channel Model

The wireless channel between UE i and gNB is characterized by the channel quality indicator (CQI), $c_i^t \in \{0, 1, 2, \ldots, 15\}$. CQI is derived from signal-to-interference-plus-noise ratio (SINR) measurements using an effective SINR mapping

Tab. 2. 5G NR-supported flexible transmission numerologies [34].

μ	SCS [kHz]	FR1	FR2	PRB BW	Slots/frame	Slot duration	Symbol duration
0	15	√	×	180 kHz	10	1000 µs	66.67 µs
1	30	√	×	360 kHz	20	500 µs	33.33 µs
2	60	√	√	720 kHz	40	250 µs	16.67 µs
3	120	×	√	1.44 MHz	80	125 µs	8.33 µs
4	240	×	√	2.88 MHz	160	62.5 µs	4.17 µs

procedure:

$$\mathrm{SINR}_{\mathrm{eff},i}^{t} = \beta_1 \mathcal{I}^{-1}\left(\frac{1}{N_s}\sum_{k=1}^{N_s}\mathcal{I}\frac{\mathrm{SINR}_{i,k}^{t}}{\beta_2}\right), \tag{4}$$

where N_s is the number of SINR samples, $\mathcal{I}(\cdot)$ is the mutual information function and β_1, β_2 are parameters specific to the modulation and coding scheme (MCS).

Furthermore, [41] specifies a mapping table that relates the resulting SINR to the CQI and the transport block size (TBS) index, and we use it.

3.3. Dynamic Fitness Function

The proposed priority-aware uplink scheduling (PAUS) algorithm employs a dynamic fitness function to maximize the utility of the system by considering channel quality, QoS priority, packet aging, and long-term fairness. For each UE $i \in \mathcal{X}_t$ at scheduling time t, the fitness function is defined as:

$$F_i^t = \omega_i^t + \alpha_i^t \cdot \left(F_i^{t-1} + \xi_i^t\right), \tag{5}$$

where ω_i^t is the channel-aware priority weight, α_i^t is a temporal decay factor, F_i^{t-1} is the historical fitness value from the previous TTI ($F_i^0 = 0$) and ξ_i^t is a fairness adjustment term. The fitness value acts as a dynamic scheduling weight that evolves over time.

To balance efficiency with QoS requirements, we define a priority weight that combines channel conditions and static QoS priority. We assume that each UE i has a static priority $p_i \in \{10, 20, \ldots, P\}$, where lower values indicate a higher priority (i.e., $p_i = 10$ is the highest priority). We normalize CQI to [0,1], as $c_{i,\mathrm{norm}}^t = c_i^t/15$. The channel-aware priority weight is formulated as:

$$\omega_i^t = (c_{i,\mathrm{norm}}^t)^{\gamma} \cdot \left(1 + \frac{P - p_i}{P}\right), \tag{6}$$

where $\gamma \in [0.5, 2]$ is a scaling parameter that controls the influence of channel quality and P is the lowest priority index. This ensures that a higher CQI combined with a higher QoS priority increases the chances of UE scheduling.

To prevent excessive delays for buffered packets and ensure timely delivery for delay-sensitive applications while maintaining system stability, a temporal decay factor is introduced:

$$\alpha_i^t = \begin{cases} e^{-\beta\cdot(t - t_i^{\mathrm{last}})} & \text{if } t - t_i^{\mathrm{last}} < D_i^{\max} - \epsilon \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where t_i^{last} is the last scheduling time for UE i, D_i is the PDB for UE i, $\beta \in [0.01, 1]$ is the decay rate parameter, $0 \leqslant \epsilon \leqslant 0.2 \times D_i$, is a tolerance threshold, and $\alpha_i^t \in [0, 1]$. The decay mechanism ensures that UEs are scheduled before their packet deadline while resetting at the edge of the deadline to mitigate QoS delay violations and enforces soft deadline scheduling.

To mitigate the starvation of low-priority UEs, we incorporate a fairness adjustment term that provides a gradually increasing boost to UEs that have not been scheduled for a long time.

$$\xi_i^t = \frac{\eta}{1 + e^{-\lambda\cdot(t - t_i^{\mathrm{last}})}}, \quad 0 \leqslant \xi_i^t \leqslant \eta, \tag{8}$$

where $\eta \in [0.1, 2]$ controls the maximum fairness boost and $\lambda \in [0.05, 2]$ determines the steepness of the sigmoid function. This term implements a bounded increase in fitness score proportional to waiting time, ensuring that no UE experiences starvation regardless of its priority or channel conditions. Sigmoid-based term increases fitness for prolonged unscheduled UEs, but is upper bounded by η, preventing instability.

The specific default values and recommended ranges for γ, β, η, and λ were determined by the R-method [42], a technique to classify multiple weighted criteria or parameters.

3.4. Optimization Problem Formulation

The dynamic uplink scheduling problem is formulated as a constrained binary integer optimization problem aimed at maximizing the aggregate fitness of scheduled UEs while respecting resource and operational constraints. The decision variables are the UE scheduling vector x_t and the PRB assignment matrix $\mathbf{s}_t$:

$$\max_{x_t, \mathbf{s}_t} \sum_{i \in \mathcal{X}_t} x_i^t \cdot F_i^t, \tag{9}$$

subject to

$$\sum_{i \in \mathcal{X}_t} x_i^t \cdot r_i^t \leqslant M, \tag{10}$$

$$x_i^t \in \{0, 1\}, \quad \forall i \in \mathcal{X}_t, \tag{11}$$

$$\sum_{i \in \mathcal{X}_t} s_{i,j}^t \leqslant 1, \quad \forall j \in \{1, \ldots, M\}, \tag{12}$$

$$\sum_{j=1}^{M} s_{i,j}^t = r_i^t \cdot x_i^t, \quad \forall i \in \mathcal{X}_t, \tag{13}$$

$$s_{i,j}^t \in \{0, 1\}, \quad \forall i \in \mathcal{X}_t, j \in \{1, \ldots, M\}. \tag{14}$$

The optimization problem forms Eqs. (9) – (14) and determines the set of UEs to be scheduled and the corresponding PRB-level allocation within each TTI. Constraint (10) enforces the PRB limitation of the uplink bandwidth by ensuring that the aggregate PRB requirement of all selected UEs does not exceed the total number of PRBs available. The binary constraint (11) specifies whether UE i is selected for transmission and manages the scheduling decision, while (12) enforces PRB exclusivity by ensuring that each PRB is assigned to no more than one UE, reflecting the orthogonality requirement of NR uplink transmissions. Completeness constraint (13) imposes an all-or-nothing allocation rule, ensuring that a UE receives either the full set of PRBs it requests or nothing, thereby avoiding partial transport block formation and simplifying HARQ operations [43].

Finally, constraint (14) ensures the binary PRB assignment and manages resource mapping. Therefore, the formulation captures the discrete, non-linear, and coupled nature of uplink scheduling in 5G NR, making the problem NP hard and motivating the use of heuristic algorithms such as the proposed PAUS Algorithm 1.

Because the uplink scheduling problem is formulated in Eqs. (9) – (14) is NP-hard, a greedy heuristic is adopted to obtain an efficient real-time solution. The algorithm computes the fitness value for each eligible UE and iteratively selects the UEs in descending order of fitness until all PRBs are exhausted. This procedure ensures that the scheduler does not forget about UEs that have been waiting, making it inherently fair while still being efficient. The uplink scheduling workflow is illustrated in Fig. 3.

The computational complexity of the PAUS algorithm is dominated by the UE fitness sorting step. Calculating the fitness function for all eligible UEs requires $O(|\mathcal{X}_t|)$ operations. The sorting step has the complexity of $O(|\mathcal{X}_t| \log |\mathcal{X}_t|)$. The greedy resource allocation loop executes at most $O(|\mathcal{X}_t|)$ iterations. Therefore, the overall time complexity is $O(|\mathcal{X}_t| \log |\mathcal{X}_t|)$, which is polynomial and feasible for real-time scheduling, where $|\mathcal{X}_t|$ may range from tens to thousands of UEs per TTI.

Algorithm 1 PAUS algorithm for dynamic scheduling

Input: $\mathcal{X}_t$, $\{p_i, c_i^t, r_i^t, D_i, t_i^{\text{last}}, F_i^{t-1}\}_{i\in\mathcal{X}_t}$, $M, \gamma, \beta, \eta, \lambda, \epsilon$
Output: scheduled UE set $\mathcal{S}_t$, updated fitness $\{F_i^t\}_{i\in\mathcal{X}_t}$, PRB allocation map $\mathbf{s}_t$
Initialize: $\mathcal{S}_t = \emptyset$, $M_{\text{rem}} = M$
for each $i \in \mathcal{X}_t$ **do**
    Compute ω_i^t using Eq. (6)
    Compute α_i^t using Eq. (7)
    Compute ξ_i^t using Eq. (8)
    Compute F_i^t using Eq. (5)
end for
Sort UEs in $\mathcal{X}_t$ in descending order of F_i^t values
for each UE i in sorted order **do**
    if $r_i^t \leqslant M_{\text{rem}}$ **then**
        $\mathcal{S}_t = \mathcal{S}_t \cup \{i\}$
        $M_{\text{rem}} = M_{\text{rem}} - r_i^t$
        Allocate r_i^t PRBs to UE i
        Set $s_{i,j}^t = 1$ for all allocated PRBs j
        Update $t_i^{\text{last}} = t$
    end if
end for
Return $\mathcal{S}_t$, $\{F_i^t\}_{i\in\mathcal{X}_t}$ and PRB allocation map $\mathbf{s}_t$

4. Performance Evaluation

4.1. Simulation Setup

We consider a network scenario with a coverage radius of 1 km, where mMTC UEs are randomly distributed within a single 5G NR macro cell, and gNB is positioned at the cell's center. UEs communicate directly with the gNB without requiring dedicated gateways. UEs with the same 5QI priorities are grouped for traffic differentiation.

The simulation is implemented using the Matlab 5G NR system simulator. This framework provides comprehensive simulation capabilities, incorporating the PHY, MAC, and RLC layers, as well as configurable traffic generation. UEs remain stationary during each simulation run to isolate scheduler performance from mobility-based channel variations.

For PHY modeling, we utilize the integrated PHY abstraction in the 5G Toolbox, which implements link-to-system mapping for performance evaluation. The channel model is the 3GPP TR 38.901 urban macro (UM) scenario [44]. RLC entities operate in an unacknowledged mode to reduce protocol overhead, a suitable choice for latency-sensitive and small-payload mMTC traffic.

Scheduling decisions are executed in the MAC layer of the gNB, which supports the integration of custom uplink schedulers. The proposed PAUS scheduler utilizes a composite metric that incorporates instantaneous CQI feedback, 5QI priority, packet aging, and long-term fairness. The key simulation parameters are summarized in Tab. 3. A duration of

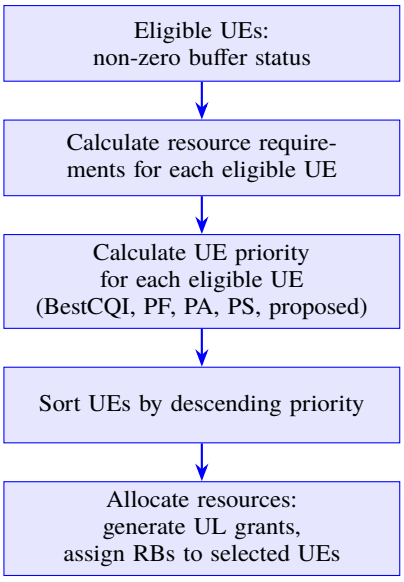

Fig. 3. Particle swarm optimization flow chart.

Tab. 3. Simulation parameters.

Parameter	Value
Channel bandwidth	5 MHz
Number of PRBs	25
Duplex mode	FDD
Number of gNBs	1
Numerology (μ)	0 (15 kHz SCS)
Subcarriers per PRB	12
OFDM symbols per slot	14
Slots per subframe	1
Subframes per frame	10
Frame duration	10 ms
Slot duration	1 ms
Channel model	Urban macro (UM)
Number of UEs	100, 200, . . . , 500
UE positions	Random
UE height	1.5 m
Simulation duration	500 frames

500 frames provides sufficient time to observe steady-state scheduler behavior under varying mMTC loads while keeping simulation runtime practical.

4.2. Application Scenario

The evaluation considers a smart city deployment characterized by high-density machine-type connectivity. mMTC UEs, such as smart meters, environmental sensors, public asset monitors, and traffic control are utilized. Their communication pattern consists mainly of sporadic small payload uplink transmissions. The primary challenge for the 5GC and NR RAN is to efficiently support massive connectivity while minimizing signaling load and satisfying heterogeneous QoS requirements. The traffic distribution across 5QI classes is presented in Tab. 4 [45].

4.3. Baseline Algorithms

To evaluate PAUS performance, we implement four benchmark algorithms for comparison purposes:

- Best CQI (BestCQI) scheduler. It allocates PRBs to UEs with the highest instantaneous channel quality, maximizing throughput but ignoring fairness, calculated from channel-dependent MCS, transmission rank, and available REs.
- Priority algorithm (PA). It schedules UEs strictly based on static 5QI priority, independently of channel conditions and fairness.
- Packet scheduling (PS). This algorithm [46] combines QoS requirements with the fairness mechanism, but fairness deteriorates drastically with network load.
- Proportional fair (PF) scheduler. It balance throughput and fairness by prioritizing UEs with high instantaneous-to-average rate ratios, using an exponential moving average.

4.4. Performance Metrics

The following key performance indicators are used to assess the efficiency of uplink scheduling algorithms.

The average cell throughput R_{cell} measures the mean data rate successfully delivered from all scheduled UEs to the gNB and is defined as:

$$R_{\text{cell}}\,[\text{Mbps}] = \frac{\left(\sum_{t=1}^{T_{\text{slot}}}\sum_{i=1}^{N} R_i^t\right)\times 8}{T_{\text{sim}}\times 10^6}\,, \tag{15}$$

where R_i^t is the data [bytes] received from UE i in slot t, T_{slot} is the total number of slots, T_{sim} ($\text{NumFrames}\times 0.01$ s/frame) is the total simulation time in seconds, and N is the total number of active UE.

Resource utilization ρ quantifies the ratio of PRBs allocated to the total PRBs available and is defined as:

$$\rho = \frac{1}{T_{\text{slot}}}\sum_{t=1}^{T_{\text{slot}}}\frac{\sum_{i=1}^{N} r_i^t}{M}, \qquad 0\leqslant\rho\leqslant 1\,, \tag{16}$$

where r_i^t is the number of PRBs allocated to UE i in the slot t and M is the total number of PRBs available per slot.

Instantaneous resource utilization is defined as:

$$\rho(t) = \frac{\sum_{i=1}^{N} r_i^t}{M}\,. \tag{17}$$

The fairness index $\mathcal{F}$ evaluates the equitable distribution of radio resources among competing UEs, using Jain's fairness index is:

$$\mathcal{F} = \frac{\left(\sum_{i=1}^{N} R_i\right)^2}{N\cdot\sum_{i=1}^{N} R_i^2}, \qquad 0\leqslant\mathcal{F}\leqslant 1\,, \tag{18}$$

where each R_i is the throughput for UE i:

$$R_i = \frac{\left(\sum_{t=1}^{T_{\text{slot}}} R_i^t\right)\times 8}{T_{\text{sim}}\times 10^6}\,. \tag{19}$$

5QI priority satisfaction evaluates the proportion of scheduling events where the scheduler fails to satisfy QoS priority differentiation among traffic classes. Lower values indicate better 5QI priority satisfaction and improved QoS differentiation.

The delay violation rate represents the average number of events when the packets exceed their PDB before successful transmission. A lower delay violation rate indicates better delay performance.

5. Results and Discussion

This section analyzes the experimental results obtained from extensive simulations for a range of mMTC network sizes

Tab. 4. mMTC traffic parameters for the cell radius r = 1000 m.

Priority	5QI	D_i^{max} [ms]	Packets/min	Traffic type	UE [%]	Example service
10	5	100	5	Event update	12	Emergency notification service
20	1	100	10	Event update	7	Intrusion detection service
30	3	50	30	Periodic update	12	Critical monitoring service
40	2	150	30	Event/periodic update	12	Environment monitoring
50	4	300	40	Event/periodic update	19	Smart home service
60	6	300	60	Periodic update	19	Utility application service
70	7	100	60	Periodic update	19	Smart meters

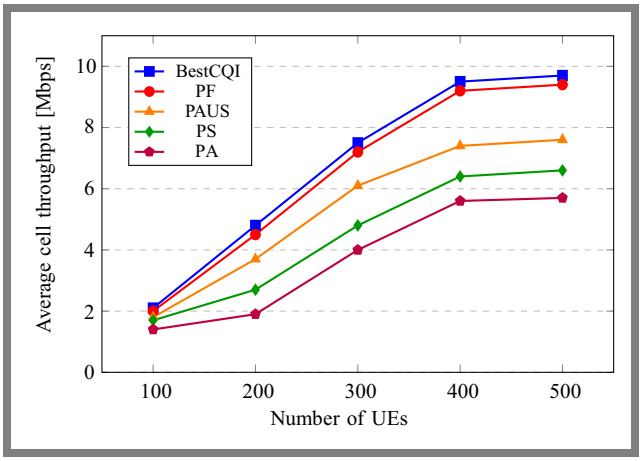

Fig. 4. Average cell throughput.

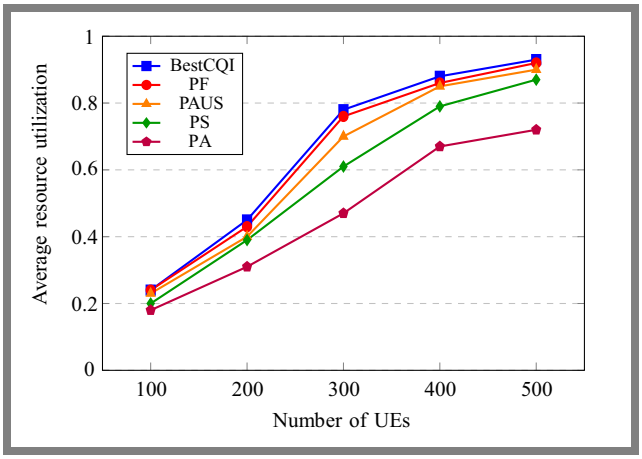

Fig. 5. Average resource utilization.

and load conditions. Each experiment is repeated for 50 independent runs. The reported results correspond to the sample mean across the runs.

5.1. Average Cell Throughput

Average cell throughput is defined in Eq. (15). Figure 4 compares the average cell throughput of the five schedulers under consideration as the number of UEs increases from 100 to 500. Channel-oriented schemes (BestCQI and PF) achieve the highest throughput because they prioritize instantaneous channel capacity. The PAUS scheduler achieves optimal throughput while improving fairness and priority satisfaction. Importantly, PAUS outperforms the PA and PS baselines across all network sizes by combining channel awareness with priority and aging information, aligning scalability and efficiency.

5.2. Resource Utilization

Resource utilization, defined in Eq. (16), quantifies the efficiency of PRB usage. Figure 5 shows that CQI-centric schedulers (BestCQI, PF) achieve the highest utilization, exceeding 90% under heavy load. PAUS achieves high utilization (up to 90%) by opportunistically exploiting channel instances while respecting priority constraints. The PA algorithm underperforms in utilization at high loads because it schedules according to static priority, often leaving spectral opportunities unused when high-priority UEs have poor channel quality, underlining the importance of channel awareness for spectral efficiency.

5.3. Fairness Index

The fairness index is defined in Eq. (18), where values close to 1 denote high fairness. Figure 6 illustrates that PAUS achieves higher fairness ≈ 0.84 than a purely channel-oriented scheduler ≈ 0.30 and static priority ≈ 0.46 at 500 UEs. PF exhibits intermediate fairness. The high fairness of PAUS is due to the sigmoid-based fairness boost ξ_i^t and the aging-aware decay factor, which together mitigate the starvation of low-priority or persistently poor-channel UEs, suitable for dense, heterogeneous IoT environments.

5.4. Priority Satisfaction

Figure 7 shows that the PA baseline delivers the lowest violation rate (highest 5QI satisfaction), but at a cost to throughput and fairness. PAUS achieves a balanced trade-off with 0.23 for 500 UE, lower than PF and BestCQI, while maintaining optimal throughput and fairness than PA. PF and BestCQI prioritize instantaneous channel conditions, accumulating more 5QI priority violations. This shows that PAUS effectively balances QoS differentiation with minimum sacrifice of efficiency.

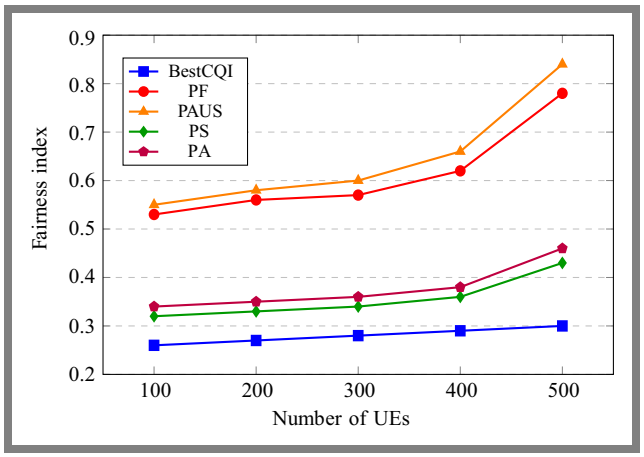

Fig. 6. Fairness index.

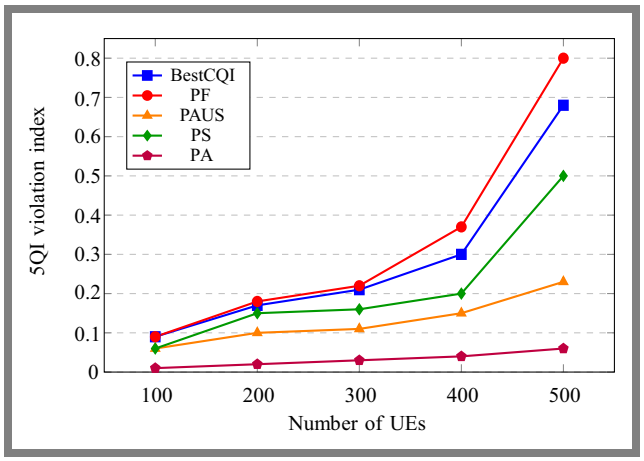

Fig. 7. 5QI priority satisfaction.

5.5. Delay Violation Rate

The delay violation rate is a critical mMTC reliability indicator. The results shown in Fig. 8 illustrate that PAUS yields the lowest average delay violation, ≈ 8.5 for 500 UE, due to the deadline-aware temporal decay α_i^t and the fairness boost that prioritizes aged packets, and is lower than all other schemes. PF and BestCQI, which ignore deadline information, incur substantially higher delay violations under heavy loads. The findings suggest the suitability of PAUS for delay-sensitive mMTC traffic.

The results demonstrate that PAUS achieves a balance between efficiency and fairness/QoS differentiation in dense mMTC deployments. Although BestCQI and PF maximize throughput, they suffer from poor fairness and higher 5QI violations. In contrast, PA satisfy priority at the cost of efficiency. PAUS achieves near-PF throughput with substantially improved fairness and priority satisfaction, making it suitable for urban mMTC scenarios.

6. Conclusions

This paper presents a priority-aware uplink radio resource scheduling algorithm to address the challenge of balancing efficiency with optimal fairness in dense 5G mMTC deployments. The novelty of the proposed algorithm lies in its dynamic fitness function, which integrates a channel-aware priority weight to utilize channel conditions, a temporal decay mechanism to mitigate delay violations based on the PDB, and a sigmoid-based fairness adjustment to prevent starvation of low-priority or poor-channel UEs.

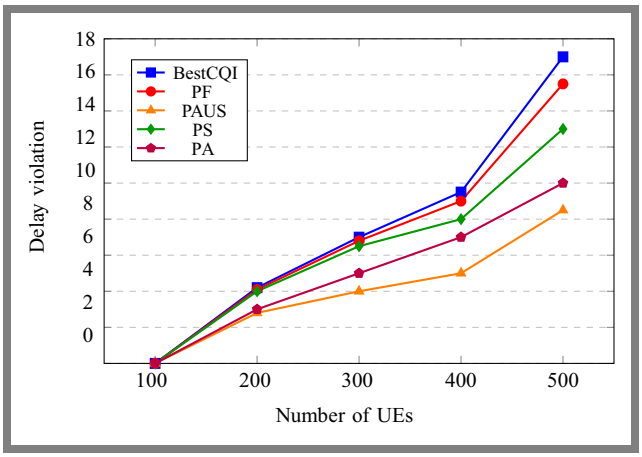

Fig. 8. 5QI priority satisfaction.

Extensive simulation results demonstrated that PAUS achieved near-optimal throughput and resource utilization, close to channel-aware schemes (PF, BestCQI). PAUS outperformed efficiency-focused schemes in QoS metrics and achieved the highest fairness and the lowest delay violation rate, which can manage delay-sensitive traffic. Furthermore, it maintained a low 5QI priority violation, demonstrating effective QoS differentiation with balanced spectral efficiency.

Polynomial-time complexity makes PAUS implementable within real-time 5G gNB MAC scheduler architectures and is suitable for supporting the large-scale, heterogeneous traffic requirements of urban smart city and industrial IoT applications.

References

[1] 3GPP, "5G; Study on New Radio (NR) Access Technology (3GPP TR 38.912 Version 18.0.0 Release 18)", Technical Report, 2024.

[2] J. Janković *et al.*, "Effects of Differentiated 5G Services on Computational and Radio Resource Allocation Performance", *IEEE Transactions on Network and Service Management*, vol. 18, pp. 2226–2241, 2021 (https://doi.org/10.1109/TNSM.2021.3060865).

[3] 3GPP, "5G; NR; NR and NG-RAN Overall Description; Stage-2 (3GPP TS 38.300 Version 17.0.0 Release 17)", Technical Specification, 2022.

[4] S.R. Pokhrel *et al.*, "Towards Enabling Critical mMTC: A Review of URLLC within mMTC", *IEEE Access*, vol. 8, pp. 131796–131813, 2020 (https://doi.org/10.1109/ACCESS.2020.3010271).

[5] S. Rezwan and W. Choi, "Priority-based Joint Resource Allocation with Deep Q-learning for Heterogeneous NOMA Systems", *IEEE Access*, vol. 9, pp. 41468–41481, 2021 (https://doi.org/10.1109/ACCESS.2021.3065314).

[6] M. Attaran, "The Impact of 5G on the Evolution of Intelligent Automation and Industry Digitization", *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 5977–5993, 2023 (https://doi.org/10.1007/s12652-020-02521-x).

[7] S. Hamdoun, A. Rachedi, and Y. Ghamri-Doudane, "Graph-based Radio Resource Sharing Schemes for MTC in D2D-based 5G Networks", *Mobile Networks and Applications*, vol. 25, pp. 1095–1113, 2020 (https://doi.org/10.1007/s11036-020-01527-1).

[8] P.K. Baheti and A. Khunteta, "QoS Aware Resource Scheduling in LTE Network for Smart City M2M Communication", *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)*, Raipur, India, 2021 (https://doi.org/10.1109/TRIBES52498.2021.9751661).

[9] T.-S. Lee, C.-H. Yang, T.-Y. Kuo, and Y.-J. Wu, "Chapter 13 - Resource Allocation in Massive Machine-type Communications", in: *Resource Optimization in Wireless Communications*, pp. 317–347, 2025 (https://doi.org/10.1016/B978-0-44-330092-9.00019-7).

[10] B.-S. Kim, "A Priority-aware Dynamic Scheduling Algorithm for Ensuring Data Freshness in 5G Networks", *Future Generation Computer Systems*, vol. 163, art. no. 107542, 2025 (https://doi.org/10.1016/j.future.2024.107542).

[11] P.K. Baheti and A. Khunteta, "Service-aware Resource Scheduling for Heterogeneous M2M Communication in 5G Networks", *International Journal of Basic and Applied Sciences*, vol. 14, pp. 137–143, 2025 (https://doi.org/10.14419/saqrkp91).

[12] P. Vidhya, K. Subashini, R. Sathishkannan, and S. Gayathri, "Dynamic Network Slicing Based Resource Management and Service Aware Virtual Network Function (VNF) Migration in 5G Networks", *Computer Networks*, vol. 259, art. no. 111064, 2025 (https://doi.org/10.1016/j.comnet.2025.111064).

[13] S. Malta, P. Pinto, and M. Fernandez-Veiga, "Optimizing 5G Network Slicing with DRL: Balancing eMBB, URLLC, and mMTC with OMA, NOMA, and RSMA", *Journal of Network and Computer Applications*, vol. 234, art. no. 104068, 2025 (https://doi.org/10.1016/j.jnca.2024.104068).

[14] W. Hamdi, O. Dağdeviren, and H. Bulut, "QoS-aware Network Slicing and Resource Management for Internet of Vehicles in 5G Networks", *Ad Hoc Networks*, vol. 178, art. no. 103976, 2025 (https://doi.org/10.1016/j.adhoc.2025.103976).

[15] E. Goshi, F. Mehmeti, T.F. La Porta, and W. Kellerer, "Modeling and Analysis of mMTC Traffic in 5G Core Networks", *IEEE Transactions on Network and Service Management*, vol. 22, pp. 409–425, 2025 (https://doi.org/10.1109/TNSM.2024.3481240).

[16] A. Samuylov *et al.*, "Performance of MAC Layer Mechanisms in DECT-2020 NR mMTC Technology", *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, Singapore, Singapore, 2024 (https://doi.org/10.1109/VTC2024-Spring62846.2024.10683658).

[17] T.N. Weerasinghe, V. Casares-Giner, I.A.M. Balapuwaduge, and F.Y. Li, "Priority Enabled Grant-free Access with Dynamic Slot Allocation for Heterogeneous mMTC Traffic in 5G NR Networks", *IEEE Transactions on Communications*, vol. 69, pp. 3192–3206, 2021 (https://doi.org/10.1109/TCOMM.2021.3053990).

[18] M. Abdullah *et al.*, "Satellite Synergy: Navigating Resource Allocation and Energy Efficiency in IoT Networks", *Journal of Network and Computer Applications*, vol. 230, art. no. 103966, 2024 (https://doi.org/10.1016/j.jnca.2024.103966).

[19] Y.L. Lee, T.C. Chuah, J. Loo, and F. Ke, "Proportional-fair Uplink Resource Allocation with Statistical QoS Provisioning for RAN Slicing", *Physical Communication*, vol. 65, art. no. 102389, 2024 (https://doi.org/10.1016/j.phycom.2024.102389).

[20] M.O. Kabaou *et al.*, "Empowering Communication Networks with MMR Scheduler: A Novel Approach to Balancing User Throughput and Fairness", *Alexandria Engineering Journal*, vol. 76, pp. 641–649, 2023 (https://doi.org/10.1016/j.aej.2023.06.042).

[21] L. Zhang, A. Liu, and X. Chen, "A WMMSE-based Contiguous Resource Scheduling Algorithm for 5G-NR Uplink", IEEE Wireless Communications Letters, vol. 13, pp. 466–470, 2024 (https://doi.org/10.1109/LWC.2023.3332327).

[22] K. Boutiba, M. Bagaa, and A. Ksentini, "Optimal Radio Resource Management in 5G NR Featuring Network Slicing", *Computer Networks*, vol. 234, art. no. 109937, 2023 (https://doi.org/10.1016/j.comnet.2023.109937).

[23] O. Elgarhy *et al.*, "Energy Efficiency and Latency Optimization for IoT URLLC and mMTC Use Cases", *IEEE Access*, vol. 12, pp. 23132–23148, 2024 (https://doi.org/10.1109/ACCESS.2024.3364349).

[24] Y. Kaura, B. Lall, R.K. Mallik, and A. Singhal, "Adaptive Scheduling of Shared Grant-free Resources for Heterogeneous Massive Machine Type Communication in 5G and Beyond Networks", *IEEE Transactions on Network and Service Management*, vol. 22, pp. 1188–1204, 2025 (https://doi.org/10.1109/TNSM.2024.3493015).

[25] P.K. Baheti and A. Khunteta, "Priority-based Resource Scheduling for Smart City M2M Communication in 5G Networks", *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, India, 2023 (https://doi.org/10.1109/ICMNWC60182.2023.10435821).

[26] S.A. AlQahtani, "Cooperative-aware Radio Resource Allocation Scheme for 5G Network Slicing in Cloud Radio Access Networks", *Sensors*, vol. 23, art. no. 5111, 2023 (https://doi.org/10.3390/s23115111).

[27] W.U. Rehman *et al.*, "Improved Resource Allocation in 5G MTC Networks", *IEEE Access*, vol. 8, pp. 49187–49197, 2020 (https://doi.org/10.1109/ACCESS.2020.2974632).

[28] D. Ivanova *et al.*, "Mathematical Framework for Mixed Reservation- and Priority-based Traffic Coexistence in 5G NR Systems", *Mathematics*, vol. 11, art. no. 1046, 2023 (https://doi.org/10.3390/math11041046).

[29] S.B. Prathiba, K. Raja, R.V. Saiabirami, and G. Kannan, "An Energy-aware Tailored Resource Management for Cellular-based Zero-touch Deterministic Industrial M2M Networks", *IEEE Access*, vol. 12, pp. 33613–33627, 2024 (https://doi.org/10.1109/ACCESS.2024.3372417).

[30] M.R.M. Anfar and J. Mwangama, "Machine Learning-based Service Differentiation in the 5G Core Network", *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju Island, South Korea, 2021 (https://doi.org/10.1109/ICAIIC51459.2021.9415263).

[31] 3GPP, "5G; System Architecture for the 5G System (5GS) (3GPP TS 23.501 Version 17.5.0 Release 17)", Technical Specification, 2022 (https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/17.05.00_60/ts_123501v170500p.pdf).

[32] S.O. Oladejo and O.E. Falowo, "Latency-aware Dynamic Resource Allocation Scheme for Multi-tier 5G Network: A Network Slicing-multitenancy Scenario", *IEEE Access*, vol. 8, pp. 74834–74852, 2020 (https://doi.org/10.1109/ACCESS.2020.2988710).

[33] B. Agarwal, M.A. Togou, M. Marco, and G.-M. Muntean, "A Comprehensive Survey on Radio Resource Management in 5G HetNets: Current Solutions, Future Trends and Open Issues", *IEEE Communications Surveys & Tutorials*, vol. 24, pp. 2495–2534, 2022 (https://doi.org/10.1109/COMST.2022.3207967).

[34] A. Yazar and H. Arslan, "Flexible Multi-numerology Systems for 5G New Radio", *Journal of Mobile Multimedia*, vol. 14, pp. 367–394, 2018.

[35] E. Engin, I. Hokelek, A. Gorcin, and H.A. Cirpan, "A Pre-emptive Scheduling Mechanism for Service Assurance of Network Slicing in Next Generation Cellular Networks", *IEEE Access*, vol. 13, pp. 23297–23311, 2025 (https://doi.org/10.1109/ACCESS.2025.3536997).

[36] 3GPP, "5G; NR; Medium Access Control (MAC) Protocol Specification (3GPP TS 38.321 Version 16.1.0 Release 16)", Technical Specification, 2020.

[37] V. Kovtun, O. Kovtun, K. Grochla, and O. Yasniy, "The Quality of Service Assessment of eMBB and mMTC Traffic in a Clustered 5G Ecosystem of a Smart Factory", *Egyptian Informatics Journal*, vol. 29, art. no. 100598, 2025 (https://doi.org/10.1016/j.eij.2024.100598).

[38] T. Wang, "Energy-efficient Resource Allocation for UAV-aided Full-duplex OFDMA Wireless Powered IoT Communication Networks", *Journal of King Saud University - Computer and Information Sciences*, vol. 36, art. no. 102225, 2024 (https://doi.org/10.1016/j.jksuci.2024.102225).

[39] P.K. Korrai *et al.*, "Joint Power and Resource Block Allocation for Mixed-numerology-based 5G Downlink under Imperfect CSI", *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1583–1601, 2020 (https://doi.org/10.1109/OJCOMS.2020.3029553).

[40] 3GPP, "5G; NR; Physical Channels and Modulation (3GPP TS 38.211 Version 16.2.0 Release 16)", Technical Specification, 2020.

[41] 3GPP, "5G; NR; Physical Layer Procedures for Data (3GPP TS 38.214 Version 16.2.0 Release 16)", Technical Specification, 2020.

[42] R.V. Rao and J. Lakshmi, "R-method: A Simple Ranking Method for Multi-attribute Decision-making in the Industrial Environment", *Journal of Project Management*, vol. 6, pp. 223–230, 2021 (https://doi.org/10.5267/j.jpm.2021.5.001).

[43] 3GPP, "5G; NR; Physical Layer Procedures for Control (3GPP TS 38.213 Version 17.1.0 Release 17)", Technical Specification, 2022.

[44] 3GPP, "5G; Study on Channel Model for Frequencies From 0.5 to 100 GHz (3GPP TR 38.901 Version 16.1.0 Release 16)", Technical Report, 2020.

[45] J. Navarro-Ortiz *et al.*, "A Survey on 5G Usage Scenarios and Traffic Models", *IEEE Communications Surveys & Tutorials*, vol. 22, pp. 905–929, 2020 (https://doi.org/10.1109/COMST.2020.2971781).

[46] A. Iqbal, T. Khurshaid, A. Nauman, and S.-B. Rhee, "Energy-aware Ultra-reliable Low-latency Communication for Healthcare IoT in Beyond 5G and 6G Networks", *Sensors*, vol. 25, art. no. 3474, 2025 (https://doi.org/10.3390/s25113474).

Prashant Kumar Baheti, Research Scholar
Department of Computer Science and Engineering
https://orcid.org/0009-0003-0728-0009
E-mail: 2020phdevenprashant9185@poornima.edu.in
Poornima University, Jaipur, Rajasthan, India
https://www.poornima.edu.in

Ajay Khunteta, Ph.D., Professor
Department of Computer Science and Engineering
https://orcid.org/0000-0002-5335-9434
E-mail: ajay.khunteta@poornima.edu.in
Poornima University, Jaipur, Rajasthan, India
https://www.poornima.edu.in

A Spectral Efficiency Design for Active IRS-assisted SWIPT System via Semidefinite Relaxation Method

Pham Viet Tuan[1], Hoang Dai Long[2], Pham Ngoc Son[3], and Mai T.P. Le[4]

[1]*University of Education, Hue University, Hue City, Vietnam,*
[2]*University of Sciences, Hue University, Hue City, Vietnam,*
[3]*Ho Chi Minh City University of Technology and Engineering, Ho Chi Minh City, Vietnam,*
[4]*University of Science and Technology, The University of Danang, Da Nang City, Vietnam*

https://doi.org/10.26636/jtit.2026.1.2267

Abstract — Active intelligent reflecting surfaces (IRS) with phase-shift and amplifier capabilities have arisen as a solution relied upon to improve spectral/energy efficiency of wireless systems, as they outperform conventional passive techniques/without IRS assistance. In this work, the simultaneous wireless information and power transfer (SWIPT) downlink is supported by an active IRS, where a multi-antenna base station (BS) broadcasts both information and power to multiple hybrid power-splitting (PS) users. The target of sum data rate maximization is to study the constraints of user energy harvesting thresholds and power transmission limitations of BS and active IRS. To tackle this complicated issue, iterative algorithms are proposed to find the optimal beamforming vector, PS coefficients, and IRS parameters, as amplification factors and phase shift. A joint optimization framework using alternating optimization, semidefinite relaxation, and non-convex approximations is used. Finally, simulation experiments are performed to assess that the proposed iterative algorithms of the active IRS scheme converge fast and achieve better sum rate results than conventional baseline schemes.

Keywords — intelligent reflecting surface, power splitting, semidefinite relaxation, simultaneous wireless information and power transfer, sum data rate

1. Introduction

Simultaneous wireless information and power transfer (SWIPT) has emerged as an effective method for powering energy-constrained devices by enabling concurrent transmission of information and energy over a single wireless channel [1]–[3]. In the meantime, active intelligent reflecting surfaces (IRS), a recent evolution of conventional passive IRS, have become crucial enablers for next-generation wireless networks, owing to their unique capability to reflect and amplify signals simultaneously [4]–[6]. Unlike passive IRS, which lacks amplification capabilities and, therefore, suffers significant cascaded path loss, active IRS effectively counteracts this limitation, enhancing signal strength during the reflection phase.

Prior research exploring IRS-assisted SWIPT primarily addressed passive IRS configurations, focusing on optimizing the reflecting elements' phase shifts to maximize energy harvesting efficiency or enhance data transmission rates [7]–[14]. For instance, several studies investigated optimizing energy efficiency and transmit power control in IRS-aided multiple-input single-output (MISO) SWIPT networks involving multiple users [7], [8]. Secure SWIPT systems employing passive IRS have also aimed at maximizing minimum secrecy rates and overall energy efficiency [11], [12]. Additionally, non-orthogonal multiple access (NOMA) schemes were employed in combination with passive IRS to further boost system performance, particularly through optimized power allocation strategies [13] and improved uplink sum rates for IoT devices [8].

Despite these advancements, the inability of passive IRS to amplify reflected signals considerably restricts their effectiveness, especially under long-distance wireless communication scenarios [2].

Recently, the introduction of active IRS has significantly mitigated these drawbacks by integrating amplification mechanisms directly with the reflecting elements. This enhancement allows active IRS to effectively overcome path loss, resulting in improvements in information transmission and energy harvesting capabilities. However, research into integrating active IRS with SWIPT remains relatively unexplored [15], [16].

A pioneering study addressing this integration, described in [15], demonstrated substantial advantages compared to passive IRS configurations, including extended wireless energy transfer range, increased sum harvested energy (SHE), and enhanced achievable data rates. Furthermore, recent findings indicate that integrating active IRS with NOMA-based SWIPT systems can further improve overall power efficiency and the minimum achievable secrecy rates [16]. Therefore, such an integration is a promising avenue for deploying robust and efficient sixth-generation (6G) architectures.

In this research, SWIPT-based communication is investigated between the transceiver of the base station (BS) with smart antennas and hybrid PS users, with the assistance of an active IRS. Two types of downlink transmit beamforming from the base station to multiple receivers may be distinguished [17], [18]. The first type is unicast transmit beamforming, where separated individual data streams are sent to multiple

Tab. 1. Summary of previous research.

Ref.	Mode	Users	IRS type	Goals	Methods	Main results
[15]	Unicast	Individual ID, EH	Active	Maximum SHE(SR) of EH(ID) users	AO, SDR, SCA, Gaussian random	Better SHE(SR) than w/o IRS
[19]	Broadcast	Only ID users	Passive	Minimum transfer power	AO, SDR, Gaussian random	Lower transfer power than w/o IRS
[20]	Unicast	Co-located ID, EH	Active	Balance SR and SHE	AO, SCA, price-based	Better results than w/o IRS
[21]	Unicast	Co-located ID, EH	Active	Minimum transfer power	AO, SDR, rank-1 approximation	Lower transfer power than passive IRS
[22]	Unicast	Individual multi-antenna ID, EH	Passive	Maximum minimum individual SR	AO, covariance matrices	Perfect/imperfect CSI, better w/o IRS
[23]	Unicast	Co-located ID, EH	Active STAR-RIS	Maximize EE	AO, SDR, FP, reinforced learning	Higher EE than baselines
Proposed	Broadcast (multicast)	Co-located ID, EH	Active	Maximize sum rate	SDR, AO, Taylor, SCA, rank-1 appr.	Higher SR than passive, w/o IRS

receivers. Otherwise, in this paper, we consider the other variety of multicast beamforming, where the base station with N antennas transmits common information $s(t)$ to multiple receivers.

In the work described in [10], only the subproblem in which the optimization of BS and hybrid user parameters is solved with fixed active IRS parameters was considered. Meanwhile, in this work, we optimize both transceivers and active IRS parameters of the overall SWIPT system, comparing the results to baselines of passive IRS assistance and a setup without IRS.

Unlike in the case of phase-shift passive IRS with unchanged signal power, both power amplification and phase change are exploited in the active IRS. Then, spectral and energy efficiency in SWIPT communication is expectedly improved with active IRS.

We present our model and the related papers in Tab. 1. The major novelty of this work is indicated as:

- The proposed SWIPT system is a combination of broadcast transmission, i.e. the same information is transferred to all users with power-splitting scheme, and active IRS assistance. Almost all other SWIPT-related works considered the unicast transmission, i.e. each user received an individual information stream.
- In the presented broadcast PS-SWIPT scheme, the sum rate target and active IRS assistance are considered, while the related broadcast IRS-aided system studied minimization of transmission power, with/without passive IRS.
- The general transformation techniques relying on semidefinite relaxation (SDR) and alternating optimization (AO) were utilized widely in the prior studies. However, we have contributed the mathematical manipulation and the efficient solution approach with Taylor and rank-1 approximations for the SDR method.

Here, we tackle the highly important issues of overcoming double fading and IoT sustainability in 6G and IoT networks. Double fading is a phenomenon where passive IRS suffers from severe path loss (product of BS-to-IRS and IRS-to-User paths). By addressing this with active IRS (with the amplification feature), we address a primary bottleneck preventing practical deployment of IRS. Additionally, SWIPT is critical for extending the lifespan of energy-constrained IoT devices without replacing their batteries. Ensuring a minimum harvested energy threshold E_k while maximizing data rates is a must for achieving sustainability in IoT.

To summarize, the main contributions of this work are outlined below.

- The goal of sum data rate maximization (SRM) (i.e., spectral efficiency) is studied, with the required parameters related to energy harvesting threshold, BS power limitation, and power budget of the active surface amplifier taken into consideration. The transceiver and active IRS parameters of BS smart beamformer and user power-splitting coefficients, the IRS phase change and amplification factors are jointly designed in the complicated non-convex SRM issue.
- We simplify the SRM problem by using the mathematical transformer and the alternating optimization (AO) technique. Then, the semidefinite relaxation (SDR) method with rank-one approximation is combined with successive convex approximation (SCA) for designing the iterative algorithm solution.
- Finally, numerical experiments are performed to analyze iteration convergence and performance compared to passive IRS and to a scenario without IRS baseline schemes. Furthermore, impacts of some system parameters are also taken into consideration.

The structure of this paper is as follows. The communication system and the SRM issue are described in Section 2. The proposed iterative algorithm for solving the SRM issue is presented in Section 3. Finally, numerical results and conclusions are shown in Sections 4 and 5, respectively.

In this paper, lowercase letters denote scalars, bold lowercase letters denote vectors, and bold uppercase letters denote matrices. In addition, $|\,x\,|$ is the absolute value of a complex scalar while $\|\,\mathbf{x}\,\|$ is the Euclidean norm of a complex vector $\mathbf{x}$. Furthemore, a diagonal matrix $\mathrm{diag}\,(\mathbf{x})$ is generated by a diagonal vector $\mathbf{x}$. The maximum eigenvalue and corresponding eigenvector of a matrix $\mathbf{X}$ are represented by $\lambda_1\,(\mathbf{X})$ and $\mathbf{v}_1$, respectively. $\mathbb{C}^{m\times n}$ is indicated as the $m \times n$ complex matrix space. Lastly, $\mathcal{CN}\left(\mu, \sigma^2\right)$ is the distribution of a circularly symmetric complex Gaussian random variable with mean μ and variance σ^2.

2. System Model Description

We study a SWIPT downlink transmission network as presented in Fig. 1. In that, a base station provided with M antennas transmits data and power to numerous hybrid PS users. The common precoding beamforming vector $\mathbf{w} \in \mathbb{C}^{M\times 1}$ is exploited at the BS and utilized for all users. The communication signal sent by the BS is indicated as $\mathbf{x} = \mathbf{w}s$ with the intended symbol s under the presumption of a zero-mean and unit variance random variable. Thus, the transmission power of the BS is calculated as $\|\mathbf{w}\|^2 \leqslant P_{\mathrm{Tx}}$ with the limited power capacity P_{Tx}.

Each PS-based hybrid user employs a power-splitting structure to simultaneously perform information decoding and energy harvesting. To improve the quality of communication, an active IRS intervention including N components is deployed. The BS-IRS link is represented by $\mathbf{G} \in \mathbb{C}^{N\times M}$ while the BS-user and IRS-user links are $\mathbf{t}_k \in \mathbb{C}^{M\times 1}$ and $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$, $\forall k \in \mathcal{K} = \{1, \ldots, K\}$, respectively.

Similar to [40], at the n-th element of the active IRS, $b_n \geqslant 0$ represents the amplification coefficient and $\varphi_n \in [0, 2\pi)$ represents the phase-shift coefficient, $\forall n \in \mathcal{N} = \{1, \ldots, N\}$. As a result, the amplification impact $\mathbf{\Psi} = \mathrm{diag}\,(b_1, \ldots, b_N)$ and the phase-shift impact $\mathbf{R} = \mathrm{diag}\left(e^{j\varphi_1}, \ldots, e^{j\varphi_N}\right)$ express the amplification and reflection of the active IRS.

Accordingly, the k-th hybrid user's arriving signal, including the direct and reflecting signals, is represented as:

$$\phi_k = \left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}s + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{z}_{\mathrm{I}} + n_k,\ \forall k \in \mathcal{K}\,. \quad (1)$$

Here, the intelligent surface with amplifying capability generates a noise $\mathbf{z}_{\mathrm{I}} \sim \mathcal{CN}\left(\mathbf{0}_N, \sigma_z^2 \mathbf{I}_N\right)$ [15], [24], [25] and the k-th hybrid user's antenna has the additive noise $n_k \sim \mathcal{CN}\left(0,\ \sigma_k^2\right)$. The k-th PS-based hybrid user splits the received signal into two streams using a power-splitting coefficient $\theta_k \in (0, 1)$, where one stream, with a power ratio of θ_k is utilized for information decoding, and the other, with a power ratio of $(1 - \theta_k)$ is used for energy harvesting [21], [26].

Although theoretical studies often model power splitting as a lossless operation, practical implementations using Wilkinson power dividers or varactor-based circuits introduce non-negligible insertion losses, such as heat on resistance elements [27], [28]. The power splitter is usually designed for a fixed impedance (e.g., 50 Ω). However, impedance mismatch incurs a loss in the harvested energy when the input impedance of the energy harvesting circuit (rectifier) changes dynamically with the input power level (i.e. is non-linear) [29]–[31]. Consequently, the ID signal part is represented as $\phi_k^{\mathrm{ID}} = \sqrt{\theta_k}\,\phi_k + v_k$ under a processing noise $v_k \sim \mathcal{CN}\left(0,\ \delta_k^2\right)$. The EH signal part is represented as $\phi_k^{\mathrm{EH}} = \sqrt{1-\theta_k}\,\phi_k$. For those reasons as [9], [40] the information rate of the k-th hybrid PS user is formulated as:

$$\Omega_k = \log_2\left(1 + \frac{\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2}{\sigma_z^2 \|\mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\|^2 + \sigma_k^2 + \frac{\delta_k^2}{\theta_k}}\right),\ \forall k \in \mathcal{K}\,, \quad (2)$$

and the received energy with the presumption of completed efficiency conversion is given by:

$$\Upsilon_k = (1 - \theta_k)\left(\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2 + \sigma_z^2 \left\|\mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\right\|^2\right), \forall k\,. \quad (3)$$

3. Joint Active IRS and Transceivers Design

Here, the sum rate maximization problem for active IRS-aided SWIPT system is introduced and an efficient iteration solution algorithm is proposed. The target is to optimize spectral efficiency by jointly designing the parameters of transceivers and active IRS.

Unlike in [9], [40], here, the BS beamformer, the hybrid users PS coefficients, and the active IRS's amplification and phase-shifting coefficients are degined jointly using a novel spectral efficiency maximization problem (P1) which can be formulated as:

$$\underset{\{\mathbf{w},\theta_k,\varphi_n,b_n\}}{\text{maximize}} \sum_{k=1}^{K} \log_2\left(1 + \frac{\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2}{\sigma_z^2 \left\|\mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\right\|^2 + \sigma_k^2 + \frac{\delta_k^2}{\theta_k}}\right) \quad (4a)$$

subject to: (4b)

$$(1 - \theta_k)\left(\begin{array}{l}\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2 \\ +\sigma_z^2 \left\|\mathbf{h}_k^H \mathbf{\Psi}\mathbf{R}\right\|^2\end{array}\right) \geqslant E_k, \forall k \quad (4c)$$

$$\|\mathbf{w}\|^2 \leqslant P_{\mathrm{Tx}} \quad (4d)$$

$$\|\mathbf{\Psi}\mathbf{R}\mathbf{G}\mathbf{w}\|^2 + \sigma_z^2 \|\mathbf{\Psi}\mathbf{R}\|^2 \leqslant P_{\mathrm{I}} \quad (4e)$$

$$1 \geqslant \theta_k \geqslant 0,\ \forall k \in \mathcal{K} \quad (4f)$$

$$\varphi_n \in [0, 2\pi)\,,\ \forall n \in \mathcal{N}. \quad (4g)$$

The target formulated in (4a) is the maximization of the total information rate of hybrid users. We consider a system suitable for energy-constrained applications, such as sensor networks, where only the base station possesses a reliable power supply. The receivers must therefore sustain their operations by replenishing energy via the signals received from the active IRS-assisted base station. In the constraint (4b), the amount of harvested energy is larger than the minimum required EH threshold, E_k which provides enough power for basic operations, similar to [13], [32], [33]. Constraints (4d)

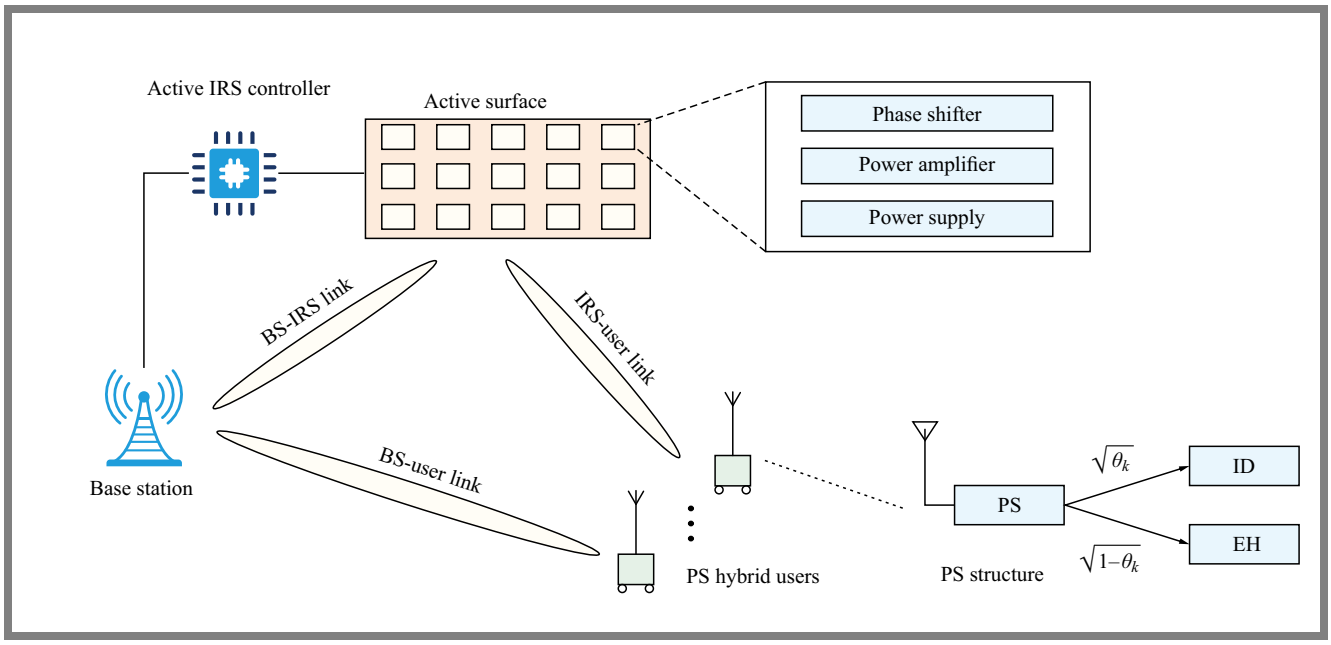

Fig. 1. Illustration of SWIPT communication with an active IRS intercession and PS hybrid users.

and (4e) introduce the maximum powers of the BS, P_{Tx}, and the active IRS, P_{I}.

In this paper, we consider that the amplifier units of active IRS elements share a common power source, as in [4], [21], [25], [34]. Thus, the amplification power of all active IRS elements is limited by the total power constraint, as illustrated in constraint (4e). Otherwise, in the other model of active IRS, each active IRS reflecting element is equipped with its own power source and independently controlled power. Therefore, the maximum amplification factor of each element is limited by its individual amplification power budget. This is a challenging problem we will aim to consider in our future research.

Formulas (4f) and (4g) present the intervals of PS coefficients at the hybrid users and the phase shifts at the active IRS components, respectively. The idea relied upon to solve the spectral efficiency maximization problem consists in combining alternating optimization, SCA iteration, and SDR with rank-one relaxation. We first separate the coupled variables $\{\mathbf{w}, \theta_k, \varphi_n, b_n\}$ into two sets of transceiver parameters $\{\mathbf{w}, \theta_k\}$ and active IRS parameters $\{\varphi_n, b_n\}$, then the SDR method is applied to solve subproblems.

3.1. Solving the Transceiver Parameters Sub-problem

As stated in (4a), the total sum rate function is very complicated due to the couple of transceiver parameters of beamforming vector, PS factors $\{\mathbf{w}, \theta_k\}$ and the SINR ratios inside $\sum_{k=1}^{K} \log_2(\cdot)$ function. Thus, we simplify the objective function by inserting the auxiliary variables $\{u_k,\ v_k\}$ as formulas (5a), (5b), and (5c) where two problems have the same maximum value (please see the proof in Appendix A). Then, we transform into the equivalent problem (P2) as:

$$\max_{\substack{\{\mathbf{w},\theta_k,\\ u_k,v_k\}}} \sum_{k=1}^{K} \log_2\left(1+\frac{u_k}{v_k+\sigma_k^2}\right) \tag{5a}$$

$$\text{s.t.: } \left|\left(\mathbf{t}_k^H+\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2 \geqslant u_k, \forall k\in\mathcal{K} \tag{5b}$$

$$\sigma_z^2\left\|\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\right\|^2+\frac{\delta_k^2}{\theta_k}\leqslant v_k, \forall k\in\mathcal{K} \tag{5c}$$

$$\left(\begin{matrix}\left|\left(\mathbf{t}_k^H+\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2\\ +\sigma_z^2\left\|\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\right\|^2\end{matrix}\right)\geqslant\frac{E_k}{(1-\theta_k)}, \forall k\in\mathcal{K} \tag{5d}$$

$$\|\mathbf{w}\|^2\leqslant P_{\mathrm{Tx}} \tag{5e}$$

$$\|\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\mathbf{w}\|^2+\sigma_z^2\|\boldsymbol{\Psi}\mathbf{R}\|^2\leqslant P_{\mathrm{I}} \tag{5f}$$

$$1\geqslant\theta_k\geqslant 0,\ \forall k\in\mathcal{K}. \tag{5g}$$

For constraint (5b), we then convert the energy harvesting constraints as:

$$\left|\left(\mathbf{t}_k^H+\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2=\mathrm{Tr}\left(\mathbf{T}_k\mathbf{w}\mathbf{w}^H\right), \tag{6}$$

where

$$\mathbf{T}_k=\left(\mathbf{t}_k^H+\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)^H\left(\mathbf{t}_k^H+\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right).$$

For constraint (5c), the function exhibits convexity with respect to $\{\theta_k, v_k\}$.

For constraints (5d), we express as follows:

$$\mathrm{Tr}\left(\mathbf{T}_k\mathbf{w}\mathbf{w}^H\right)+\sigma_z^2\left\|\mathbf{h}_k^H\boldsymbol{\Psi}\mathbf{R}\right\|^2\geqslant\frac{E_k}{(1-\theta_k)}. \tag{7}$$

For constraint (5e), $\mathrm{Tr}\left(\mathbf{w}\mathbf{w}^H\right)\leqslant P_{\mathrm{Tx}}$ and constraint (5f), we denote $\mathbf{H}=\left(\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)^H\left(\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)$ and formulate:

$$\mathrm{Tr}\left(\left(\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)^H\left(\boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\mathbf{w}^H\right)+\sigma_z^2\|\boldsymbol{\Psi}\mathbf{R}\|^2\leqslant P_{\mathrm{I}}. \tag{8}$$

For the objective function, we analyze:

$$\log_2\left(1+\frac{u_k}{v_k+\sigma_k^2}\right) = \log_2\left(v_k+u_k+\sigma_k^2\right)-\log_2\left(v_k+\sigma_k^2\right). \tag{9}$$

By exploiting SDR method, we transform the beamforming vector to the matrix variables $\mathbf{w}$ by the useful transformation $\mathbf{W}=\mathbf{w}\mathbf{w}^H$, $\mathbf{w}\neq\mathbf{0}$ where $\mathbf{W}\succeq 0$, $\text{rank}\,(\mathbf{W})=1$ [35].

$$\max_{\substack{\mathbf{W},\theta_k,\\ u_k,v_k}}\ \sum_{k=1}^{K}\left[\log_2\left(v_k+u_k+\sigma_k^2\right)-\log_2\left(v_k+\sigma_k^2\right)\right] \tag{10a}$$

$$\text{s.t.: } u_k-\text{Tr}\left(\begin{array}{c}\left(\mathbf{t}_k^H+\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\mathbf{G}\right)^H\\ \left(\mathbf{t}_k^H+\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{W}\end{array}\right)\leqslant 0,\forall k\in\mathcal{K} \tag{10b}$$

$$\sigma_z^2\left\|\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\right\|^2+\frac{\delta_k^2}{\theta_k}-v_k\leqslant 0,\forall k\in\mathcal{K} \tag{10c}$$

$$\frac{E_k}{(1-\theta_k)}-\left(\text{Tr}\left(\mathbf{T}_k\mathbf{W}\right)+\sigma_z^2\left\|\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\right\|^2\right)\leqslant 0,\forall k\in\mathcal{K} \tag{10d}$$

$$\text{Tr}\left(\mathbf{W}\right)-P_{\text{Tx}}\leqslant 0 \tag{10e}$$

$$\text{Tr}\left(\left(\mathbf{\Psi}\mathbf{R}\mathbf{G}\right)^H\left(\mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{W}\right)+\sigma_z^2\|\mathbf{\Psi}\mathbf{R}\|^2-P_{\text{I}}\leqslant 0 \tag{10f}$$

$$1\geqslant\theta_k\geqslant 0,\ \ \forall k\in\mathcal{K} \tag{10g}$$

$$u_k\geqslant 0, v_k\geqslant 0,\ \forall k\in\mathcal{K} \tag{10h}$$

$$\mathbf{W}\succeq 0 \tag{10i}$$

$$\text{rank}\,(\mathbf{W})=1. \tag{10j}$$

Since $\mathbf{W}$ is positive semidefinite matrix (10i), the rank-one constraint (10j) is equivalently expressed as $\text{Tr}\,(\mathbf{W})-\lambda_1(\mathbf{W})\leqslant 0$, where λ_1 is maximum eigenvalue of $\mathbf{W}$. Then via the exact penalty function method [37], we move the novel condition into the objective to formulate the penalty function with a large enough penalty parameter κ as:

$$\max_{\substack{\mathbf{W},\theta_k,\\ u_k,v_k}}\ \sum_{k=1}^{K}\left[\log_2\left(v_k+u_k+\sigma_k^2\right)-\log_2\left(v_k+\sigma_k^2\right)\right] -\kappa\left(\text{Tr}\left(\mathbf{W}\right)-\lambda_1\left(\mathbf{W}\right)\right)$$

We use Taylor approximation for the concave function:

$$\log_2\left(v_k+\sigma_k^2\right)\leqslant\left(\begin{array}{l}\log_2\left(v_k^{(l)}+\sigma_k^2\right)+\\ \frac{1}{\ln 2\cdot\left(v_k^{(l)}+\sigma_k^2\right)}\left(v_k-v_k^{(l)}\right)\end{array}\right). \tag{11}$$

Moreover, the convex function $\lambda_1(\mathbf{W})$ has a subgradient $\mathbf{q}_1\mathbf{q}_1^H$ with the unit-norm eigenvector $\mathbf{q}_1$ corresponding the maximum eigenvalue λ_1 [38]:

$$\lambda_1(\mathbf{W})\geqslant\lambda_1\left(\mathbf{W}^{(l)}\right)+\mathbf{q}_1^{(l)H}\left(\mathbf{W}-\mathbf{W}^{(l)}\right)\mathbf{q}_1^{(l)}=\mathbf{q}_1^{(l)H}\mathbf{W}\mathbf{q}_1^{(l)}. \tag{12}$$

Then, at the fixed point

$$\left\{v_k^{(l)},\mathbf{W}^{(l)},\lambda_1\left(\mathbf{W}^{(l)}\right),\mathbf{q}_1^{(l)}\right\}$$

the convex optimization subproblem is rewritten as follows:

$$\max_{\substack{\mathbf{W},\theta_k,\\ u_k,v_k\}}}\ \sum_{k=1}^{K}\left[\begin{array}{l}\log_2\left(v_k+u_k+\sigma_k^2\right)-\\ \left(\begin{array}{l}\log_2\left(v_k^{(l)}+\sigma_k^2\right)+\\ \frac{1}{\ln 2\cdot\left(v_k^{(l)}+\sigma_k^2\right)}\left(v_k-v_k^{(l)}\right)\end{array}\right)\end{array}\right] -\kappa\left(\text{Tr}\left(\mathbf{W}\right)-\mathbf{q}_1^{(l)H}\mathbf{W}\,\mathbf{q}_1^{(l)}\right) \tag{13a}$$

$$\text{s.t.: (10b), ..., (10i).} \tag{13b}$$

Therefore, this convex optimization problem may be solved by the Matlab CVX tool [36]. Then, the optimal solution of this convex optimization problem is used to update the fixed point

$$\left\{v_k^{(l)},\mathbf{W}^{(l)},\lambda_1\left(\mathbf{W}^{(l)}\right),\mathbf{q}_1^{(l)}\right\}$$

and the procedure is repeated until satisfying the convergent conditions. To find a point for the initialization, we solve the problem as:

$$\underset{\{\mathbf{W},\theta_k\}}{\text{maximize}}\ \ 0 \tag{14a}$$

$$\text{subject to:}\ \ \text{(10d), (10e), (10f), (10g), (10i).} \tag{14b}$$

The optimal solution of the problem is assigned for $\left\{\mathbf{W}^{(0)},\theta_k^{(0)}\right\}$, then $\left\{v_k^{(0)}\right\}$ is obtained as

$$v_k^{(0)}=\sigma_z^2\left\|\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\right\|^2+\frac{\delta_k^2}{\theta_k^{(0)}},\ \forall k\in\mathcal{K}.$$

Algorithm 1 The proposed algorithm with the SDR-based approach to optimize the spectral efficiency of problem (5).

1: **Initialization:** Starting fixed phase shifts $\{\varphi_n\}$, active IRS element power, $\{b_n\}$, largest iteration number, I_1
2: **For** $l=1$ **to** I_1 **do**
3: Solving problem (14) to achieve the initial feasible SCA point
4: Solving problem (13) with SDR-based method to achieve $\{\mathbf{W}^*,\theta_k^*,v_k^*,\mathbf{q}_1^*\}$ with fixed $\{b_n,\varphi_n\}$
5: Update $\left\{v_k^{(l)},\mathbf{q}_1^{(l)}\right\}\leftarrow\{v_k^*,\mathbf{q}_1^*\}$
6: **If** accuracy is obtained **then**
7: Break
8: **End If**
9: **End For**
10: **Output:** Optimal solution $\{\mathbf{w},\theta_k\}\leftarrow\{\mathbf{w}^*,\theta_k^*\}$, optimal value $\sum_{k=1}^{K}\log_2\left(1+\frac{\left|\left(\mathbf{t}_k^H+\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2}{\sigma_z^2\left\|\mathbf{h}_k^H\mathbf{\Psi}\mathbf{R}\right\|^2+\sigma_k^2+\frac{\delta_k^2}{\theta_k}}\right)$

3.2. Solving Subproblem of Active IRS Parameters

Next, we find the amplification factors and the phase shifts at the active IRS when the beamforming vector and the PS

factor are fixed. We consider (P3) as follows:

$$\max_{\{b_n,\varphi_n\}} \ 0 \tag{15a}$$

$$\text{s.t.:} \left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi R G}\right)\mathbf{w}\right|^2 \geqslant u_k,\ \forall k \in \mathcal{K} \tag{15b}$$

$$\sigma_z^2 \left\|\mathbf{h}_k^H \mathbf{\Psi R}\right\|^2 + \frac{\delta_k^2}{\theta_k} \leqslant v_k,\ \forall k \in \mathcal{K} \tag{15c}$$

$$\left(\begin{array}{l}\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi R G}\right)\mathbf{w}\right|^2 \\ +\sigma_z^2 \left\|\mathbf{h}_k^H \mathbf{\Psi R}\right\|^2\end{array}\right) \geqslant \frac{E_k}{(1-\theta_k)}, \forall k \in \mathcal{K} \tag{15d}$$

$$\left\|\mathbf{\Psi R G w}\right\|^2 + \sigma_z^2 \left\|\mathbf{\Psi R}\right\|^2 \leqslant P_1. \tag{15e}$$

For constraint (15b), we denote

$$\mathbf{f} = \left[b_1 e^{j\varphi_1}, b_2 e^{j\varphi_2}, \cdots, b_N e^{j\varphi_N}\right]^H \in \mathbb{C}^{N\times 1}$$

and thus derive

$$\mathbf{h}_k^H \mathbf{\Psi R G w} = \mathbf{f}^H diag\left(\mathbf{h}_k^H\right) \mathbf{G w}\ .$$

we rewrite as:

$$\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \mathbf{\Psi R G}\right)\mathbf{w}\right|^2 = \left|\mathbf{f}^H diag\left(\mathbf{h}_k^H\right)\mathbf{G w} + \mathbf{t}_k^H \mathbf{w}\right|^2 . \tag{16}$$

With $\mathbf{x}_k = diag\left(\mathbf{h}_k^H\right)\mathbf{G w}$, $y_k = \mathbf{t}_k^H \mathbf{w}$, we formulate:

$$\begin{aligned}\left|\mathbf{f}^H \mathbf{x}_k + y_k\right|^2 &= \left(\mathbf{f}^H \mathbf{x}_k + y_k\right)\left(\mathbf{f}^H \mathbf{x}_k + y_k\right)^H \\ &= \mathbf{f}^H \mathbf{x}_k \mathbf{x}_k^H \mathbf{f} + y_k \mathbf{x}_k^H \mathbf{f} + \mathbf{f}^H \mathbf{x}_k y_k^H + y_k^H y_k \\ &= [\mathbf{f}\ ,\ 1]^H \begin{bmatrix} \mathbf{x}_k \mathbf{x}_k^H & \mathbf{x}_k y_k^H \\ y_k \mathbf{x}_k^H & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ 1 \end{bmatrix} + y_k^H y_k\ ,\end{aligned} \tag{17}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{x}_k \mathbf{x}_k^H & \mathbf{x}_k y_k^H \\ y_k \mathbf{x}_k^H & 0 \end{bmatrix}\ .$$

We assign $\mathbf{F} = \tilde{\mathbf{f}}\tilde{\mathbf{f}}^H$, $\tilde{\mathbf{f}} = [\mathbf{f}; 1]$, Then, constraint (15b) is rewritten as:

$$\mathrm{Tr}\left(\mathbf{A}_1 \mathbf{F}\right) + y_k^H y_k \geqslant u_k,\ \forall k.$$

For constraint (15c), first we transform as:

$$\begin{aligned}\left\|\mathbf{h}_k^H \mathbf{\Psi R}\right\|^2 &= \left\|\mathbf{f}^H diag\left(\mathbf{h}_k^H\right)\right\|^2 = \left\|diag\left(\mathbf{h}_k\right)\mathbf{f}\right\|^2 \\ &= \mathbf{f}^H diag\left(\mathbf{h}_k^H\right) diag\left(\mathbf{h}_k\right)\mathbf{f} = \mathbf{f}^H \mathbf{H}_k \mathbf{f} \\ &= \tilde{\mathbf{f}}^H \tilde{\mathbf{H}}_k \tilde{\mathbf{f}} = \mathrm{Tr}\left(\tilde{\mathbf{H}}_k \mathbf{F}\right)\end{aligned} \tag{18}$$

where:

$$\begin{aligned}\mathbf{H}_k &= diag\left(\mathbf{h}_k^H\right) diag\left(\mathbf{h}_k\right), \\ \tilde{\mathbf{H}}_k &= \left[\mathbf{H}_k, \mathbf{0}_{N\times 1}; \mathbf{0}_{1\times N}, 0\right].\end{aligned}$$

Then, constraint (15c) is converted as:

$$\sigma_z^2 \mathrm{Tr}\left(\tilde{\mathbf{H}}_k \mathbf{F}\right) + \frac{\delta_k^2}{\theta_k} \leqslant v_k,\ \forall k \in \mathcal{K}\ . \tag{19}$$

For constraint (15d), from equation (18) we derive the equivalent form as:

$$\mathrm{Tr}\left(\mathbf{A}_1 \mathbf{F}\right) + y_k^H y_k + \sigma_z^2 \mathrm{Tr}\left(\tilde{\mathbf{H}}_k \mathbf{F}\right) \geqslant \frac{E_k}{(1-\theta_k)},\ \forall k \in \mathcal{K}\ . \tag{20}$$

For constraint (15e), we transform the first term on the left side as:

$$\begin{aligned}\left\|\mathbf{\Psi R G w}\right\|^2 &= \left\|\mathbf{f}^H diag\left(\mathbf{G w}\right)\right\|^2 \\ &= \mathbf{f}^H diag\left(\mathbf{G w}\right)\left(diag\left(\mathbf{G w}\right)\right)^H \mathbf{f} \\ &= \mathbf{f}^H \bar{\mathbf{G}} \mathbf{f} \\ &= \tilde{\mathbf{f}}^H \tilde{\mathbf{G}} \tilde{\mathbf{f}} = \mathrm{Tr}\left(\tilde{\mathbf{G}}\mathbf{F}\right)\end{aligned}, \tag{21}$$

where:

$$\bar{\mathbf{G}} = diag\left(\mathbf{G w}\right)\left(diag\left(\mathbf{G w}\right)\right)^H\ ,$$

$$\tilde{\mathbf{G}} = \left[\bar{\mathbf{G}}, \mathbf{0}_{N\times 1}; \mathbf{0}_{1\times N}, 0\right]\ .$$

Moreover, the second term is expressed as:

$$\left\|\mathbf{\Psi R}\right\|^2 = \left\|\mathbf{f}\right\|^2 = \left\|\tilde{\mathbf{f}}\right\|^2 - 1 = \tilde{\mathbf{f}}^H \tilde{\mathbf{f}} - 1 = \mathrm{Tr}\left(\mathbf{F}\right) - 1\ . \tag{22}$$

Similar to (12), the rank-one condition is approximated by:

$$\lambda_1\left(\mathbf{F}\right) \geqslant \lambda_1\left(\mathbf{F}^{(i)}\right) + \mathbf{e}_1^{(i)H}\left(\mathbf{F} - \mathbf{F}^{(i)}\right)\mathbf{e}_1^{(i)} = \mathbf{e}_1^{(i)H}\mathbf{F}\mathbf{e}_1^{(i)}\ . \tag{23}$$

The final problem is expressed as:

$$\min_{\{\mathbf{F}\}} \left(\mathrm{Tr}\left(\mathbf{F}\right) - \mathbf{e}_1^{(i)H}\mathbf{F}\mathbf{e}_1^{(i)}\right) \tag{24a}$$

$$\text{s.t.:}\ \mathrm{Tr}\left(\mathbf{A}_1\mathbf{F}\right) + y_k^H y_k \geqslant u_k,\ \forall k \in \mathcal{K} \tag{24b}$$

$$\sigma_z^2 \mathrm{Tr}\left(\tilde{\mathbf{H}}_k \mathbf{F}\right) + \frac{\delta_k^2}{\theta_k} \leqslant v_k,\ \forall k \in \mathcal{K} \tag{24c}$$

$$\left(\begin{array}{l}\mathrm{Tr}\left(\mathbf{A}_1\mathbf{F}\right) + y_k^H y_k \\ +\sigma_z^2 \mathrm{Tr}\left(\tilde{\mathbf{H}}_k \mathbf{F}\right)\end{array}\right) \geqslant \frac{E_k}{(1-\theta_k)},\ \forall k \in \mathcal{K} \tag{24d}$$

$$\mathrm{Tr}\left(\tilde{\mathbf{G}}\mathbf{F}\right) + \sigma_z^2\left(\mathrm{Tr}\left(\mathbf{F}\right) - 1\right) \leqslant P_1. \tag{24e}$$

To start at an initial feasible point, we solve the SDP problem to obtain $\mathbf{e}_1^{(0)}$ as follows:

$$\underset{\{\mathbf{F}\}}{\text{minimize}}\ \ 0 \tag{25a}$$

$$\text{subject to}\ \ (24b), (24c), (24d), (24e). \tag{25b}$$

Then, the initial values for variables $\mathbf{F}^{(0)}$ and $\mathbf{e}_1^{(0)}$ are obtained by the optimal solution. After that, the detailed solution is presented as Algorithm 2.

Finally, the proposed iterative algorithm is illustrated via flowchart form in Fig. 2 and the detailed steps in Algorithm 3.

4. Numerical Experiments

In this part of the paper, the numerical experiments of the iterative algorithm designed for solving the transceiver and active IRS subproblems and the total system problem are presented to evaluate the effect of the proposed method.

For setting system parameters similar to [10], [39], the hybrid user number is assigned $K = 3$, the noise powers of the active IRS and the PS hybrid users are $\sigma_{\mathrm{I}}^2 = -40$ dBm, $\sigma_k^2 = -50$ dBm, and $\delta_k^2 = -40$ dBm, $\forall k \in \mathcal{K}$. The distance-dependent pathloss for every link is expressed as $PL = [-30 - 10\,\beta \log_{10}(d)]$ dB, where is the path loss at

Algorithm 2 The iterative algorithm with the SDR-based approach to achieve optimal amplification and phase shifts coefficients.

Initialization: Fixed $\{\mathbf{w}, \theta_k, u_k, v_k\}$, setting iteration number, I_2
Finding initial values for variables $\mathbf{F}^{(0)}$ and $\mathbf{e}_1^{(0)}$ by solving the initial problem (25)
For $l = 1$ **to** I_2 **do**
  Solving tractable convex problem (24) with SDR-based method to achieve $\mathbf{F}^*$ and $\mathbf{e}_1^*$
  Update $\left\{\mathbf{F}^{(i)}, \mathbf{e}_1^{(i)}\right\} \leftarrow \{\mathbf{F}^*, \mathbf{e}_1^*\}$
  If accuracy is obtained **then**
    Break
  End If
End For
Decompose: $\mathbf{F}^{(i)} = \tilde{\mathbf{f}}^{(i)}\tilde{\mathbf{f}}^{(i)H}$
Derive $\mathbf{f}^{(i)} = \left[\tilde{\mathbf{f}}^{(i)}\right]_{1:N} / \tilde{f}_{N+1}^{(i)}$ and phase-shift coefficients $\varphi_n^{(i)} = -\arg\left(f_n^{(i)}\right), \forall n \in \mathcal{N}$, amplification coefficients $b_n^{(i)} = \left|f_n^{(i)}\right|, \forall n \in \mathcal{N}$
Output: optimal solution $\{b_n, \varphi_n\} \leftarrow \left\{b_n^{(i)}, \varphi_n^{(i)}\right\}$

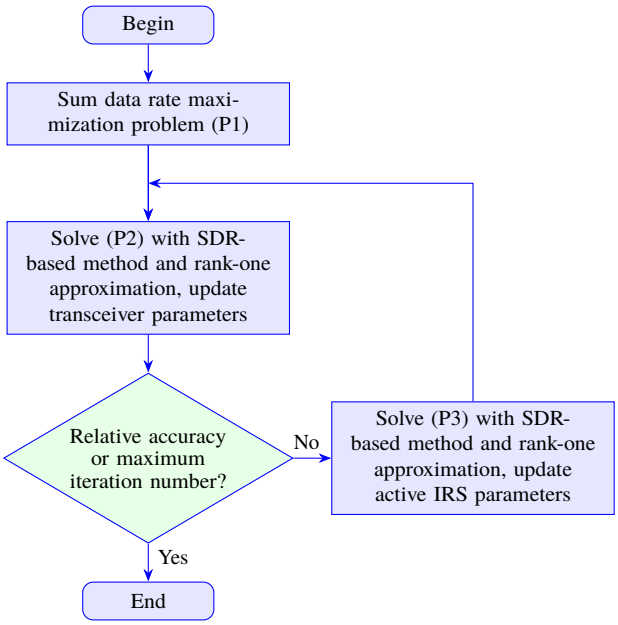

Fig. 2. Particle swarm optimization flow chart.

the reference distance –30 dB, d denotes the link range, and β indicates the pathloss factor. The factor β is allotted 2.2 for the links of BS-active IRS and active IRS-users, and 3.8 for the links of BS-users.

We apply the Rician fading formula for the small-scale model with the Rician coefficient $K_R = 3$ dB. The BS location is placed at (0 m, 0 m) and the active IRS is placed at (3 m, 2.5 m). Moreover, the hybrid user locations are uniformly placed at the circle which has the centre point (6 m, 0 m) and the radius 0.5 m.

We apply the Rician fading formula for the small-scale model with the Rician coefficient of $K_R = 3$ dB. The BS is placed

Algorithm 3 The final iterative algorithm for optimizing the transceivers and active IRS parameters.

Initialization: Starting initial phase-shift coefficients $\left\{\varphi_n^{(0)}\right\}$, active amplification coefficients, $\left\{b_n^{(0)}\right\}$, largest number of iterations, I_3
Finding initial values for variables $\mathbf{F}^{(0)}$ and $\mathbf{e}_1^{(0)}$ by solving the initial problem (25)
For $p = 1$ **to** I_3 **do**
  Solving problem (5) via SDR-based approach to achieve $\left\{\mathbf{w}^{(p)}, \theta_k^{(p)}\right\}$ under the given active IRS parameters
  Solving problem (15) via SDR-based approach to achieve $\left\{b_n^{(p)}, \varphi_n^{(p)}\right\}$ under the given transceiver parameters
  If accuracy is obtained **then**
    Break
  End If
End For
Output: optimal solution $\{\mathbf{w}, \theta_k, b_n, \varphi_n\} \leftarrow \left\{\mathbf{w}^{(p)}, \theta_k^{(p)}, b_n^{(p)}, \varphi_n^{(p)}\right\}$, and optimal value $\sum_{k=1}^{K} \log_2\left(1 + \frac{\left|\left(\mathbf{t}_k^H + \mathbf{h}_k^H \boldsymbol{\Psi}\mathbf{R}\mathbf{G}\right)\mathbf{w}\right|^2}{\sigma_z^2\left\|\mathbf{h}_k^H \boldsymbol{\Psi}\mathbf{R}\right\|^2 + \sigma_k^2 + \frac{\delta_k^2}{\theta_k}}\right)$

at (0 m, 0 m) and the active IRS is placed at (3 m, 2.5 m). Moreover, the hybrid user locations are uniformly placed along a circle with a center point of (6 m, 0 m) and a radius of 0.5 m.

Figure 3 expresses the iteration results of the SDR-based iterative procedure completed using Algorithm 2 under several fixed phase shifts and amplification coefficients of the active intelligent surface. The subproblem (13) provides the target result which rises and goes to the fixed value with lower five repetitions. In addition, the result of $\left(\mathrm{Tr}\left(\mathbf{W}\right) - \mathbf{z}_1^{(l)H}\mathbf{W}\mathbf{z}_1^{(l)}\right)$ is approximate zero with the high value penalty coefficient κ. Consequently, SDR-based formula of rank-1 condition is satisfied, then the BS beamforming solution is obtained via decomposition technique.

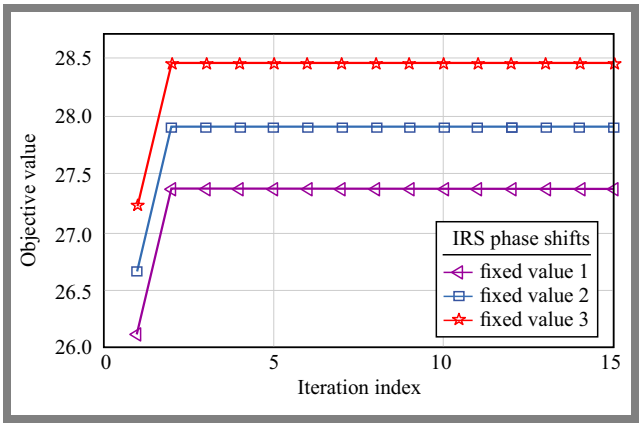

Fig. 3. Convergent behavior of proposed iterative Algorithm 1.

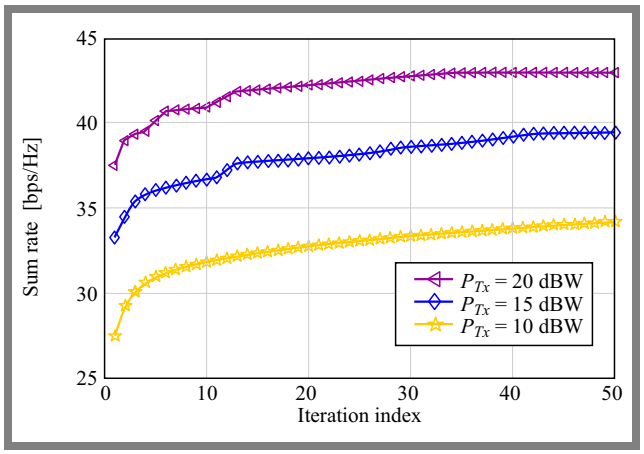

Fig. 4. Convergent behavior of proposed iterative Algorithm 3.

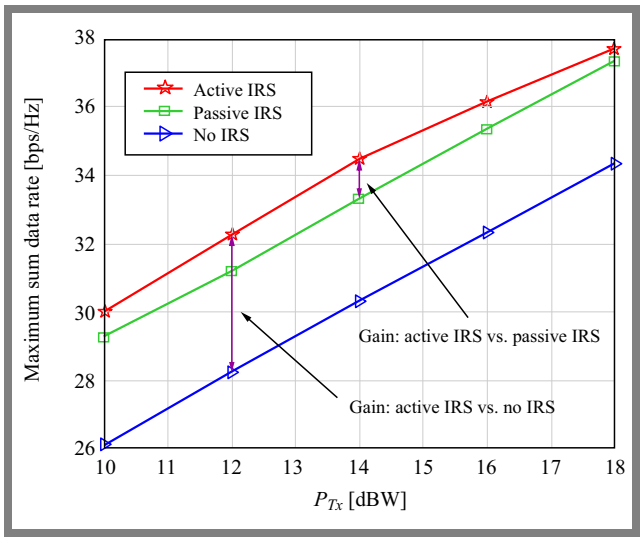

Fig. 5. Maximum sum data rate versus BS transmission power.

Figure 4 expresses the iteration results of the proposed alternating optimization Algorithm 3 under some BS transmission power values. As seen in Fig. 3, the sum data rate increases gradually after each iteration until convergence is achieved. Then, we set the largest iteration of AO iteration number is $I_3 = 50$ and the relative accuracy is $\Delta = 10^{-2}$ for terminating loop.

Figure 5 presents the maximum sum data rate via BS transmission power for three schemes, including the proposed active IRS, passive IRS, and a solution without IRS, when the number of BS antennas is $M = 5$, number of IRS components is $N = 10$, active IRS power is $P_I = 5$ dBW, and the required energy harvesting threshold is $E_k = -25$ dBm.

We observe that the support of active IRS provides the highest total sum rate in comparison with a scenario with passive IRS and a solution without IRS schemes. For instance, when $P_{\mathrm{Tx}} = 12$ dBW, the proposed active IRS scheme achieves 32.23 bps/Hz, while the passive IRS and the solution without IRS obtain 31.21 bps/Hz and 28.24 bps/Hz, respectively. Furthermore, the increase of maximum BS power provides a higher total sum rate.

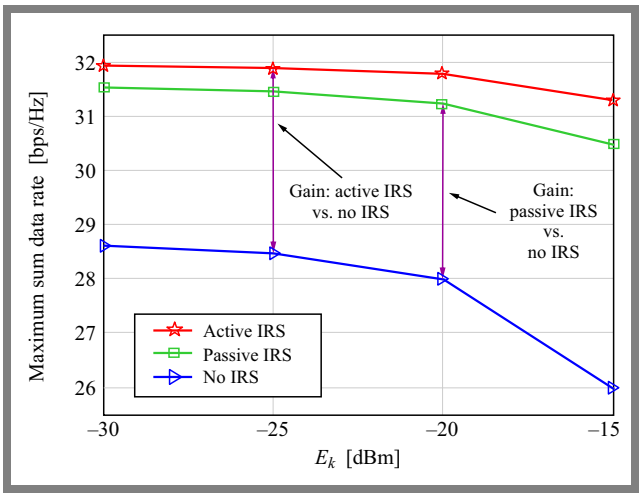

Fig. 6. Maximum sum data rate versus required energy harvesting value.

Figure 6 presents the maximum sum data rate via the required EH value, E_k in three schemes, including the proposed active IRS, passive IRS, and a scheme without IRS, when the number of BS antennas is $M = 8$, number of IRS components is $N = 10$, active IRS power is $P_I = 1$ dBW, and maximum BS power is $P_{\mathrm{Tx}} = 10$ dBW.

It can be observed that the sum data rate of PS hybrid users goes down with the increase of the required EH threshold. The reason is that when the EH threshold is higher, PS hybrid users have to split more energy for EH circuits, and then keep less energy for decoding data. In addition, the result of the considered active IRS scheme is also higher than that of the passive IRS and the scheme without IRS when both amplification and phase-shift coefficients are optimized.

5. Conclusions

In this study, the total data rate optimization for SWIPT with PS hybrid users and an active IRS aid is investigated via an SDR-based approach, rank-one approximations, and AO method. The SDR-based iterative algorithm shows fast convergence with a few iterations. Under the impact of amplification characteristics, the sum rate result of the active IRS-assisted scheme is better than that of the passive IRS-assisted SWIPT with only phase shifts and the solution scheme without IRS.

Moreover, the sum rate results are higher with an increase in the number of BS antennas, number of IRS components or BS transmission power. In conclusion, future research should focus on rendering the system more robust under the influence of imperfect channel-side information of the active IRS.

Acknowledgments

This research is funded by University of Education, Hue University, under grant number NCTB-T.24-TN.103.02. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED), under grant number 102.04-2023.40.

Appendix A: Proof that problems (4) and (5) have the same maximum value

We denote the objective functions of problems (4) and (5) as $f_4\left(\mathbf{w}, \theta_k\right)$ and $f_5\left(\mathbf{w}, \theta_k, u_k, v_k\right)$, respectively.

We assume that the maximum values of problems (4) and (5) are $P^*_{(4)}$ at $\{\mathbf{w}^*, \theta^*_k\}$ and $P^*_{(5)}$ at $\{\tilde{\mathbf{w}}^*, \tilde{\theta}^*_k, \tilde{u}^*_k, \tilde{v}^*_k\}$, respectively. We have:

$$\begin{aligned} P^*_{(5)} &= f_5\left(\tilde{\mathbf{w}}^*, \tilde{\theta}^*_k, \tilde{u}^*_k, \tilde{v}^*_k\right) \leqslant f_4\left(\tilde{\mathbf{w}}^*, \tilde{\theta}^*_k\right) \\ &\leqslant f_4\left(\mathbf{w}^*, \theta^*_k\right) = P^*_{(4)} \end{aligned} \tag{26}$$

due to (5a), (5b), and (5c). Conversely, we have:

$$\begin{aligned} P^*_{(4)} &= f_4\left(\mathbf{w}^*, \theta^*_k\right) = f_5\left(\mathbf{w}^*, \theta^*_k, u^\dagger_k, v^\dagger_k\right) \\ &\leqslant f_5\left(\tilde{\mathbf{w}}^*, \tilde{\theta}^*_k, \tilde{u}^*_k, \tilde{v}^*_k\right) = P^*_{(5)} \end{aligned} \tag{27}$$

where

$$u^\dagger_k = \left|\left(\mathbf{t}^H_k + \mathbf{h}^H_k \mathbf{\Psi R G}\right)\mathbf{w}^*\right|^2$$

and

$$v^\dagger_k = \sigma^2_z \left\|\mathbf{h}^H_k \mathbf{\Psi R}\right\|^2 + \frac{\delta^2_k}{\theta^*_k}, \forall k \in \mathcal{K}$$

Therefore, we conclude $P^*_{(4)} = P^*_{(5)}$.

References

[1] I. Krikidis *et al.*, "Simultaneous Wireless Information and Power Transfer in Modern Communication Systems", *IEEE Communications Magazine*, vol. 52, pp. 104–110, 2014 (https://doi.org/10.1109/MCOM.2014.6957150).

[2] C. Psomas *et al.*, "Wireless Information and Energy Transfer in the Era of 6G Communications", *Proceedings of the IEEE*, vol. 112, pp. 764–804, 2024 (https://doi.org/10.1109/JPROC.2024.3395178).

[3] T.D.P. Perera *et al.*, "Simultaneous Wireless Information and Power Transfer (SWIPT): Recent Advances and Future Challenges", *IEEE Communications Surveys & Tutorials*, vol. 20, pp. 264–302, 2017 (https://doi.org/10.1109/COMST.2017.2783901).

[4] Z. Kang, C. You, and R. Zhang, "Active-IRS-aided Wireless Communication: Fundamentals, Designs and Open Issues", *IEEE Wireless Communications*, vol. 31, pp. 368–374, 2024 (https://doi.org/10.1109/MWC.003.2300013).

[5] Z. Peng *et al.*, "Beamforming Optimization for Active RIS-aided Multiuser Communications with Hardware Impairments", *IEEE Transactions on Wireless Communications*, vol. 23, pp. 9884–9898, 2024 (https://doi.org/10.1109/TWC.2024.3367131).

[6] R. Patra and A. Mahapatro, "Reconfigurable Intelligent Surface-based Propagation Control in FBMC/OQAM Systems", *Journal of Telecommunications and Information Technology*, vol. 95, pp. 83–90, 2024 (https://doi.org/10.26636/jtit.2024.1.1326).

[7] J. Tang *et al.*, "Energy Efficiency Optimization for a ultiuser IRS-aided MISO System with SWIPT", *IEEE Transactions on Communications*, vol. 71, pp. 5950–5962, 2023 (https://doi.org/10.1109/TCOMM.2023.3296631).

[8] M.R. Camana, C.E. Garcia, and I. Koo, "Rate-splitting Multiple Access in a MISO SWIPT System Assisted by an Intelligent Reflecting Surface", *IEEE Transactions on Green Communications and Networking*, vol. 6, pp. 2084–2099, 2022 (https://doi.org/10.1109/TGCN.2022.3196048).

[9] Q. Wu and R. Zhang, "Weighted Sum Power Maximization for Intelligent Reflecting Surface Aided SWIPT", *IEEE Wireless Communications Letters*, vol. 9, pp. 586–590, 2019 (https://doi.org/10.1109/LWC.2019.2961656).

[10] P.V. Tuan, M.T. Le, T.T. Duy, and I. Koo, "An SDR-based Approach for Efficient Simultaneous Wireless Information Power Transfer with Active IRS Assistance", *Proc. of International Conference on Industrial Networks and Intelligent Systems*, pp. 57–66, 2025 (https://doi.org/10.1007/978-3-032-02362-9_5).

[11] Y. Shen *et al.*, "Outage Constrained Max-min Secrecy Rate Optimization for IRS-aided SWIPT Systems with Artificial Noise", *IEEE Internet of Things Journal*, vol. 11, pp. 9814–9828, 2023 (https://doi.org/10.1109/JIOT.2023.3325362).

[12] J. Liu *et al.*, "Energy Efficiency in Secure IRS-aided SWIPT", *IEEE Wireless Communications Letters*, vol. 9, pp. 1884–1888, 2020 (https://doi.org/10.1109/LWC.2020.3006837).

[13] Z. Li *et al.*, "Joint Beamforming Design and Power Splitting Optimization in IRS-assisted SWIPT NOMA Networks", *IEEE Transactions on Wireless Communications*, vol. 21, pp. 2019–2033, 2021 (https://doi.org/10.1109/TWC.2021.3108901).

[14] M.B. Goktas, Y. Dursun, and Z. Ding, "IRS and SWIPT-assisted Full-duplex NOMA for 6G umMTC", *IEEE Transactions on Green Communications and Networking*, vol. 7, pp. 1957–1970, 2023 (https://doi.org/10.1109/TGCN.2023.3289505).

[15] Y. Gao *et al.*, "Beamforming Optimization for Active Intelligent Reflecting Surface-aided SWIPT", *IEEE Transactions on Wireless Communications*, vol. 22, pp. 362–378, 2022 (https://doi.org/10.1109/TWC.2022.3193845).

[16] Q. Zhai, L. Dong, Y. Li, and W. Cheng, "Secure Communications via Active IRS Assisted SWIPT NOMA Networks", *IEEE Systems Journal*, vol. 18, pp. 1032–1043, 2024 (https://doi.org/10.1109/JSYST.2024.3365590).

[17] A.B. Gershman *et al.*, "Convex Optimization-based Beamforming", *IEEE Signal Processing Magazine*, vol. 27, pp. 62–75, 2010 (https://doi.org/10.1109/MSP.2010.936015).

[18] A.M. Elbir, K.V. Mishra, S.A. Vorobyov, and R.W. Heath, "Twenty-five Years of Advances in Beamforming: From Convex and Nonconvex Optimization to Learning Techniques", *IEEE Signal Processing Magazine*, vol. 40, pp. 118–131, 2023 (https://doi.org/10.1109/MSP.2023.3262366).

[19] H. Han *et al.*, "Reconfigurable Intelligent Surface Aided Power Control for Physical-layer Broadcasting", *IEEE Transactions on Communications*, vol. 69, pp. 7821–7836, 2021 (https://doi.org/10.1109/TCOMM.2021.3104871).

[20] A. Mohamed, A. Zappone, and M. Di Renzo, "Bi-objective Optimization of Information Rate and Harvested Power in RIS-aided SWIPT Systems", *IEEE Wireless Communication Letter,* vol. 11, pp. 2195–2199, 2022 (https://doi.org/10.1109/LWC.2022.3196906).

[21] S. Zargari, A. Hakimi, C. Tellambura, and S. Herath, "Multiuser MISO PS-SWIPT Systems: Active or Passive RIS?", *IEEE Wireless Communications Letters*, vol. 11, pp. 1920–1924, 2022 (https://doi.org/10.1109/LWC.2022.3187671).

[22] Z. Yang and Y. Zhang, "Optimal SWIPT in RIS-aided MIMO Networks", *IEEE Access*, vol. 9, pp. 112552–112560, 2021 (https://doi.org/10.1109/ACCESS.2021.3099698).

[23] S. Faramarzi *et al.*, "Energy Efficient Design of Active STAR-RIS-aided SWIPT Systems", *IEEE Transactions on Wireless Communications*, vol. 24, pp. 3209–3224, 2025 (https://doi.org/10.1109/TWC.2025.3528959).

[24] C. You and R. Zhang, "Wireless Communication Aided by Intelligent Reflecting Surface: Active or Passive?", *IEEE Wireless Communications Letters*, vol. 10, pp. 2659–2663, 2021 (https://doi.org/10.1109/LWC.2021.3111044).

[25] Z. Zhang *et al.*, "Active RIS vs. Passive RIS: Which Will Prevail in 6G?", *IEEE Transactions on Communications*, vol. 71, pp. 1707–1725, 2023 (https://doi.org/10.1109/TCOMM.2022.3231893).

[26] R. Zhang and C.K. Ho, "MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer", *IEEE Transactions on Wireless Communications*, vol. 12, pp. 1989–2001, 2013 (https://doi.org/10.1109/TWC.2013.031813.120224).

[27] Y. Wu *et al.*,"Analytical Design Method of Multiway Dual-band Planar Power Dividers with Arbitrary Power Division", *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 3832–3841, 2010 (https://doi.org/10.1109/TMTT.2010.2086712).

[28] Agilent 11667A,B power splitters, Agilent RF and Microwave Test Accessories, (https://www.testequipmenthq.com/datasheets/Keysight-11667A-Datasheet.pdf).

[29] C. He *et al.*, "A Novel Dual-band Power Divider with Controllable Power Ratio and Phase Difference", *IEEE Access*, vol. 12, pp. 153146–153153, 2024 (https://doi.org/10.1109/ACCESS.2024.3480134).

[30] S. Trovarello, G. Paolini, D. Masotti, and A. Costanzo, "A Modular System of Rectifiers for Energy Harvesting with Wide Dynamic Input-range", *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*, Bol and Split, Croatia, 2021 (https://doi.org/10.23919/SpliTech52315.2021.9566459).

[31] E. Boshkovska, D.W.K. Ng, N. Zlatanov, and R. Schober, "Practical Non-linear Energy Harvesting Model and Resource Allocation for SWIPT Systems", *IEEE Communications Letters*, vol. 19, pp. 2082–2085, 2015 (https://doi.org/10.1109/LCOMM.2015.2478460).

[32] Q. Wu and R. Zhang, "Joint Active and Passive Beamforming Optimization for Intelligent Reflecting Surface Assisted SWIPT Under QoS Constraints", *IEEE Journal on Selected Areas in Communications*, vol. 38, pp. 1735–1748, 2020 (https://doi.org/10.1109/JSAC.2020.3000807).

[33] C. Psomas *et al.*, "Wireless Information and Energy Transfer in the Era of 6G Communications", *Proceedings of the IEEE*, vol. 112, pp. 764–804, 2024 (https://doi.org/10.1109/JPROC.2024.3395178).

[34] R. Long, Y.-C. Liang, Y. Pei, and E.G. Larsson, "Active Reconfigurable Intelligent Surface Aided Wireless Communications", *IEEE Transactions on Wireless Communications*, vol. 20, pp. 4962–4975, 2021 (https://doi.org/10.1109/TWC.2021.3064024).

[35] Z.-Q. Luo *et al.*, "Semidefinite Relaxation of Quadratic Optimization Problems", *IEEE Signal Processing Magazine*, vol. 27, pp. 20–34, 2010 (https://doi.org/10.1109/MSP.2010.936019).

[36] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.2", CVX Research, 2020 (https://cvxr.com/cvx/).

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 727 p., 2004 (ISBN 9780521833783).

[38] A.H. Phan, H.D. Tuan, H.H. Kha, and D.T. Ngo, "Nonsmooth Optimization for Efficient Beamforming in Cognitive Radio Multicast Transmission", *IEEE Transactions on Signal Processing*, vol. 60, pp. 2941–2951, 2012 (https://doi.org/10.1109/TSP.2012.2189857).

[39] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming", *IEEE Transactions on Wireless Communications*, vol. 18, pp. 5394–5409, 2019 (https://doi.org/10.1109/TWC.2019.2936025).

[40] P.V. Tuan, V.-Q.-B. Ngo, P. N. Son, and V.H. Vu, "An Energy Harvesting Fairness Maximization for Active/passive IRS-assisted PS-SWIPT System", *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 49, pp. 215–224, 2025 (https://doi.org/10.1504/IJAHUC.2025.147751).

Pham Viet Tuan, Ph.D.
Faculty of Physics
https://orcid.org/0000-0002-7060-2272
E-mail: phamviettuan@dhsphue.edu.vn
E-mail: pvtuan@huenuni.edu.vn
University of Education, Hue University, Hue City, Vietnam
https://dhsphue.edu.vn

Hoang Dai Long, Ph.D.
Faculty of Electronics, Electrical Engineering and Material Technology
https://orcid.org/0000-0003-4056-6570
E-mail: hdlong@hueuni.edu.vn
University of Sciences, Hue University, Hue City, Vietnam
https://husc.edu.vn

Pham Ngoc Son, Associate Professor
Faculty of Electrical and Electronics Engineering
https://orcid.org/0000-0002-9698-4221
E-mail: sonpndtvt@hcmute.edu.vn
Ho Chi Minh City University of Technology and Engineering, Ho Chi Minh City, Vietnam
https://en.hcmute.edu.vn

Mai T.P. Le, Associate Professor
Faculty of Electronics and Telecommunication Engineering
https://orcid.org/0000-0001-9657-5172
E-mail: lpmai@dut.udn.vn
University of Science and Technology, The University of Danang, Da Nang City, Vietnam
https://en.dut.udn.vn

Determining Speed and Reliability of Transmitting Indicator Information in Residual Class Systems

Matin Hadzhyiev, Nick Odegov, and Dmytro Stepanov

State University of Intelligent Technologies and Telecommunications, Odesa, Ukraine

https://doi.org/10.26636/jtit.2026.1.2403

Abstract — The article is devoted to the analysis and assessment of the efficiency of fiber optic systems whose primary objective is to transmit data. The efficiency of information transmission systems depends on numerous indicators, such as interference immunity, speed, energy efficiency, cost, development time, and design. However, from the user's point of view, quality of service is determined primarily by transmission speed and reliability. This article compares two systems. The first corresponds to the modern paradigm: one user – one transmission channel. In the other, the number of users is provided with a complex channel for transmitting symbols of the alphabet of a certain system of residual classes. At the same time, the transmission speed in the residual class system – compared to the classical multiplexing method – decreases slightly to 28/32, while the reliability (determined based on the probability of failures) increases by several orders of magnitude. The work proves a lemma that determining the optimal alphabets of residual class systems allows to optimally approximate the transmission speed of modules to binary coding systems. An analysis of the non-linear loss function, which considers the speed parameters and probabilistic reliability indicators, is performed as well.

Keywords — efficiency criteria, fiber optic transmission systems, reliability, residual class systems, transmission speed

1. Introduction

Joint consideration of energy efficiency (with such factors as reliability and speed taken into consideration) is of paramount importance in various technical systems. Speed and reliability may be contradictory requirements, while losses depend on other factors as well. For example, if a vehicle is operated at a supercritical speed over a significant amount of time, then it is logical that its reliability will decrease. On the contrary, if the vehicle is equipped with a heavy duty, damage-resistant suspension system, this will lead to a potential decrease in its maximum speed.

This work considers both transmission speed and reliability in analyzing the efficiency of fiber optic transmission systems (FOTS). The principles presented in this article can be applied to any transmission system. However, FOTS was chosen as an example solution, as this technology is used for transmitting the largest amounts of traffic and such systems may rely on a very large number of different multiplexing methods.

It should be noted that modern global data transmission systems are based on a hierarchical architecture: international traffic exchange nodes constitute the top level of the system. Below there are national node subsystems, operator nodes, and so on, all the way to the "last mile", i.e. to the end user taking advantage of a given service. Channels connecting these nodes at all levels of the network must be reliable, meaning they need to minimize the probability of failures and ensure sufficient data transfer speeds for a given network node.

For a FOTS channel, a failure is the worst-case scenario, as it is tantamount to a complete loss of efficiency.

The classic approach to information distribution consists in providing each user, temporarily, with a separate data transmission channel. This principle is referred to as the classical principle of multiplexing (CPM). In this case, a failure of one transmission channel does not exert any impact on most other users.

However, the specific user affected by the failure may be extremely dissatisfied with the service. The relevance of information may be lost due the insufficient quality of service, which can lead to discontinuation of subscription. If the channel affected by the failure is used for transmitting operational information of national importance, the problem of minimizing the probability of failure becomes particularly significant.

The aim of this work is to develop an approach that combines FOTS-specific principles, according to which efficiency is analyzed based on such criteria as transmission speed and reliability. To achieve the goal, we study the transmission speed and failure probability of a complex channel used for transmitting symbols of the alphabet of a certain system of residual classes.

2. Literature Review

Efficiency, determined based on a set of probabilistic reliability and transmission speed (bandwidth) indicators, was not considered in depth in recent research dealing with FOTS. Reliability [1], [2] and bandwidth [3]–[5] were considered in different contexts, as mutually independent factors. Only work [6] took a comprehensive approach to assessing the ef-

ficiency of FOTSs, as it considered the criteria of reliability and bandwidth, and thus may be considered relevant.

Paper [6] presented a framework in which FOTS channels were multiplexed the logic of residual class systems (RCS). In this concept, the residues from different modules are transmitted over different channels (physical or logical). This method increases the reliability of FOTS, but achieves worse results for CPMs-based systems on binary arithmetic (SBA) logic in total bandwidth.

3. Transmission Methods in RCS

The research presented in this work is based on the theory of timer signal structures (TSS). TSS was initially developed for the purpose of transmitting signals using RCS-based sequences [7]–[8].

The use of RCS allows to increase cryptographic protection of transmission systems. This is caused by the fact that no codes are used in RCS, unlike in systems based on the transmission of SBA codes. However, at the channel level and in the subsystems tasked with routing and processing information flows, modern transmission systems are based on SBA in some way or another. Therefore, the problem of reducing the redundancy of RCS, compared to SBA, is becoming increasingly important.

When analyzing RCS, the focus is placed on the Chinese remainder theorem (CRT). CRT considers systems of mutually prime integers $m_1, m_2, \ldots, m_M$, which are called RCS modules. The capacity of the RCS alphabet is defined as the following product: $E = m_1 \cdot m_2, \ldots, m_M$. The property of the CRT is that any number N within range $0 \leqslant N < E$ can be uniquely represented by the remainder of dividing this number N by the module.

However, due to the difference between RCS and SBA systems (SBA is a positional system, while RCS is not) during the transmission of RCS symbols in SBA, a certain number of additional bits is required, which can be determined for a separate module:

$$b(m) = \begin{cases} \log_2(m), & \text{fract}[\log_2(m)] \equiv 0 \\ \log_2(m) + 1, & \text{fract}[\log_2(m)] > 0 \end{cases}, \qquad (1)$$

where $\text{fract}[\cdot]$ determines the fractional part of a number.

The proof of Eq. (1) is simple and follows from the fact that only for modules that are powers of two do the numbers of bits in the RCS and the SBA coincide. Therefore, there can be only one module that is an even number in the RCS. Information redundancy of the RCS with M modules, in the general case, will be approximately m bits per transmitted symbol.

For an SBA system with base n, the alphabet contains $D = 2n$ possible messages. Compared to SBA, a number of "residual" bits $\Delta B = M$ is added to the RCS if there are no even numbers between the RCS modules, if one of the modules is an even number, $\Delta B = M - 1$. In this case, the capacity of the RCS alphabet E, in the general case, is less than the capacity of the SBA alphabet D. Otherwise, if $E > D$, then an optimization task, based on the minimum redundancy in terms of bits lost, is necessary. Simultaneously, the optimal RCS should transmit the minimum bit sequence and the number of symbols that are as close as possible to the D value.

Based on these considerations, the RCS optimization problem can be formulated as a system based on the following criteria:

$$\Delta B \to \min; \;\; \Delta E = D - E \to \min; \;\; M > 1, \qquad (2)$$

where ΔB is bit redundancy and ΔE stands for information deficiency.

The number of RCS alphabets that satisfy condition (2) is very limited. This conclusion is proved in the following manner.

Let us consider an RCS system with base n.

If n is an even number ($n = 2k$), then the optimal RCS in the sense of criterion (2) will be:

$$m_1 = 2^k - 1, \;\; m_2 = 2^k + 1. \qquad (3)$$

If n is an odd number ($n = 2k + 1$), then the optimal RCS is the following.

$$m_1 = 2, \;\; m_2 = 2^k - 1, \;\; m_3 = 2^k + 1. \qquad (4)$$

The proof is in:

$$E = (2^k - 1) \cdot (2^k - 1) = 2^n - 1 = D - 1.$$

Compared to the SBA alphabet, the RCS alphabet has only one symbol of information deficiency. Therefore, if $E = D$, then the RCS alphabet will include modules that are not mutually prime. Modules m_1 and m_2 are mutually prime numbers, since the remainder of the division by will never assume the value of 0, since:

$$\frac{m_2}{m_1} = \frac{m_1 + 2}{m_1} = \frac{1 + 2}{m_1} \Rightarrow \frac{2}{m_1 \neq 0}.$$

In this case, we have once again 1 bit of redundancy, and the information deficiency increases to 2 symbols.

Examples of RCS characteristics according to Eqs. (3) and (4), compared with the characteristics of the SBA, are given in Tab. 1.

One may notice that the relative difference of characteristics ΔB and ΔE compared to the alphabet capacity decreases with increasing base n of the RCS. For comparative analysis of the SBA and RCS in the general case, the relative indicators of bit redundancy and information deficiency can be used in the following form:

$$br = \frac{n - 1}{n}, \quad ir = \frac{E}{D}.$$

According to the definitions of modules form Eqs. (3) and (4), the graph of these characteristics is presented in Fig. 1.

4. Comparison of FOTS Models for SBA and RCS

An example of a CRT based on RCS is presented in Fig. 2. The figure shows two physically and logically separated channels for two different data transmission users. Each user is the owner of a separate optical fiber and transmits, in parallel code, 3 bits per clock interval only. The signals are transmitted

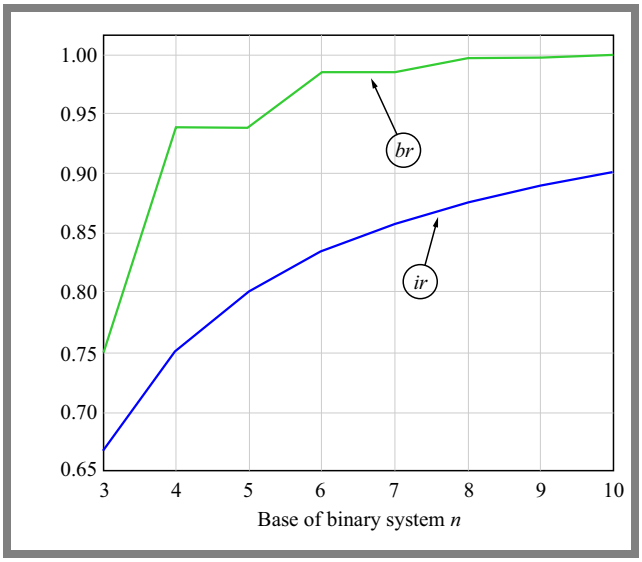

Fig. 1. Dependence of the relative indicators br and ir on SBA n.

by an elementary binary code (0, 1), and the subsequent bits of code packets B_0, B_1, B_2 are transmitted, in parallel code, over 3 frequency channels $F_{ch}0$, $F_{ch}1$, $F_{ch}2$, respectively.

In fact, each of the users has a separate physical channel (OF 1 or OF 2), in which 3 logical subchannels are organized by multiplexing in frequency domain. This scheme does not fully reflect a modern multichannel FOTS, where signals in 30 ... 90 frequency channels are transmitted simultaneously over a single optical fiber. However, only the basic principles are considered in this paper.

In the time domain, the transmission is also organized discretely, i.e. by using the clock signal. According to Fig. 2, in a CRT-based RCS system, each of the users can transmit, in one clock interval, one of the symbols of the alphabet with $D = 8$ possible values, and both users can transmit a total of 16 symbols in one clock interval.

The same multiplexing channels can be used to organize data transmission in the RCS. The schematic diagram of the RCS, which is close to the CRT-RCS from Fig. 2, is shown in Fig. 3. Here, the frequency channels and transmission channels over optical fibers are aggregated into a single common channel which is provided to both users.

Paths along two different fibers can be treated as one bit of information. Therefore, this subsystem can transmit the values of residue module 2. The frequency channels in an aggregated system can be in eight states. The RCS modulus closest to eight, which is mutually prime to 2, has the value

Tab. 1. Comparative characteristics of SBA and RCS.

n	D	m_1	m_2	m_3	E	ΔB	ΔE
4	16	3	5	–	15	1	1
5	32	2	3	5	30	1	2
6	64	7	9	–	63	1	1
7	128	2	7	9	126	1	2
8	256	15	17	–	255	1	1

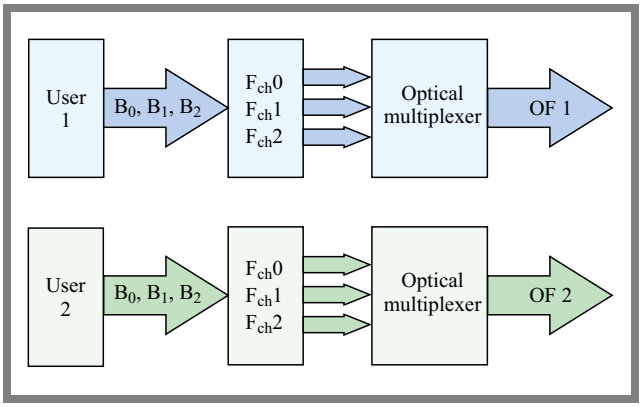

Fig. 2. Conventional system with two complex FOTS channels for a CRT-based RCS.

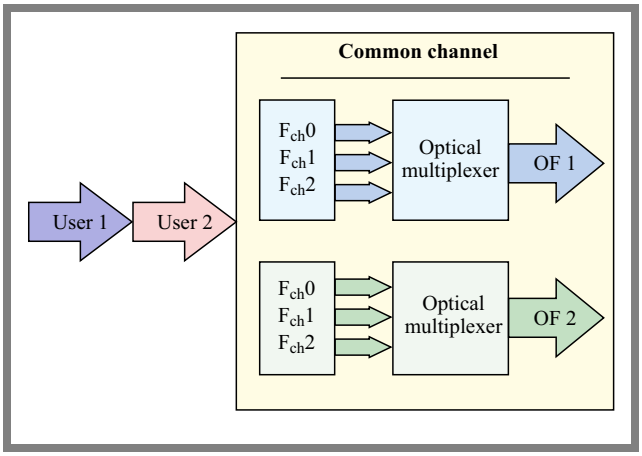

Fig. 3. System with an aggregated channel in RCS codes.

of 7. Therefore, in one clock cycle, it is possible to transmit symbols of the alphabet with module (2, 7), and thus the capacity of the alphabet is 14 variants in this case.

As a result, one RCS user transmits 14 variants in each clock cycle, while the SDC user only transmits 8. This difference does not mean that RCS is more efficient in terms of transmission speed. To compare RCS and SBA, it is necessary to consider the fact that users on the receiving end must act differently.

For the SBA case (Fig. 2), users transmit data by logically separated channels. Here, the distribution of data streams follows the TSC theory [8], where the transmission of user messages occurs sequentially over time. This method of distributing the total data stream is illustrated in Fig. 3, where users sequentially transmit their data, alternating according to FOTS clocks.

Within two clock periods, each SBA user will transmit symbols using an alphabet with a capacity of 16 symbols, while an RCS user will only transmit from an alphabet which contains 14 symbols. Accordingly, both users of the SBA can transmit one symbol out of 32 possible options in 2 clock cycles, while users of the RCS can transmit only 28. Therefore, transmitting the RCS alphabets using the SBA codes leads to losses in terms of the maximum speed achieved.

5. Comparison of FOTS Models Based on Reliability Criteria

FOTS system failures may be caused by various reasons:

- failure of frequency channels due to aging of active elements of optoelectronic subsystems,
- aging of optical fiber,
- damage to optical cable, individual fibers or patchcords,
- above-specification operation of attenuators,
- violation of optical fiber switching schemes,
- lightning strikes affecting the cable,
- intentional damage to FOTS components.

Due to the high complexity of these systems and optical components, such reasons must be analyzed during design, production, and operation of FOTS. Most likely, it is impossible to derive any generalized formulas for analyzing reliability at the system level. Simulation modeling methods and the determination of probabilistic reliability indicators can be helpful and effective. However, in this work, we consider a simple probabilistic model, where the probabilities of the operational state of all FOTS components are known according to the scheme shown in Fig. 2.

The reliability parameter of a separate frequency channel is defined as probability p of its operating state at a specific time during operation, for example, over a month, year, or decade. Consequently, the probability of failure in this component is: $q = 1 - p$. A malfunction of any frequency channel affecting each user means a critical service failure. In this case, the SBA signals cannot be transmitted with the base of the 3-bit binary system. Similarly, the failure of the entire channel for each user may be caused by damage to the optical fiber. In this case, the reliability of the system is defined as a linear function of the reliability of frequency channels $p* = a \cdot p$, where a is the proportionality coefficient.

We assume that frequency channel failures and optical fiber damage are statistically independent. Then, according to Fig. 2, the logic of potential channel failures affecting one user is determined by the fact that specific system components, namely 3 frequency channels and 1 optical channel, are connected in series. Then, the probability of failure-free operation of the transmission channel for each user P_U and the probability of failure Q_U, can be formulated as follows:

$$P_U = a \cdot p^4, \quad Q_U = 1 - P_U \, . \tag{5}$$

In the case of RCS, the system failure logic is different and more complex. The failure model can be defined as the parallel connection of channels for the first and second users, according to Fig. 2. Then, the probability of a serviceable state P_R and the probability of failure Q_R for RCS is determined as follows:

$$\begin{aligned} P_R &= 1 - Q_U^2 = 1 - (1 - P_U)^2 = 2a\,p^4 - a^2 p^8 \\ Q_R &= 1 + a^2 p^8 - 2a\,p^4 \end{aligned} \, . \tag{6}$$

The proportionality coefficient determining the comparative reliability of frequency channels and the optical channel may not be very intuitively acceptable, since the probabilities of operating states are large. In other words, if the probability of failure of frequency channel q and optical fiber $q*$ were defined as a proportional formula $q* = b \cdot q$, then the clear description of the states of the system is described in example as: the probability of damage to the optical fiber is 10 times greater than the probability of failure of the frequency channel ($b = 10$). Moreover, coefficients a and b are related by a simple relationship, such as:

$$a = 1 + \frac{q\,(1-b)}{1-q} \, , \tag{7}$$

As a result of the fact that parameter b can vary significantly depending on the class and operating conditions of FOTS, the conventional model values of this parameter are subjected to further analysis using Eqs. (5)–(7), modeling of the probability of failures for the SBA and RCS was performed.

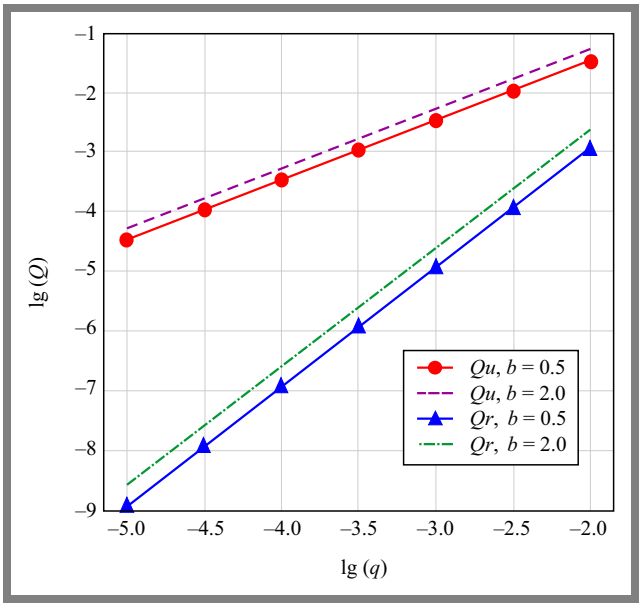

Fig. 4. Comparison of the probability of failure for SBA and RCS channels.

Figure 4 shows two examples of the probability failure of frequency channels and failures along the optical fiber, when $b = 0.5$ and $b = 2$. Analysis of these graphs shows that the failure probabilities of the SBA are significantly higher when compared to the RCS and differ by orders of magnitude. In this case, for the SBA, $Q_U > q$, while for the RCS, $Q_R < q$ is always true.

6. Effectiveness of FOTS in Terms of Speed and Reliability

The effectiveness of FOTS in terms of service quality is perceived differently by operators and users. For operators, a decrease in quality may result in lost subscribers or corporate clients. For users, a decrease in quality of service depends on the reduction in the transmission speed V (percentage-wise) compared to the maximum achievable value $Vmax$. Therefore, the speed reduction coefficient ($K = Vmax/V$)

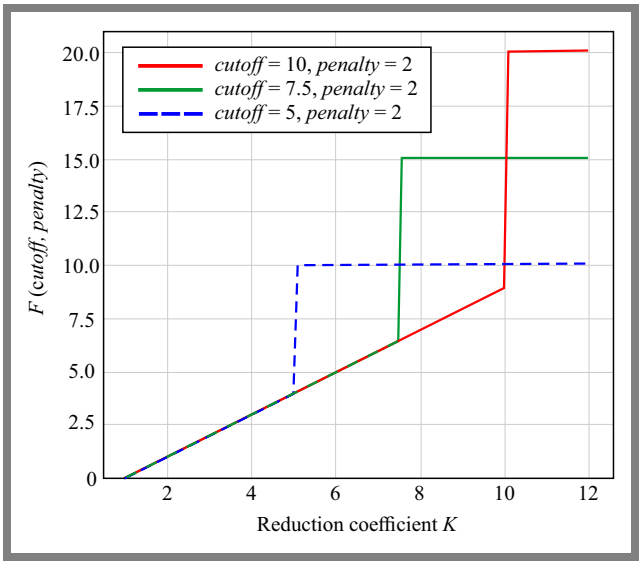

Fig. 5. Examples of penalty functions $F(V)$.

can be considered one of the important quality measures. The decrease in quality can be characterized by various factors, e.g., penalty function $F(V)$, with $F(Vmax) = 0$.

Some operators base their sales policies on the type of quality of service (QoS). In such a business model, the user enjoys a guaranteed transmission speed V that is expected to remain within specific limits ($Vmin \leqslant V \leqslant Vmax$). Speed is reduced to the guaranteed minimum level $Vmin$ during periods in which peak network loads are experienced, for instance. In a situation in which the contract with the operator provides for a minimum transfer speed value $cutoff = Vmin$, when the transmission rate falls below that threshold, the user may perceive this to constitute a critical failure. If the cutoff level is exceeded, an increase factor sanction > 1 is added to the penalty. For example, the penalty function corresponds to the following conditions:

$$F(V) = \begin{cases} \dfrac{Vmax}{V}, & V \leqslant cutoff \\ \dfrac{penalty \cdot Vmax}{Vmin}, & V > cutoff \end{cases} . \qquad (8)$$

An example of $F(V)$ is shown in Fig. 5, where the value of the speed reduction coefficient K is plotted on the x-axis.

Using Eq. (8), an integral indicator of efficiency losses is formulated, which takes into account the total penalties over a given period of operation:

$$L(F,t) = \frac{1}{T}\int_0^T F\big(V(t)\big)\,\mathrm{d}t\,. \qquad (9)$$

Equations (8) and (9) are suitable for assessing the efficiency of a FOTS system during operation and may serve as a foundation for analysis using simulation modeling methods. Then, considering the statistics concerning potential failures and the lead time required to restore the operation of the system after incidents, Eqs. (8) and (9) are extended by applying appropriate weighting factors that characterize the probabilities of the serviceable state and the state of parametric or critical failures.

7. Results and Discussion

Solving optimization problems using multiple criteria that may contradict each other requires using non-trivial approaches, especially in problems that require the development of non-standard indicators based on various informal requirements for the system.

This article proposes the basics of a methodology used for a comprehensive analysis of the efficiency of FOTS from the point of view of speed and reliability. The method is based on the theory of timer signal structures, the theory of reliability, and the arithmetic principles of residual class systems.

For the transmission of residues of different modules in the RCS, it is proposed to use different multiplication methods. Here, the simplest option is considered, where residues module 2 are transmitted over two different optical fibers (spatial multiplexing), and residues modulo 3 are transmitted using frequency multiplexing.

The analysis conducted shows that the transmission speed is in the RCS. Therefore, in the development of the theory of timer signal structures, optimal RCS have been determined, which are optimally close to the SBA based on the speed criterion. For the example considered, the ratio of the RCS and SBA transmission speed in bits was found to equal 28/32 bits per clock interval. Under equal conditions, the RCS according to the reliability criterion exceeds the RCS by orders of magnitude. To solve the complex problem of evaluating the effectiveness of FOTS in terms of quality of service, a non-linear penalty function has been proposed.

This study shows that the use of RCS relying on various multiplexing methods can promise a significant improvement in probabilistic reliability indicators of FOTS, with a very slight loss of transmission speed.

References

[1] O.V. Bondarenko *et al.*, "Thermomechanical Influence on Optical Fibers during the Production and Operation of Optical Cable", *Scientific Works of ONAS*, pp. 23–32, 2019 (https://doi.org/10.33243/2518-7139-2019-1-1-23-32).

[2] D.M. Stepanov, "Development of the Method of Improving the Selection of Dielectrical Strength Elements of Optical Cable Design", *Herald of Khmelnytskyi National University. Technical Sciences*, vol. 343, pp. 315–322, 2024 (https://heraldts.khmnu.edu.ua/index.php/heraldts/en/article/view/570).

[3] N.A. Odegov and I.S. Baleyev, "Number-analytical Method for the Synthesis of Optimal Irregular DWDM Frequency Plans", *Scientific Works of ONAS*, pp. 68–80, 2020 (https://doi.org/10.33243/2518-7139-2020-1-2-70-81).

[4] N.A. Odegov, "Evaluation of the Limit Transmission Speed in the Magistral Fiber Optical System", *Scientific Works of ONAS*, pp. 61–76, 2019 (https://doi.org/10.33243/2518-7139-2019-1-2-61-76).

[5] N.A. Odegov *et al.*, "Algorithm for The Synthesis of Optimal FOTS Frequency Plans for Nonlinear Transmission Environment (Boat Method)", *Herald of Khmelnytskyi National University. Technical Sciences*, vol. 327, pp. 88–97, 2023 (https://heraldts.khmnu.edu.ua/index.php/heraldts/en/article/view/509).

[6] D. Bahachuk *et al.*, "Multiplex Technique of Data Transmission in Residual Class Systems", *Eastern-European Journal of Enterprise*

Technologies, vol. 6, pp. 23–31, 2023 (https://doi.org/10.15587/1729-4061.2023.292504).

[7] M.V. Zakharchenko *et al.*, "Data Transmission Systems: Effectiveness of Block Coding", *Scientific Works of ONAS*, 2014 (https://method.onat.edu.ua/download/555).

[8] M. Zakharchenko *et al.*, "Comparison of Positional and Timer Coding in the System of Residual Classes", *2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering*, Lviv-Slavske, Ukraine, 2022 (https://doi.org/10.1109/TCSET55632.2022.9766866).

Matin Hadzhyiev, D.Sc., Professor
Department of Software Engineering
https://orcid.org/0000-0001-7280-3863
E-mail: gadjievmm@ukr.net
State University of Intelligent Technologies and Telecommunications, Odesa, Ukraine
https://en.suitt.edu.ua

Nick Odegov, Ph.D., Assoc. Professor
Department of Software Engineering
https://orcid.org/0000-0001-5526-2487
E-mail: onick_64@ukr.net
State University of Intelligent Technologies and Telecommunications, Odesa, Ukraine
https://en.suitt.edu.ua

Dmytro Stepanov, Ph.D., Assoc. Professor
Department of Electronic Communications Switching Systems
https://orcid.org/0009-0002-8299-5402
E-mail: dmstepanov@ukr.net
State University of Intelligent Technologies and Telecommunications, Odesa, Ukraine
https://en.suitt.edu.ua

DOA Estimation of Linear Dipole Arrays Based on Horse Herd Optimization Algorithm

Mohamed Bensalem and Ouarda Barkat

University of Freres Mentouri Constantine 1, Constantine, Algeria

https://doi.org/10.26636/jtit.2026.1.2422

Abstract — Subspace-based direction of arrival (DOA) estimation algorithms, such as MUSIC and ESPRIT, are designed for adaptive smart antenna arrays. However, these subspace methods require a large number of signal snapshots and sufficient angular separation between signals to provide an accurate DOA estimation of RF signal sources. Moreover, their resolution degrades significantly in severe noise scenarios. This study proposes a swarm intelligence (SI) algorithm, known as horse herd optimization (HOA), to address these limitations. An optimizer is employed as a direction-finding method to estimate the directions of arrival (DoAs) of incident signals impinging on a linear array of half-wavelength dipole (HWD) antennas by examining the global minimum of a non-linear cost function. This cost function is defined as the difference between the actual and estimated angles and is used to evaluate candidate solutions. Simulation results of the proposed algorithm have been compared with other recognized algorithms, including ESPRIT, root-MUSIC, and PSO, to verify estimation accuracy, convergence behavior, robustness against the number of elements, noise, and snapshots over Monte Carlo trials. It has been observed that the suggested HOA achieves better performance with a few snapshots, outperforms PSO and subspace-based methods when it comes to estimating DOA of incoming signals, particularly in a low signal-to-noise ratio (SNR) environment, and even when only fewer snapshots are available.

Keywords — DOA estimation, horse herd optimization, linear dipole arrays

1. Introduction

Over the past two decades, we have witnessed impressive growth in wireless communication systems. In light of this growth, advanced-generation wireless communication technologies, known as beyond 5G (B5G) or 6G networks, are emerging as powerful solutions to enhance system capacity, ensure reliable connectivity, and increase spectral efficiency [1], [2]. In such systems, smart dipole array antennas play a crucial role in supporting higher data rates and wider bandwidths, while enhancing system performance and capacity by increasing the gain in the desired direction of arrival (DOA) of the signal [3]–[5].

In general, a smart antenna system consists of an array of antenna elements arranged following a specific geometric pattern and equipped with a signal processor to accomplish two essential tasks: DOA estimation of all incoming signals and computation of optimal weights for adaptive beamforming [6], [7]. Before the beamforming process, the DOAs must be accurately estimated from the information received by the antenna array elements using an appropriate DOA estimation algorithm. The estimated directions then enable the beamforming technique to extract the signal of interest in the presence of interference and noise by forming a narrow beam toward the signal of interest (SOI) and placing nulls toward interfering signals, referred to as signals not of interest (SNOIs). Therefore, DOA estimation is a critical process in smart antenna systems, as it provides accurate directional information about signal sources in wireless communication networks [8], [9].

Estimation is a research domain in array signal processing, and DOA algorithms are an important part of this field. In general, these algorithms can be classified into two categories: those of deterministic and stochastic variety [10]. Although they differ in terms of accuracy, computational complexity, and convergence speed, several deterministic DOA estimation algorithms have been investigated in the literature for adaptive smart antenna arrays.

These algorithms can be mainly divided into three classes: conventional methods, subspace-based methods, and maximum likelihood techniques [11]. Owing to their simplicity and ability to accurately and efficiently resolve incoming signals, subspace-based DOA algorithms, such as multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT), have achieved significant success within the scientific community. These high-resolution algorithms share similar underlying principles [12]–[14]. They estimate the DOA by applying eigen-decomposition to the sample covariance matrix (SCM) of the received signals, exploiting orthogonality between the signal and noise subspaces. However, ESPRIT relies solely on the signal subspace.

Despite their advantages, these algorithms fail to provide accurate estimates in challenging scenarios, such as low signal-to-noise ratio (SNR) conditions and closely spaced sources [9], [15]. Moreover, they require a large number of signal snapshots to reliably estimate the SCM [9], which may not always be available in dynamic environments. Therefore, stochastic algorithms are employed to address these limitations.

Stochastic algorithms are predominantly metaheuristic methods. Techniques of this kind are generally flexible, provide reliable solutions, are easy to implement and are adaptable to a wide range of optimization problems [16], [17]. High-level metaheuristic search methods explore the search space randomly to avoid local optima and to identify the global optimum solution. They are typically inspired by natural behaviors observed in humans and animals, physical phenomena, evolutionary concepts, or other optimization paradigms [18], [19]. For this reason, they are referred to as nature-inspired metaheuristic algorithms and are commonly classified into four main categories according to their sources of inspiration: evolution-based, physics-based, swarm-based, and human-based algorithms [18], [20].

The swarm intelligence (SI) algorithms have gained significant popularity among researchers in recent years. These algorithms mimic the collective behavior of animals by employing multiple interacting agents that cooperate to solve optimization problems. The search process usually begins with a population of randomly generated candidate solutions which are subsequently refined through an iterative improvement process.

In the field of array signal processing (ASP), several SI algorithms have been proposed for DOA estimation, e.g. particle swarm optimization (PSO) [21], genetic algorithms (GA) [22], ant colony optimization (ACO) [23], bee colony optimization (BCO) [24], flower pollination algorithm (FPA) [25], squirrel search algorithm (SSA) [26], grey wolf optimization (GWO) [27], and chicken swarm optimization (CSO) [28].

SI algorithms typically begin with a population of randomly generated candidate solutions. Each algorithm employs a fitness function to evaluate the population, update candidate solutions, and search for the optimal solution using stochastic mechanisms.

According to the no free lunch (NFL) theorem [29], no single algorithm is capable of solving all optimization problems effectively, despite the wide diversity of existing metaheuristic approaches. Consequently, the primary challenge for these algorithms lies in achieving an appropriate balance between exploration and exploitation throughout the search process.

A recently proposed algorithm, known as the horse herd optimization algorithm (HOA), is inspired by the natural behaviors of horses belonging to different age groups [30]–[32]. This algorithm has demonstrated excellent performance in solving several real-world optimization problems. Owing to its effective exploration–exploitation strategy, HOA shows strong potential for addressing complex optimization problems across various application domains.

It is important to emphasize that the proposed approach strictly follows the conventional mathematical formulation of DOA estimation, while HOA is employed solely as a numerical global optimization tool to solve the resulting non-linear optimization problem. In this framework, each candidate solution represents a possible set of DOA estimates, and the optimization process aims at minimizing a signal-model-based cost function that quantifies the mismatch between the received signals and their reconstructed counterparts.

The novelty of this paper lies in the use of a HOA metaheuristic algorithm that is deployed as an estimator for determining the DOAs of multiple incident signals. Performance of the proposed algorithm is evaluated using simulation and measurements, with a particular emphasis placed on estimation accuracy, resolution capability, robustness to noise, robustness with respect to the number of snapshots through Monte Carlo simulations, and convergence speed.

Simulation results show that the proposed HOA-based estimator achieves excellent DOA estimation performance under extremely challenging conditions, including closely spaced sources, a limited number of snapshots, and severe noise, when applied to a linear array of HWD antennas. Moreover, the proposed method is compared with particle swarm optimization (PSO), root-MUSIC, and ESPRIT algorithms. The results demonstrate that the HOA-based approach outperforms these conventional methods in estimating the DOAs of incident signals impinging on a linear HWD antenna array.

In addition, a computational complexity analysis is provided to clarify the practical usefulness of the proposed method. Unlike subspace-based techniques, the proposed approach does not require the computation of the spatial covariance matrix nor spectral peak searching, which contributes to a reduced computational burden.

2. Problem Formulation

In a smart antenna system (SAS), DOA estimation refers to the process of extracting the directional information of multiple electromagnetic waves from the outputs of several receiving antennas using array signal processing techniques. These techniques exploit the fact that an electromagnetic wave impinging on an antenna array reaches each array element at different time instants.

In this work, an adaptive dipole array system is considered, which represents a type of smart antenna equipped with adaptive beamforming and DOA estimation capabilities, as illustrated in Fig. 1. The system consists of an array of half-wavelength dipole antennas, complex weighting coefficients, a summation unit, and a signal processing module [6], [9]. The smart antenna system employs a dipole array combined with advanced signal processing algorithms to estimate the DOAs of the incoming signals and subsequently compute the beamforming vectors, enabling electronic steering of the array output.

2.1. Objective Function

We consider L uncorrelated narrowband source signals impinging on an array of M antennas with different time delays. These sources generate plane waves, denoted by $s_l(t)$, arriving from distinct directions $\{\theta_1, \theta_2, \ldots, \theta_L\}$, and they also include additive zero-mean Gaussian noise (AWGN). For narrowband signals, the propagation time delays across the antenna elements can be equivalently represented as phase shifts.

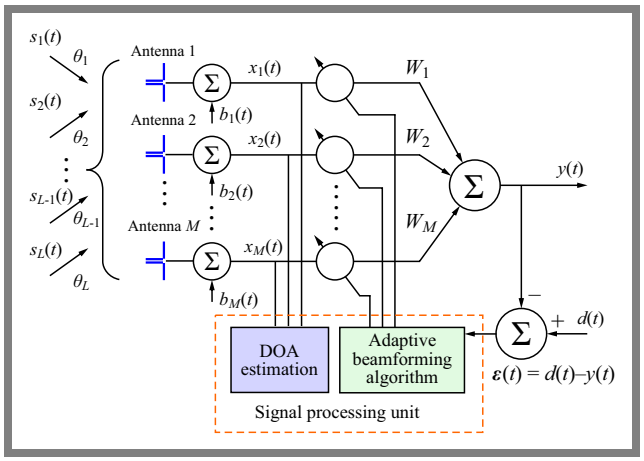

Fig. 1. Functional block diagram of an adaptive dipole array system.

Under ideal conditions, the steering vector is assumed to be perfectly known and depends solely on the array geometry and the directions of arrival. Consequently, the received signal vector of dimension $M \times 1$ can be written as:

$$x_m(t) = \sum_{i=1}^{L} e^{-j\frac{2\pi d(m-1)\sin(\theta_i)}{\lambda}} s_i(t) + b_m(t), \quad m = 1, 2, \ldots, M \,. \tag{1}$$

We can express Eq. (1) in matrix form as follows:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{s}(t) + \mathbf{b}(t)\,, \tag{2}$$

where $\bar{\mathbf{s}}(t)$ represents the original signals from the L sources with directions of arrival (DOAs) $\theta_1, \theta_2, \ldots, \theta_L$, and $\bar{\mathbf{b}}(t)$ denotes an additive noise vector whose components have zero-mean and variance σ^2.

The matrix $\bar{\mathbf{A}}(\boldsymbol{\theta})$ represents the steering vectors corresponding to the L sources for an array of M antennas, defined as:

$$\bar{\mathbf{x}}(t) = [\,x_1(t),\, x_2(t),\, \ldots,\, x_M(t)\,]^T\,, \tag{3}$$

$$\bar{\mathbf{b}}(t) = [\,b_1(t),\, b_2(t),\, \ldots,\, b_M(t)\,]^T\,, \tag{4}$$

$$\bar{\mathbf{s}}(t) = [\,s_1(t),\, s_2(t),\, \ldots,\, s_L(t)\,]^T\,, \tag{5}$$

$$\bar{\mathbf{A}}(\boldsymbol{\theta}) = [\,\bar{\mathbf{a}}(\theta_1),\, \bar{\mathbf{a}}(\theta_2),\, \ldots,\, \bar{\mathbf{a}}(\theta_L)\,]\,. \tag{6}$$

The steering vector $\bar{\mathbf{a}}(\theta_i)$ is given by:

$$\bar{\mathbf{a}}(\theta_i) = \left[1,\; e^{-j\frac{2\pi d\sin(\theta_i)}{\lambda}},\; \ldots,\; e^{-j\frac{2\pi d(M-1)\sin(\theta_i)}{\lambda}}\right]^T\,. \tag{7}$$

This study examines an array of M dipole elements, each half a wavelength long, spaced equally at a distance d. When an incoming wave carrying a baseband signal $s(t)$ impinges on the antenna array at an incident angle θ_i, the steering matrix $\bar{\mathbf{A}}(\boldsymbol{\theta})$ of the linear array corresponding to the source arriving from direction θ_i $(i = 1, \ldots, L)$ is given by:

$$\bar{\mathbf{A}}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\frac{2\pi d\sin(\theta_1)}{\lambda}} & \cdots & e^{-j\frac{2\pi d\sin(\theta_L)}{\lambda}} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi d(M-1)\sin(\theta_1)}{\lambda}} & \cdots & e^{-j\frac{2\pi d(M-1)\sin(\theta_L)}{\lambda}} \end{bmatrix}_{M\times L}\,, \tag{8}$$

where the dimension of $\bar{\mathbf{A}}(\boldsymbol{\theta})$ is an $M \times L$ matrix, with each column corresponding to one of the M directional antennas.

The spatial covariance matrix $\mathbf{R}_{xx}$ of the observed signal vector $\bar{\mathbf{x}}(t)$, based on N signal samples at the array output, is given by:

$$\bar{\mathbf{R}}_{xx} = \mathbb{E}\left[\bar{\mathbf{x}}\,\bar{\mathbf{x}}^H\right] = \frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{x}}(i)\,\bar{\mathbf{x}}^H(i)\,, \tag{9}$$

where $(\cdot)^H$ denotes the Hermitian transpose operation.

By substituting Eq. (2) into Eq. (9), and assuming that $\bar{s}(t)$ and $\bar{b}(t)$ are uncorrelated and statistically independent, the covariance matrix can be written as:

$$\bar{\mathbf{R}}_{xx} = \bar{\mathbf{A}}(\boldsymbol{\theta})\,\bar{\mathbf{R}}_{ss}\,\bar{\mathbf{A}}^H(\boldsymbol{\theta}) + \sigma_b^2\,\mathbf{I}_M\,, \tag{10}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\bar{\mathbf{R}}_{ss} = \mathbb{E}[\bar{\mathbf{s}}(t)\bar{\mathbf{s}}^H(t)]$ is the signal covariance matrix.

To solve any problem using the HOA algorithm, a fitness function has to be defined first, which will be used to evaluate the quality of possible solutions. For the purpose of DOA estimation, HOA can be employed to optimize the directions of arrival of the incident signals impinging on the antenna array by making the estimated angles approximately match the actual angles.

Accordingly, the following fitness function is applied for the evaluation, defined as:

$$f(\boldsymbol{\theta}) = \|\bar{\mathbf{x}}_a(t) - \bar{\mathbf{x}}_e(t,\boldsymbol{\theta})\|^2\,, \tag{11}$$

where $\bar{\mathbf{x}}_a(t)$ represents the actual received signal, which is expressed as:

$$\bar{\mathbf{x}}_a(t) = \bar{\mathbf{A}}(\boldsymbol{\theta}_a)\,\bar{\mathbf{s}}(t) + \bar{\mathbf{b}}(t)\,. \tag{12}$$

The estimated received signal, denoted by $\bar{\mathbf{x}}_e(t,\boldsymbol{\theta})$, can be expressed after obtaining the optimal solution as follows:

$$\bar{\mathbf{x}}_e(t,\boldsymbol{\theta}) = \bar{\mathbf{A}}(\boldsymbol{\theta})\,\hat{\mathbf{s}}(t)\,. \tag{13}$$

It should be noted that the vector $\bar{\mathbf{s}}(t)$ is unknown but common to both the actual and the estimated received signals. Its contribution is implicitly accounted for through the minimization of the fitness function $f(\boldsymbol{\theta})$. Accordingly, the SI-based estimator aims to determine the optimal DOA parameter vector $\boldsymbol{\theta}$ by searching for the values that minimize the objective function.

Therefore, the DOA estimation problem formulated in Eq. (11) can be rewritten as follows:

$$\min_{\boldsymbol{\theta}\in\mathcal{S}}\; f(\boldsymbol{\theta})\,, \tag{14}$$

where $\mathcal{S} \subset \mathbb{R}^L$ denotes the solution space.

The performance of each candidate solution is evaluated based on its corresponding fitness value. The initial population is randomly generated with a predefined population size. The HOA evaluation process is performed iteratively over T iterations using a population of N individuals until the stopping criterion is satisfied when the maximum number of iterations is reached $(t = T)$. Accordingly, the population at the t-th iteration can be represented as a collection of

L-dimensional vectors, given by:

$$\boldsymbol{\Theta}^{(t)} = \{\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, \boldsymbol{\theta}_3^{(t)}, \ldots, \boldsymbol{\theta}_N^{(t)}\} . \tag{15}$$

The HOA-based approach is well suited for solving multidimensional optimization problems. Therefore, the position $\boldsymbol{\theta}^*$ corresponding to the minimum value of the fitness function $f(\boldsymbol{\theta})$ represents the estimated directions of arrival of the impinging signals. The DOA estimate at iteration t can thus be expressed as:

$$\boldsymbol{\theta}_{best}^{(t)} = \arg \min_{\boldsymbol{\theta}_i^{(t)} \in \{\boldsymbol{\theta}^{(t)}\}} f(\boldsymbol{\theta}_i^{(t)}) . \tag{16}$$

2.2. DOA Estimation with HOA

Horse herd optimization algorithm (HOA) is a modern swarm intelligence-based metaheuristic approach proposed in [30]. This algorithm is inspired by the behavioral patterns of horses in a herd, in their natural habitat, throughout their lifetime. During different stages of their lives, horses exhibit distinct behaviors that can be categorized into six primary groups: grazing (G), hierarchy (H), sociability (S), imitation (I), defense mechanism (D), and roaming (R). Their typical lifespan is approximately 25 – 30 years [33]. In this context, horses are labeled based on their age categories: δ represents horses between 0 and 5 years old, γ represents ages from 5 to 10, β represents ages ranging from 10 to 15, while α denotes horses that are over 15 years old. The emulation of the horses movement during each iteration is described in [30].

$$\mathbf{X}_n^{(\mathrm{iter},\mathrm{age})} = \mathbf{V}_n^{(\mathrm{iter},\mathrm{age})} + \mathbf{X}_n^{(\mathrm{iter}-1,\mathrm{age})}, \quad \mathrm{age} = \alpha, \beta, \gamma, \delta , \tag{17}$$

where $\mathbf{X}_n^{(\mathrm{iter},\mathrm{age})}$ indicates the position of the n-th horse, $\mathbf{V}_n^{(\mathrm{iter},\mathrm{age})}$ denotes the velocity vector of the considered horse during movement, while $\mathrm{iter}, \mathrm{age}$ represent the current iteration and the age category of each horse, respectively.

In each iteration, horses within the herd are organized, in the response matrix, from best to worst based on their fitness values. The selection process of horse ages in each iteration is guided by the rules of the HOA algorithm according to the following criteria [30]:

- α group horses are the most experienced and achieve the best response – they comprise the top 10% of the sorted population,
- β group horses follow, representing 20% of the total horse population,
- γ and δ group horses account for 30% and 40% of the remaining population, respectively.

Six distinct horse behaviors are specifically used to identify the velocity vector. Accordingly, the velocity vector of horses at different ages during every iteration of the algorithm is formulated as [30]:

$$\mathbf{V}_n^{(\mathrm{iter},\alpha)} = \mathbf{G}_n^{(\mathrm{iter},\alpha)} + \mathbf{D}_n^{(\mathrm{iter},\alpha)}, \tag{18}$$

$$\mathbf{V}_n^{(\mathrm{iter},\beta)} = \mathbf{G}_n^{(\mathrm{iter},\beta)} + \mathbf{H}_n^{(\mathrm{iter},\beta)} + \mathbf{S}_n^{(\mathrm{iter},\beta)} + \mathbf{D}_n^{(\mathrm{iter},\beta)}, \tag{19}$$

$$\begin{aligned}\mathbf{V}_n^{(\mathrm{iter},\gamma)} = \mathbf{G}_n^{(\mathrm{iter},\gamma)} + \mathbf{H}_n^{(\mathrm{iter},\gamma)} + \mathbf{S}_n^{(\mathrm{iter},\gamma)} \\ + \mathbf{I}_n^{(\mathrm{iter},\gamma)} + \mathbf{D}_n^{(\mathrm{iter},\gamma)} + \mathbf{R}_n^{(\mathrm{iter},\gamma)},\end{aligned} \tag{20}$$

$$\mathbf{V}_n^{(\mathrm{iter},\delta)} = \mathbf{G}_n^{(\mathrm{iter},\delta)} + \mathbf{I}_n^{(\mathrm{iter},\delta)} + \mathbf{R}_n^{(\mathrm{iter},\delta)} . \tag{21}$$

In the following, we describe the primary steps related to individual and social intelligence of horses, including the mathematical model for each behavior pattern.

2.3. Grazing

As herbivorous animals, horses primarily obtain their food from plants, grasses, and forages. The grazing behavior is exhibited by horses throughout their entire lives, at all age stages. As part of their daily routine, they spend most of their time grazing on pastures, devoting up to 16 – 20 h per day to this activity [33], and rest for only a very short period. The HOA optimizer emulates the grazing region around every horse by using a coefficient denoted as g, such that every horse grazes at specific locations. Equations (19) and (20) are employed to describe the grazing behavior of horses mathematically [30].

$$\begin{aligned}\mathbf{G}_n^{(\mathrm{iter},\mathrm{age})} = & g_n^{(\mathrm{iter},\mathrm{age})} (\hat{\mathbf{u}} + \hat{\mathbf{P}}_l) \times \mathbf{X}_n^{(\mathrm{iter}-1,\mathrm{age})}, \\ & \mathrm{age} = \alpha, \beta, \gamma, \delta ,\end{aligned} \tag{22}$$

$$g_n^{(\mathrm{iter},\mathrm{age})} = g_n^{(\mathrm{iter}-1,\mathrm{age})} \times \omega_g . \tag{23}$$

Here $\mathbf{G}_n^{(\mathrm{iter},\mathrm{age})}$ refers to the parameter of describing grazing movement of the horse, which reflects the horse's proclivity to graze. This parameter decreases linearly with a reduction coefficient ω_g per iteration.

P is a randomly generated number between $[0, 1]$, while $\hat{\mathbf{l}}$ and $\hat{\mathbf{u}}$ denote the lower and upper boundaries of the grazing area, respectively, as recommended in [30], and are set to 0.95 and 1.05. Also, the coefficient g is set to 1.5 at the start of the search process for all age ranges.

2.4. Hierarchy

The second behavior of horses considered by the HOA algorithm is *hierarchy*. Horses are social animals and typically live under the guidance of a leader, similar to certain human social structures. In the wild, the herd is led by the dominant leader, who is the most experienced and strongest horse, often a mature stallion or mare, responsible for supervising and guiding the herd. In the HOA algorithm, the parameter h illustrates the proclivity of horses to pursue the most experienced and strongest horse. Studies have shown that horses in the middle age groups, β and γ, adhere to the hierarchy rule. The mathematical expressions describing this behavior are as follows [30]:

$$\begin{aligned}\mathbf{H}_n^{(\mathrm{iter},\mathrm{age})} = & h_n^{(\mathrm{iter},\mathrm{age})} \times \left(\mathbf{X}_*^{(\mathrm{iter}-1)} - \mathbf{X}_n^{(\mathrm{iter}-1)}\right) \\ & \mathrm{age} = \alpha, \beta, \gamma ,\end{aligned} \tag{24}$$

$$h_n^{(\mathrm{iter},\mathrm{age})} = h_n^{(\mathrm{iter}-1,\mathrm{age})} \times \omega_h . \tag{25}$$

Here, $\mathbf{H}_n^{(\text{iter,age})}$ signifies the level of the horse's hierarchy, indicating how the best horse's location impacts the velocity parameter. $\mathbf{X}_*^{(\text{iter}-1)}$ denotes the position of the best horse in the previous iteration. At the start of the search process, the parameter h is initialized to 1.5, 0.9, and 0.5 for horses of ages α, β, and γ, respectively, while ω_h represents the hierarchy reduction coefficient [30].

2.5. Sociability

Horses typically exhibit social behavior, allowing them to coexist with other livestock, including cattle and sheep, and they often enjoy their company rather than being solitary in their natural environment. Since the hunting habits of predators threaten the safety of wild horses, herd life significantly enhances their security, reduces their likelihood of predation, and facilitates escape from predators.

Horses may occasionally engage in disputes and can display irritability due to their individuality. Notably, the *sociability* behavior is most commonly observed in horses aged between 5 and 15 years. In the HOA algorithm, this behavior is modeled as the tendency of a horse to move toward the average position of other horses, as expressed in the following mathematical formulas [30]:

$$\mathbf{S}_n^{(\text{iter,age})} = s_n^{(\text{iter,age})} \left[\frac{1}{N} \sum_{j=1}^{N} \mathbf{X}_j^{(\text{iter}-1)} - \mathbf{X}_n^{(\text{iter}-1)} \right], \quad \text{age} \in \{\beta, \gamma\}, \tag{26}$$

$$s_n^{(\text{iter,age})} = s_n^{(\text{iter}-1,\text{age})}\, \omega_s . \tag{27}$$

In Eqs. (26) and (27), $\mathbf{S}_n^{(\text{iter,age})}$ represents the social movement vector of the n-th horse, while $s_n^{(\text{iter,age})}$ denotes the horse's tendency to move toward the herd during the i-th iteration. The parameter $s_n^{(\text{iter,age})}$ is reduced in each iteration by a sociability reduction coefficient, ω_s. Here, N represents the total number of horses.

2.6. Imitation

The social nature of horses within a herd allows them to share and learn both positive and negative behaviors by imitating one another [34], such as identifying suitable grazing locations. Juvenile horses have the propensity to mimic the more senior ones, and such an instinct pattern lasts until they reach adulthood. This behavior is commonly observed in horses up to 10 years old (i.e., γ horses). In the HOA algorithm, the parameter i represents the imitation behavior of horses, as described below [30]:

$$\mathbf{I}_n^{(\text{iter,age})} = i_n^{(\text{iter,age})} \left[\frac{1}{pN} \sum_{j=1}^{pN} \hat{\mathbf{X}}_j^{(\text{iter}-1)} - \mathbf{X}_n^{(\text{iter}-1)} \right], \quad \text{age} = \gamma , \tag{28}$$

$$i_n^{(\text{iter,age})} = i_n^{(\text{iter}-1,\text{age})} \times \omega_i . \tag{29}$$

where $\mathbf{I}_n^{(\text{iter, age})}$ denotes the imitation movement vector of the n-th horse toward the average position of the best horses with estimated locations $\hat{\mathbf{X}}$, and ω_i represents the imitation reduction coefficient. Moreover, pN denotes the number of horses with the best positions, where p is typically set to 10% of the herd size. It is also suggested that ω_i be set to 0.3 for age γ [30].

2.7. Defense Mechanism

Another behavioral aspect of horses considered in the HOA algorithm is *defense*. Horses' defensive reactions reflect their natural status as prey animals. Instinctively, horses compete for water and nourishment to eliminate rivals and avoid dangerous areas and predators. When confronted with a threatening situation, horses exhibit a *fight-or-flight* response to protect themselves, with fleeing being their primary reaction. They may also buck violently and neigh loudly when caught in a trap.

In the HOA optimizer, the horses' defense mechanism is modeled such that a horse moves away from other horses with inferior positions that are far from the optimal solution. This defensive behavior is maintained throughout the entire lifespan of both young and adult horses. Equations (30) and (31) describe the defense mechanism of horses using a negative parameter d, which forces the current horse to stay away from unsuitable locations [30].

$$\vec{D}_n^{(\text{iter,age})} = -d_n^{(\text{iter,age})} \left(\frac{1}{qN} \sum_{j=1}^{qN} \check{\mathbf{X}}_j^{(\text{iter}-1)} - \mathbf{X}_n^{(\text{iter}-1)} \right), \quad \text{age} \in \{\alpha, \beta, \gamma\}, \tag{30}$$

$$d_n^{(\text{iter,age})} = d_n^{(\text{iter}-1,\text{age})}\, \omega_d , \tag{31}$$

where $\vec{D}_n^{(\text{iter, age})}$ represents the escape vector of the n-th horse from those with the worst locations, which are indicated by the vector $\check{\mathbf{X}}$. The term qN denotes the number of horses occupying the worst positions in the current population, with a recommended value of $0.2N$ according to [30]. The parameter ω_d represents the defense reduction coefficient at each iteration. It is also suggested that the initial value of the parameter d for α, β, and γ be set to 0.5, 0.2, and 0.1, respectively, at the beginning of the search process [30].

2.8. Roam

The final behavioral pattern incorporated in the HOA algorithm is *roaming*. Horses naturally graze and move from one pasture to another in search of nourishment. Although the majority of horses are retained in stables, they preserve an innate tendency to roam. A horse may unexpectedly change its grazing location. Their curious nature drives them to explore new grasslands and learn their positions. To accommodate this curiosity, the side walls of stables are designed to allow horses to see one another [30]. A stochastic movement is more pronounced in young horses and gradually decreases as they reach full maturity. The HOA algorithm models this behavior as a random movement characterized by a parameter r. This process is expressed as follows [30]:

$$\vec{R}_n^{(\text{iter,age})} = r_n^{(\text{iter,age})}\, \mathbf{P}\mathbf{X}_n^{(\text{iter}-1,\text{age})}, \quad \text{age} = \gamma, \delta \tag{32}$$

$$r_n^{(\mathrm{iter,age})} = r_n^{(\mathrm{iter}-1,\mathrm{age})}\ \omega_r\ . \tag{33}$$

In the above equations, $\vec{R}_n^{(\mathrm{iter,age})}$ represents the random velocity vector of the n-th horse used for local search and to escape from local minima. The parameter ω_r denotes the imitation reduction coefficient for $r_n^{(\mathrm{iter,age})}$ at each iteration. For ages γ and δ, it is recommended to set the initial values of r to 0.05 and 0.1, respectively [30]. By substituting Eqs. (23) – (33) into Eqs. (19) – (22), the velocity vector for each age group can be obtained.

Velocity of δ horses at the age of 0 – 5 years:

$$\begin{aligned}\vec{V}_m^{(\mathrm{iter},\delta)} = &\left[g_m^{(\mathrm{iter}-1,\delta)}\,\omega_g\,(\hat{\mathbf{u}}+\hat{\mathbf{P}}_l)\,\mathbf{X}_m^{(\mathrm{iter}-1)}\right]\\ &+\left[i_m^{(\mathrm{iter}-1,\delta)}\,\omega_i\left(\frac{1}{pN}\sum_{j=1}^{pN}\hat{\mathbf{X}}_j^{(\mathrm{iter}-1)}-\mathbf{X}_m^{(\mathrm{iter}-1)}\right)\right]\\ &+\left[r_m^{(\mathrm{iter}-1,\delta)}\,\omega_r\,\mathbf{P}\mathbf{X}_m^{(\mathrm{iter}-1)}\right].\end{aligned} \tag{34}$$

Velocity of γ horses at the age of 5 – 10 years:

$$\begin{aligned}\vec{V}_m^{(\mathrm{iter},\gamma)} = &\left[g_m^{(\mathrm{iter}-1,\gamma)}\,\omega_g\,(\hat{\mathbf{u}}+\hat{\mathbf{P}}_l)\,\mathbf{X}_m^{(\mathrm{iter}-1)}\right]\\ &+\left[h_n^{(\mathrm{iter}-1,\gamma)}\,\omega_h\,(\mathbf{X}_*^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)})\right]\\ &+\left[s_n^{(\mathrm{iter}-1,\gamma)}\,\omega_s\left(\frac{1}{N}\sum_{j=1}^{N}\mathbf{X}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right)\right]\\ &+\left[i_n^{(\mathrm{iter}-1,\gamma)}\,\omega_i\left(\frac{1}{pN}\sum_{j=1}^{pN}\hat{\mathbf{X}}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right)\right]\\ &-\left[d_n^{(\mathrm{iter}-1,\gamma)}\,\omega_d\left(\frac{1}{qN}\sum_{j=1}^{qN}\check{\mathbf{X}}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right)\right]\\ &+\left[r_n^{(\mathrm{iter}-1,\gamma)}\,\omega_r\,\mathbf{P}\mathbf{X}_n^{(\mathrm{iter}-1)}\right].\end{aligned} \tag{35}$$

Velocity of β horses at the age of 10 – 15 years:

$$\begin{aligned}\vec{V}_n^{(\mathrm{iter},\beta)} = &\left[g_n^{(\mathrm{iter}-1,\beta)}\,\omega_g\,(\hat{\mathbf{u}}+\hat{\mathbf{P}}_l)\,\mathbf{X}_n^{(\mathrm{iter}-1)}\right]\\ &+\left[h_n^{(\mathrm{iter}-1,\beta)}\,\omega_h\,(\mathbf{X}_*^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)})\right]\\ &+\left[s_n^{(\mathrm{iter}-1,\beta)}\,\omega_s\left(\frac{1}{N}\sum_{j=1}^{N}\mathbf{X}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right)\right]\\ &-\left[d_n^{(\mathrm{iter}-1,\beta)}\,\omega_d\left(\frac{1}{qN}\sum_{j=1}^{qN}\check{\mathbf{X}}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right)\right].\end{aligned} \tag{36}$$

Velocity of α horses older than 15 years:

$$\begin{aligned}\vec{V}_n^{(\mathrm{iter},\alpha)} = &\,g_n^{(\mathrm{iter}-1,\alpha)}\,\omega_g\,(\hat{\mathbf{u}}+\hat{\mathbf{P}}_l)\,\mathbf{X}_n^{(\mathrm{iter}-1)}\\ &-d_n^{(\mathrm{iter}-1,\alpha)}\,\omega_d\left(\frac{1}{qN}\sum_{j=1}^{qN}\check{\mathbf{X}}_j^{(\mathrm{iter}-1)}-\mathbf{X}_n^{(\mathrm{iter}-1)}\right).\end{aligned} \tag{37}$$

The pseudo-code and the flowchart of the HOA algorithm are illustrated as Algorithm 1 and in Fig. 2.

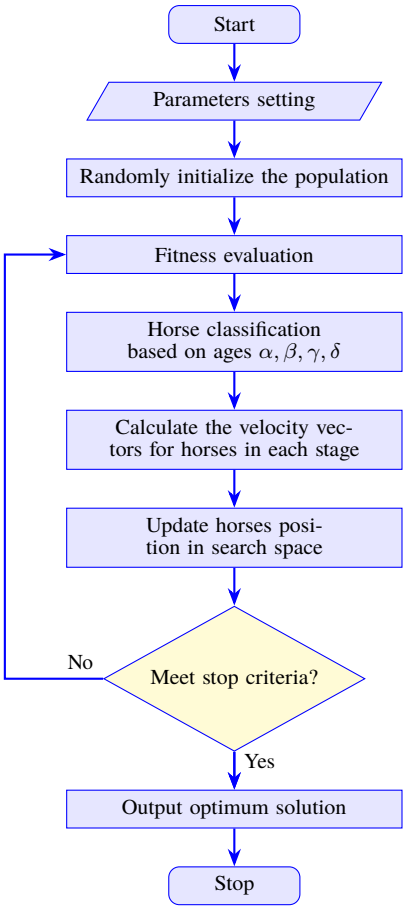

Fig. 2. Flowchart of the HOA algorithm.

3. Numerical Results and Discussions

A comprehensive comparative analysis is performed against root-MUSIC, ESPRIT, and PSO algorithms to validate the effectiveness of the proposed optimizer. This study includes four major evaluation criteria: estimation accuracy, convergence behavior, robustness against the number of elements, noise, and snapshots, and RMSE performance based on Monte Carlo simulations.

To perform these simulations, a uniform linear array (ULA) of isotropic point sources consisting of $M = 8$ half-wavelength dipoles is considered, with an inter-element spacing of half a wavelength. This spacing is selected to mitigate mutual coupling effects between adjacent dipoles. In all scenarios, the narrowband sources impinging on the ULA from different directions are assumed to be uncorrelated, and the added noise is Gaussian white noise. The simulations are carried out using 20 snapshots and 2400 Monte Carlo trials. For DOA estimation using the HOA algorithm, the fitness values are evaluated by searching for the global minimum of a non-linear cost function corresponding to the adopted ULA array, as defined in Eq. (11).

The HOA and PSO algorithms are implemented in Matlab with a maximum of 100 iterations. The search space for the uniform linear array is defined over the angular range of 0 to $180°$. For the HOA algorithm, the number of horses is set

to $N = 50$, and the control parameters h^{β}, h^{γ}, s^{β}, s^{γ}, i^{γ}, d^{α}, d^{β}, d^{γ}, r^{δ}, and r^{γ} are assigned values of 0.9, 0.5, 0.2, 0.1, 0.3, 0.5, 0.2, 0.1, 0.1, and 0.05, respectively. For the PSO algorithm, the swarm size is set to 200 particles, the cognitive and social acceleration coefficients are fixed at $C_1 = C_2 = 2$, and the inertia weight is linearly varied between $W_1 = 0.9$ and $W_2 = 0.4$.

In the first simulation, histograms of the DOA estimates obtained using the HOA optimizer are presented for $L = 3$ sources with DOAs at $[30°, 60°, 120°]$, and for $L = 4$ sources with DOAs at $[30°, 50°, 70°, 90°]$, as illustrated in illustrated in Fig. 3. These histograms represent the probability of a signal arriving from a given direction. Averaging the histograms over consecutive data blocks improves both the accuracy and reliability of the analysis. The peaks of the resulting angular probability distributions provide estimates of the DOAs.

Algorithm 1 Pseudo code of the horse herd optimization algorithm

Start

1: Define input parameters specific to the problem, including their corresponding constraints and algorithm parameters
2: Set $Itr = 1$
3: Initialization: Generate initial horse positions randomly and uniformly distributed throughout the search space between θ_{min} and θ_{max}
4: Fitness evaluation: Evaluate the fitness of each horse's position according to the objective function described in Eq. (11)
5: **while** $Itr \leqslant Itr_{max}$ **do**
6: **for** $i = 1$ to total number of horses **do**
7: Sort the locations of horses in ascending order based on their fitness values
8: Organize the horses into α, β, γ, and δ classes according to their age ranges
9: Compute the velocity vectors for horses in each class using Eqs. (22) to (25)
10: Update the positions of all horses by applying the corresponding movement vectors according to Eq. (21).
11: **end for**
12: Fitness evaluation: Evaluate the fitness of each horse's updated position using the objective function in Eq. (11).
13: Update best solution:
14: **if** new fitness value < the previous best fitness value **then**
15: • Set the new position as the best position
16: • Set the new fitness value as the best fitness value
17: **end if**
18: $Itr = Itr + 1$
19: **end while**
20: Return the best solution

End

The results indicate that the proposed algorithm yields highly accurate joint DOA estimates, with distinct peaks occurring precisely at the true angles. Moreover, estimation accuracy of the HOA algorithm is comparable to that reported in [12]. However, unlike subspace-based methods which require a large number of snapshots to construct the covariance matrix (CM), the proposed approach achieves accurate DOA estimation with a significantly smaller number of snapshots.

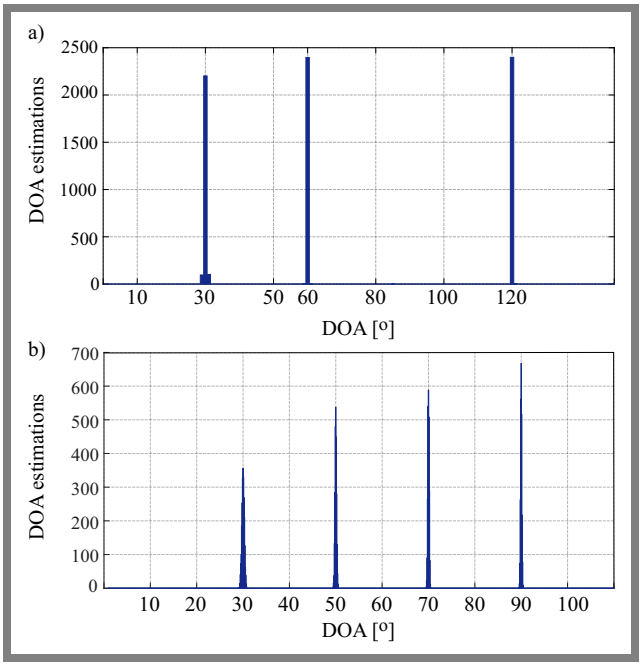

Fig. 3. Histogram of DOA estimations for HOA algorithm: a) $L = 3$ sources located at [30°, 60°, 120°] with $SNR = [5, 5, 5]$ dB, inter-element spacing $d = \lambda/2$, and an array of $M = 8$ elements, and b) $L = 4$ sources located at [30°, 50°, 70°, 90°] with $SNR = [5, 5, 5, 5]$ dB.

In the second simulation experiment, a statistical analysis is performed for the HOA, PSO, root-MUSIC, and ESPRIT algorithms to evaluate their ability to precisely and efficiently resolve incoming signals. The performance assessment is based on several statistical metrics, including the best, worst, mean, standard deviation (STD), and variance, which enhance the reliability of the experimental results.

This analysis examines the DOA estimation accuracy of the considered algorithms under three different scenarios: two sources with directions $L = [60°, 65°]$, three sources with directions $L = [60°, 70°, 80°]$, and four sources with directions $L = [60°, 70°, 80°, 90°]$, all at an SNR of -5 dB. The corresponding histograms are illustrated in Figs. 4, 5, while the statistical results obtained from 2400 Monte Carlo trials are summarized in Tabs. 1 – 3.

From Fig. 4a, it is evident that distinct peaks appear around 60°and 65°without any estimation failure when using the HOA algorithm. In contrast, the PSO algorithm successfully estimates the two DOAs, as indicated by the presence of two clear peaks around 60° and 65° in Fig. 4b. However, an undesired spurious peak appears at 115°, suggesting that PSO converges to a suboptimal solution during the DOA estimation process.

Furthermore, no dominant peaks corresponding to the true directions of arrival are observed in Figs. 4c-d. Instead, the estimated angles are widely dispersed over the range of 0 to 180°, indicating that the ESPRIT and root-MUSIC algorithms fail to provide reliable DOA estimates under this scenario.

The statistical results reported in Tab. 1 further confirm the superiority of the proposed approach. The HOA algorithm consistently yields more accurate DOA estimates across mul-

Tab. 1. Statistical results obtained from the comparative algorithms with two signal sources.

Angle	Algorithm	Best [°]	Average [°]	Worst [°]	STD [°]	Variance
60°	HOA	60.0003218	60.0084690	62.0928421	0.50195	0.251962
	PSO	60.0002743	59.9963548	63.5189249	0.87261	0.761442
	ESPRIT	60.0030129	55.1508742	1.1694029	11.8465	140.3399
	root-MUSIC	60.0228375	55.1782856	1.2533743	14.0110	196.3082
65°	HOA	64.9993400	64.9979923	63.4006353	0.46555	0.216737
	PSO	65.0006115	85.7792229	118.1794532	24.6773	608.9697
	ESPRIT	64.9922174	82.5666223	179.1887797	26.8622	721.5804
	root-MUSIC	65.0031988	100.3000588	179.5411878	31.1882	972.7084

Tab. 2. Statistical results obtained from the comparative algorithms with three signal sources.

Angle	Algorithm	Best [°]	Average [°]	Worst [°]	STD [°]	Variance
70°	HOA	70.0003199	70.0027179	71.9030490	0.483988	0.234245
	PSO	69.9997522	70.0841697	74.8310570	1.330736	1.770858
	ESPRIT	69.9993590	66.9298423	139.974860	11.81534	139.60224
	root-MUSIC	69.9980060	60.7203797	1.2675782	17.10490	292.57776
80°	HOA	80.0004090	79.9888340	81.8906260	0.472660	0.223407
	PSO	79.9985383	80.9648580	102.110770	4.156118	17.27331
	ESPRIT	80.0003768	80.7862460	157.564900	8.075658	65.21624
	root-MUSIC	80.0205634	82.0559765	131.482520	7.951755	63.23040
90°	HOA	90.0002380	90.0080280	92.1217950	0.439226	0.192919
	PSO	89.9999996	90.3872096	103.745127	4.528096	20.50365
	ESPRIT	90.0013500	95.1284890	176.143860	14.40636	207.54317
	root-MUSIC	89.9975326	110.933836	179.261060	24.60050	605.18449

Tab. 3. Statistical results obtained from the comparative algorithms with four signal sources.

Angle	Algorithm	Best [°]	Average [°]	Worst [°]	STD [°]	Variance
60°	HOA	59.999858	60.015577	58.064286	0.521405	0.271863
	PSO	60.000583	60.001719	57.074942	0.855451	0.731797
	ESPRIT	60.017033	55.614013	0.346814	11.88656	141.2903
	root MUSIC	60.000541	50.870238	0.665556	16.57344	274.6788
70°	HOA	69.999971	69.999257	68.186362	0.484881	0.235110
	PSO	69.999927	70.054767	75.944602	1.299854	1.689620
	ESPRIT	69.998388	70.449175	45.331282	6.557652	43.00281
	root MUSIC	70.011308	71.892475	95.026589	8.179528	66.90468
80°	HOA	79.999460	80.003784	81.610049	0.454403	0.206482
	PSO	79.999663	80.793881	101.15362	4.088790	16.71820
	ESPRIT	80.004837	82.124547	105.91682	6.549457	42.89539
	root MUSIC	79.987492	87.897832	132.00127	10.37003	107.5375
90°	HOA	90.000030	90.106479	154.89177	2.632250	6.928739
	PSO	90.000004	90.550814	103.67308	4.535886	20.57426
	ESPRIT	89.996200	101.91026	179.00447	20.29290	411.8016
	root MUSIC	90.000406	121.50774	178.77402	23.78936	565.9337

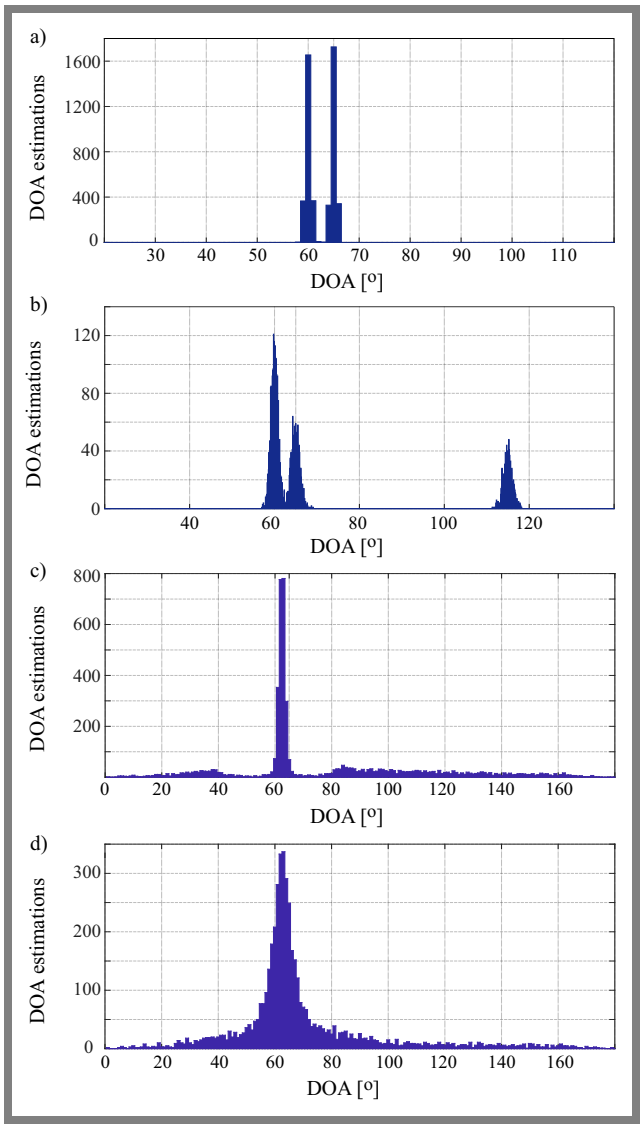

Fig. 4. Histogram of DOA estimations for $L = 2$ sources located at [60°, 65°], with SNR = [–5, –5] dB, inter-element spacing $d = \lambda/2$, and an array of $M = 8$ elements, using: a) HOA, b) PSO, c) root-MUSIC, and d) ESPRIT.

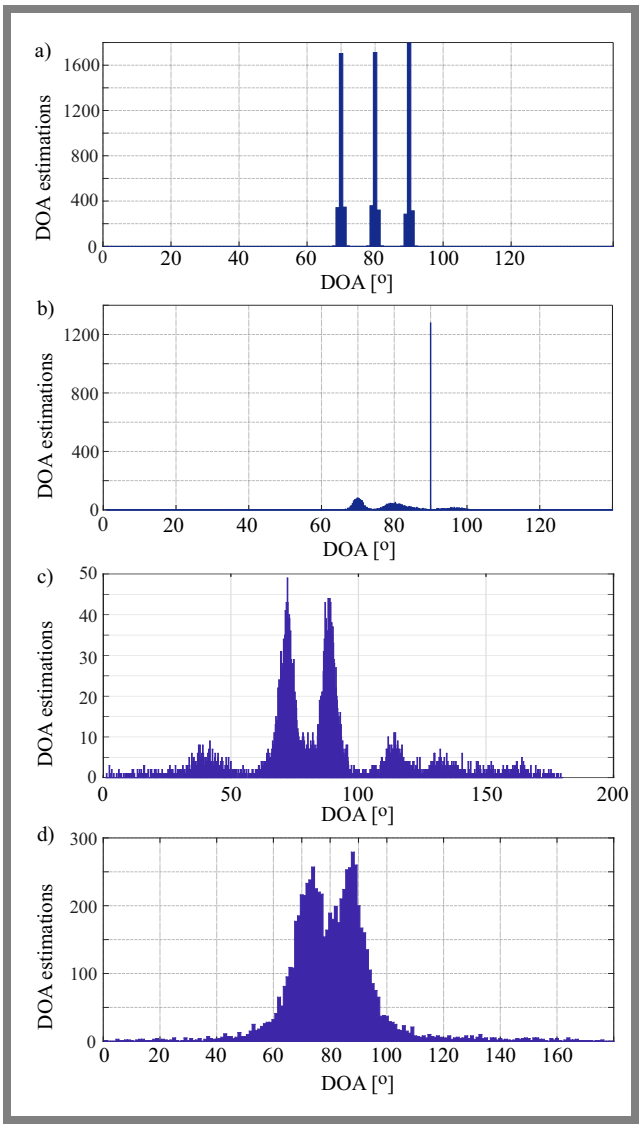

Fig. 5. Histogram of DOA estimations for $L = 3$ sources located at [70°, 80°, 90°], with SNR = [–5, –5, –5] dB, inter-element spacing $d = \lambda/2$, and an array of $M = 8$ elements, using: a) HOA, b) PSO, c) root-MUSIC, and d) ESPRIT algorithms.

tiple independent Monte Carlo simulations, achieving the best performance in terms of best, average, and worst estimated angles, while also exhibiting the smallest variance and standard deviation relative to PSO, ESPRIT and root-MUSIC.

Figure 5a illustrates that the proposed SI-based algorithm provides the most accurate DOA estimations with high resolution, where three distinct and well-defined peaks are clearly observed around 70°, 80°, and 90°. In comparison, the PSO algorithm is able to estimate the three DOAs, as shown in Fig. 5b. However, an undesired spurious peak appears at an unintended angle, indicating suboptimal convergence.

On the other hand, the root-MUSIC algorithm does not consistently yield accurate DOA estimates, for scattered estimates are observed around the true DOA peaks, as illustrated in Fig. 5c. Moreover, the ESPRIT algorithm completely fails to provide reliable DOA estimations in this scenario, since no dominant peaks are observed near the actual angles and the estimated directions are widely dispersed, as shown in Fig. 5d.

Table 3 summarizes the performance comparison among PSO, root-MUSIC, and ESPRIT for this scenario. Once again, the HOA algorithm demonstrates significantly better performance by achieving the most accurate DOA estimates, with best, average, and worst estimated angles closely matching the true directions. Furthermore, it exhibits the lowest variance and standard deviation.

Figure 6a shows that the proposed algorithm achieves highly accurate joint DOA estimation by producing sharp histograms for the corresponding DOA estimates. As can be clearly observed, the exact peaks occur at [60°, 70°, 80°, 90°] without any estimation failure. In contrast, the results obtained using the PSO algorithm, shown in Fig. 6b, are inferior to those of the HOA algorithm, as an additional spurious peak appears around 97°, indicating suboptimal estimation performance.

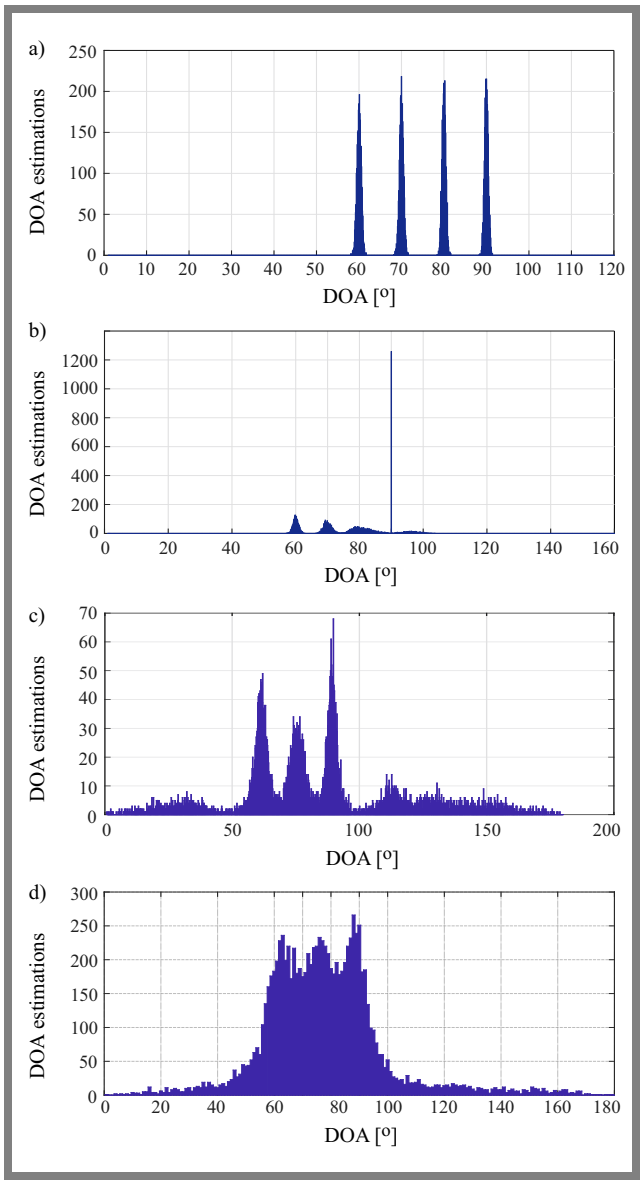

Fig. 6. Histogram of DOA estimates for $L = 4$ sources located at [60°, 70°, 80°, 90°], with SNRs of [–5,–5,–5, –5] dB, inter-element spacing $d = \lambda/2$, and $M = 8$ antenna elements, using: a) HOA, b) PSO, c) root-MUSIC, and d) ESPRIT algorithms.

Furthermore, as shown in Figs. 6c-d, no clear peaks are observed around the true angles of arrival, revealing that the DOA estimation performance of the ESPRIT and root-MUSIC algorithms is very poor and suffers from large estimation errors.

The statistical results presented in Tab. 4 further confirm the superiority of the HOA algorithm.

Tab. 4. Computational complexity analysis.

Algorithm	Computational complexity
HOA / PSO	$O(M \cdot K \cdot L \cdot \text{Iter} \cdot N)$
ESPRIT	$O(M^3 + M^2K)$
MUSIC	$O(M^3 + M^2K + M^2Q)$
root-MUSIC	$O(M^3 + M^2K)$

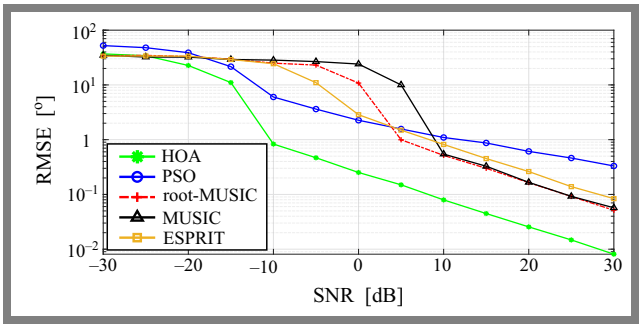

Fig. 7. Root mean square error versus SNR for DOA.

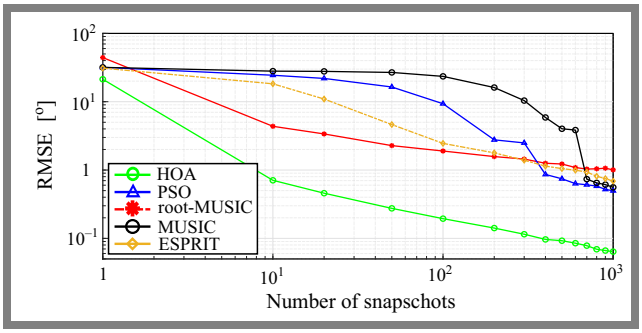

Fig. 8. Root mean square error versus snapshots for DOA.

In the third simulation, we evaluate the performance of the proposed algorithm using the root mean square error (RMSE) as the performance metric, which quantifies how close the estimated angles are to the true angles of the signal sources. The RMSE is calculated using the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{VL}\sum_{v=1}^{V}\sum_{l=1}^{L}(\hat{\theta}_{l,v} - \theta_l)^2}\,, \tag{38}$$

where $\theta_{l,v}$ denotes the estimated DOA of the l-th source, and V represents the number of Monte Carlo trials, set to 200, used to average the performance.

The evaluation is conducted by varying the signal-to-noise ratio (SNR), the number of snapshots, and the number of array elements.

In the first case, the effect of SNR on the performance of the algorithms is evaluated. Figure 7 illustrates the RMSE curves of the algorithms under different SNR conditions. We consider a scenario where the SNR varies in the range of [–30, 30] dB, with 20 snapshots and $M = 8$ antenna elements.

According to the RMSE values, the error rates of the different algorithms are very high when SNR is –30 dB. As the SNR increases, the RMSE of all algorithms gradually decreases, leading to more accurate DOA estimates. Moreover, the proposed algorithm consistently exhibits lower RMSE than the PSO, root-MUSIC, and ESPRIT algorithms under the same noise conditions. Specifically, under low SNR conditions (SNR < 0 dB), the proposed algorithm outperforms all other algorithms, demonstrating excellent robustness against noise.

In the second case, the impact of the number of snapshots on the performance of each algorithm is examined. Plots of the RMSE as a function of the number of snapshots for all algorithms are illustrated in Fig. 8, where the number of

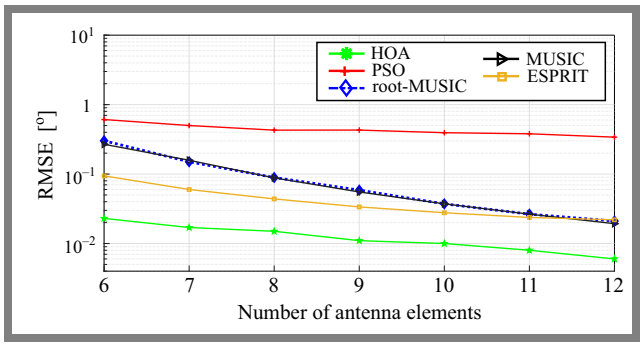

Fig. 9. Root mean square error versus number of antenna elements for DOA.

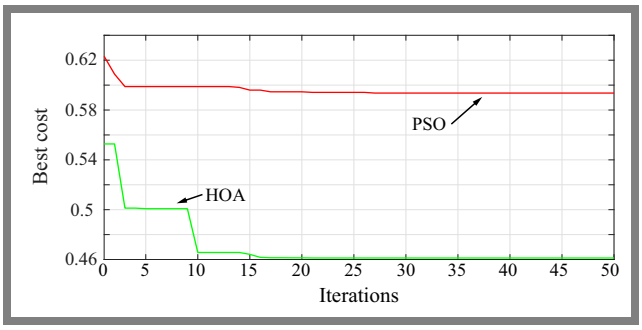

Fig. 10. Convergence characteristics of the fitness value versus the number of iterations.

snapshots varies in the range of [20, 1000], with $M = 8$ and SNR = –5 dB.

The results indicate that the accuracy of DOA estimation using the HOA, PSO, root-MUSIC, MUSIC, and ESPRIT algorithms is heavily influenced by the number of snapshots. Under conditions with few snapshots, the error rate of the subspace algorithms is very high, demonstrating a significant deterioration in estimation resolution. In contrast, the proposed HOA algorithm consistently demonstrates greater robustness when fewer snapshots are available for DOA estimation, with its error reduced to zero above just 20 snapshots. This makes the method particularly well-suited for situations with limited data. As the number of snapshots increases (beyond 10 snapshots), RMSE gradually decreases, indicating improved estimation accuracy. This improvement is expected because subspace methods require a large number of stationary time samples to capture the statistical properties necessary for accurately estimating the covariance matrix.

In this third scenario, we evaluate and compare the estimation accuracy of all previously mentioned methods by monitoring RMSE against the number of antenna elements. This scenario considers seven different values for antenna elements ($M = 6, 7, 8, 9, 10, 11, 12$) at SNR of 25 dB, using 20 snapshots. The relationship between RMSE and the number of antenna elements M is depicted in Fig. 9. It is evident that as the number of antenna elements M incresase, the error decreases, exhibiting substantial improvement in estimation accuracy for all techniques. Furthermore, the HOA and CSA algorithms provide better estimation resolution, outperforming the PSO, ESPRIT, and root-MUSIC methods across the simulated range of antenna elements.

Finally, the process of finding the optimal solution using the proposed algorithm is illustrated through convergence graphs which track the progression of the best solution at each iteration. Figure 10 presents the normalized average fitness values against the number of generations for the HOA and PSO algorithms. The figure shows that the proposed HOA algorithm converges faster than the PSO algorithm and can typically obtain near-optimal solutions more quickly when estimating the angles of incoming signals.

Table 4 provides a comparative summary of computational complexity, including subspace-based algorithms (ESPRIT, MUSIC, and root-MUSIC), PSO, and the proposed HOA approach. The main contributors to the computational complexity of the subspace-based algorithms are covariance matrix estimation, eigenvalue decomposition, and angular spectrum scanning.

However, ESPRIT and root-MUSIC avoid the spectral peak searching required by MUSIC. Therefore, the parameters associated with these algorithms are the number of array elements M, the number of signal snapshots K, the number of sources L, and the number of scanning angles Q. The complexity of ESPRIT primarily comes from the covariance estimation $O(M^2K)$ and the eigen-decomposition process $O(M^3)$, resulting in an overall complexity of $O(M^3 + M^2K)$. On the other hand, the computational complexity of the MUSIC algorithm mainly stems from covariance matrix estimation $O(M^2K)$, eigenvalue decomposition $O(M^3)$, and angular spectrum scanning $O(M^2Q)$, giving a total complexity of $O(M^3 + M^2K + M^2Q)$. Root-MUSIC avoids spectral peak searching by converting the MUSIC spectrum into a polynomial rooting problem. Its computational complexity is dominated by covariance estimation and polynomial rooting, resulting in $O(M^3 + M^2K)$. Consequently, MUSIC has higher computational complexity due to its peak-searching approach in the spatial pseudo-spectrum.

In contrast, swarm intelligence (SI) algorithms such as PSO and HOA avoid matrix eigen-decomposition and spectral peak searching. These approaches are primarily governed by the number of times the cost function is evaluated and the size of the operations in each evaluation.

In each iteration of the SI-based algorithm, a set of candidate solutions is generated. Evaluating each candidate requires constructing a steering matrix based on the estimated angles, estimating the source signals, and calculating the received signal reconstruction error. These operations involve matrix operations whose complexity is proportional to M and K. Consequently, the overall complexity increases linearly with the number of antenna elements M, snapshots K, and sources L, and also depends on the number of iterations and population size, resulting in $O(M \cdot K \cdot L \cdot \mathrm{Iter} \cdot N)$

Although the HOA-based approach incurs higher complexity than the subspace algorithms, the overall complexity remains polynomial and can be adjusted by selecting the population size and the number of iterations. This offers a favorable trade-off between estimation accuracy and computational cost, particularly in multi-source and low-SNR scenarios. According to the literature, a qualitative comparison among

the maximum likelihood (ML) algorithm [35], subspace-based methods, swarm-intelligence-based approaches, and the proposed HOA algorithm can be summarized as follows.

The ML algorithm achieves very high accuracy under ideal conditions but suffers from low robustness in noisy environments and has very high computational complexity. Subspace-based methods, such as MUSIC, root-MUSIC, and ESPRIT, provide high accuracy under favorable conditions. However, their performance degrades at low SNR and with a limited number of snapshots due to their reliance on covariance matrix estimation.

Swarm-intelligence-based approaches, such as PSO, offer moderate accuracy and robustness but remain computationally demanding because of their iterative optimization process. In contrast, the proposed HOA algorithm demonstrates high accuracy with strong robustness in low-SNR and limited-snapshot scenarios while maintaining moderate computational complexity, thus providing a favorable trade-off between performance and computational cost.

4. Conclusions

The performance of most DOA estimation algorithms, such as MUSIC and ESPRIT, deteriorates significantly in challenging scenarios, such as high noise levels or when sources are closely spaced. Moreover, these algorithms require a large number of snapshots to maintain the accuracy of the estimated DOAs of incident signals.

In this paper, we propose a DOA estimation method to address these challenges. Through extensive simulations, the proposed method is compared with several well-known algorithms, including ESPRIT, root-MUSIC, and PSO. Simulation results show that HOA achieves the highest resolution, particularly at low SNRs and for small angular separations between sources. Additionally, experimental results confirm the robustness and effectiveness of the HOA algorithm. They also prove that is consistently outperforms these methods in terms of RMSE across various scenarios, including different SNR levels, numbers of snapshots, and numbers of array elements.

Overall, the main advantage of the proposed HOA-based DOA estimation lies in its ability to accurately detect signal angles with a limited number of snapshots, making it particularly suitable for real-time implementation when data collection is constrained.

References

[1] N. Ruan, H. Wang, F. Wen, and J. Shi, "DOA Estimation in B5G/6G: Trends and Challenges", *Sensors*, vol. 22, art. no. 5125, 2022 (https://doi.org/10.3390/s22145125).

[2] A.E. Zorkun, M.A. Salas-Natera, R.M. Rodrıguez-Osorio, and S. Chatzinotas, "Energy Efficient Low-complexity RIS-aided 3-D DoA Estimation and Target Tracking Algorithm via Matrix Completion", *IEEE Access*, vol. 12, pp. 197929–197941, 2024 (https://doi.org/10.1109/ACCESS.2024.3511717).

[3] V. Dakulagi and M. Bakhar, "Smart Antenna System for DOA Estimation Using Single Snapshot", *Wireless Personal Communications*, vol. 107, pp. 81–93, 2019 (https://doi.org/10.1007/s11277-019-06241-0).

[4] O. Barkat, "Capon DOD/DOA Estimation Algorithm for Bistatic MIMO Radar Using Dipole Antenna Arrays with Known Mutual Coupling", *Journal of Telecommunications and Information Technology*, vol. 101, pp. 1–7, 2025 (https://doi.org/10.26636/jtit.2025.3.2121).

[5] S.K. Imtiaj, I.S. Misra, and S. Bhattacharya, "Revisiting Smart Antenna Array Design with Multiple Interferers Using Basic Adaptive Beamforming Algorithms: Comparative Performance Study with Testbed Results", *Engineering Reports*, vol. 3, pp. 1–32, 2020 (https://doi.org/10.1002/eng2.12295).

[6] F. Gross, *Smart Antenna for Wireless Communication*, New York: McGraw-Hill, 288 p., 2005 (ISBN: 9780071447898).

[7] N. Boughaba and O. Barkat, "LMS and RLS Beamforming Algorithms Based Linear Antenna Array with Known Mutual Coupling", *Journal of Electromagnetic Waves and Applications*, vol. 37, pp. 1449–1462, 2024 (https://doi.org/10.1080/09205071.2023.2251979).

[8] A. Kuchar *et al.*, "A Robust DOA-Based Smart Antenna Processor for GSM Base Stations", *IEEE International Conference on Communications*, Vancouver, Canada, 1999 (https://doi.org/10.1109/ICC.1999.767867).

[9] M. Bensalem and O. Barkat, "DOA Estimation of Linear Dipole Array with Known Mutual Coupling Based on ESPRIT and MUSIC", *Radio Science*, vol. 57, pp. 1–15, 2022 (https://doi.org/10.1029/2021RS007294).

[10] K. Hameed *et al.*, "DOA Estimation in Low SNR Environment through Coprime Antenna Arrays: An Innovative Approach by Applying Flower Pollination Algorithm", *Applied Sciences*, vol. 11, art. no. 7985, 2021 (https://doi.org/10.3390/app11177985).

[11] T.S. Bird, *Mutual Coupling Between Antennas*, UK: Wiley, 480 p., 2021 (ISBN: 9781119564980).

[12] K. Gowri and P. Palanisamy, "Two Dimensional Direction of Arrival Estimation Algorithm for Coherent Signals Using Three Parallel Uniform Linear Arrays", *Journal of Communications Technology and Electronics*, vol. 64, pp. 1383–1390, 2019 (https://doi.org/10.1134/S106422691912009X).

[13] R. Roy and T. Kailath, "ESPRIT–Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 37, pp. 984–995, 1989 (https://doi.org/10.1109/29.32276).

[14] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, 1986 (https://doi.org/10.1109/TAP.1986.1143830).

[15] J. Yao *et al.*, "Satellite Interference Source Direction of Arrival (DOA) Estimation Based on Frequency Domain Covariance Matrix Reconstruction", *Sensors*, vol. 23, art. no. 7575, 2023 (https://doi.org/10.3390/s23177575).

[16] E. Talbi, *Metaheuristics from Design to Implementation*, John Wiley & Sons, 624 p., 2009 (ISBN: 978-0470278581).

[17] B. Morales-Castaneda *et al.*, "A Better Balance in Metaheuristic Algorithms: Does It Exist?", *Swarm and Evolutionary Computation*, vol. 54, art. no. 100671, 2020 (https://doi.org/10.1016/j.swevo.2020.100671).

[18] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm", *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016 (https://doi.org/10.1016/j.advengsoft.2016.01.008).

[19] F. Fausto *et al.*, "From Ants to Whales: Metaheuristics for All Tastes", *Artificial Intelligence Review*, vol. 53, pp. 753–810, 2020 (https://doi.org/10.1007/s10462-018-09676-2).

[20] R. Rajakumar, P. Dhavachelvan, and T. Vengattaraman, "A Survey on Nature Inspired Meta-heuristic Algorithms with its Domain Specifications", *International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2016 (https://doi.org/10.1109/CESYS.2016.7889811).

[21] J.C. Chang, "DOA Estimation for Local Scattered CDMA Signals by Particle Swarm Optimization", *Sensors*, vol. 12, pp. 3228–3242, 2012 (https://doi.org/10.3390/s120303228).

[22] N. Ahmed *et al.*, "Performance Analysis of Efficient Computing Techniques for Direction of Arrival Estimation of Underwater Multi Targets", *IEEE Access*, vol. 9, pp. 33284–33298, 2021 (https://doi.org/10.1109/ACCESS.2021.3060819).

[23] W. Shi, J. Huang, and Y. Hou, "Fast DOA Estimation Algorithm for MIMO Sonar Based on Ant Colony Optimization", *Journal of Systems Engineering and Electronics*, vol. 23, pp. 173–178, 2012 (https://doi.org/10.1109/JSEE.2012.00022).

[24] S.A.H. Parsa, A.E. Zadeh, and S.J. Kazemitabar, "A Novel Modified Artificial Bee Colony for DOA Estimation", *International Journal of Sensors Wireless Communications and Control*, vol. 11, pp. 96–106, 2021 (https://doi.org/10.2174/2210327909666191209154508).

[25] S. Akbar *et al.*, "Flower Pollination Heuristics for Parameter Estimation of Electromagnetic Plane Waves", *Computer Modeling in Engineering & Sciences*, vol. 68, pp. 2529–2543, 2021 (https://doi.org/10.32604/cmc.2021.016097).

[26] M. Jain, V. Singh, and A. Rani, "A Novel Nature-inspired Algorithm for Optimization: Squirrel Search Algorithm", *Swarm and Evolutionary Computation*, vol. 44, pp. 148–175, 2019 (https://doi.org/10.1016/j.swevo.2018.02.013).

[27] N. Ahmed, H. Wang, M.A.Z. Raja, and W. Ali, "Novel Design of Grey Wolf Optimization Heuristics for High Resolution Direction of Arrival Estimation in Acoustic Plane Waves", *Wireless Personal Communications*, vol. 128, pp. 2507–2529, 2022 (https://doi.org/10.1007/s11277-022-10057-w).

[28] A. Sharma *et al.*, "Maximum Likelihood Direction of Arrival Estimation Using Chicken Swarm Optimization Algorithm", *International Journal of Mathematical Engineering and Management Sciences*, vol. 6, pp. 621–635, 2021 (https://doi.org/10.33889/IJMEMS.2021.6.2.038).

[29] D.H. Wolpert and W.G. Macready, "No Free Lunch Theorems for Optimization", *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 67–82, 1997 (https://doi.org/10.1109/4235.585893).

[30] F. MiarNaeimi, G. Azizyan, and M. Rashki, "Horse Herd Optimization Algorithm: A Nature-inspired Algorithm for High-dimensional Optimization Problems", *Knowledge-Based Systems*, vol. 213, art. no. 106711, 2021 (https://doi.org/10.1016/j.knosys.2020.106711).

[31] D.A. Elmanakhly, M. Saleh, E.A. Rashed, and M. Abdel-Basset, "BinHOA: Efficient Binary Horse Herd Optimization Method for Feature Selection: Analysis and Validations", *IEEE Access*, vol. 10, pp. 26795–26816, 2022 (https://doi.org/10.1109/ACCESS.2022.3156593).

[32] A. Refaat *et al.*, "Extraction of Maximum Power from PV System Based on Horse Herd Optimization MPPT Technique under Various Weather Conditions", *Renewable Energy*, vol. 220, art. no. 119718, 2024 (https://doi.org/10.1016/j.renene.2023.119718).

[33] K. Krueger and J. Heinze, "Horse Sense: Social Status of Horses (Equus Caballus) Affects Their Likelihood of Copying Other Horses' Behavior", *Animal Cognition*, vol. 11, pp. 431–439, 2008 (https://doi.org/10.1007/s10071-007-0133-0).

[34] F. Bogner, "A Comprehensive Summary of the Scientific Literature on Horse Assisted Education in Germany", Bachelor Thesis, Van Hall Larenstein University, 2011.

[35] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Rao Bound", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 720–741, 1989 (https://doi.org/10.1109/29.17564).

Mohamed Bensalem, M.Sc.
Electronic Department
https://orcid.org/
E-mail: mbensalem481@yahoo.fr
University of Freres Mentouri Constantine 1, Constantine, Algeria
http://www.umc.edu.dz

Ouarda Barkat, Professor
Electronic Department
https://orcid.org/0000-0001-6784-8338
E-mail: barkat.ouarda@umc.edu.dz
University of Freres Mentouri Constantine 1, Constantine, Algeria
http://www.umc.edu.dz

ANN-enabled Gain Prediction and Optimization in Dual-band SIW Antenna Design for 5G Networks

Md Mahabub Alam, Md Raihanul Islam Tomal, Ahmad Afif Mohd Faudzi, and Nurhafizah Abu Talip Yusof

Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia

https://doi.org/10.26636/jtit.2026.1.2424

Abstract — Artificial neural networks (ANNs) help improve antenna design process by enabling adaptive optimization strategies that address important challenges in 5G wireless systems, including signal interference, limited coverage, and high user density. This study presents an AI-assisted design methodology for a compact dual-band substrate integrated waveguide (SIW) antenna optimized for 5G operation at 28 and 38 GHz. The antenna is implemented on a Rogers RT/Duroid 5880 substrate using a novel slot configuration with strategically positioned vias to enhance radiation characteristics. The fabricated prototype achieves gains of 8.05 dBi at 28 GHz and 7.89 dBi at 38 GHz, with fractional bandwidths of 6.41% (27.491 – 29.277 GHz) and 1.81% (37.496 – 38.179 GHz), while maintaining a return loss below –10 dB across both operating bands. The pivotal contribution of this work is the development of an ANN-based predictive model capable of accurately estimating antenna gain and radiation efficiency from main parameters such as slot dimensions, via size and feedline width. The proposed model demonstrates excellent predictive accuracy, achieving mean squared error values in the range of 0.00 to 0.001 for gain prediction and 0.018 to 0.066 for radiation efficiency estimation. This AI-driven framework significantly reduces design iterations, computational overhead, and prototyping requirements, offering an automated framework for efficient antenna development in next-generation 5G communication networks.

Keywords — 5G, ANN, gain prediction, machine learning, mmWave, SIW antenna

1. Introduction

To cope with limitations, including restricted bandwidth, high latency, substantial power consumption, limited scalability for the Internet of Things (IoT), and security vulnerabilities, 5G networks utilize three primary frequency bands including sub-6 GHz, millimeter wave (mmWave), and unlicensed bands [1], [2]. The mmWave spectrum, particularly the 28 and 38 GHz bands, is most important, as it offers unprecedented data rates and ultra-low latency required for next-generation wireless applications [3]. However, realizing this potential remains challenging due to severe propagation impairments at high frequencies, including increased loss of free space path and atmospheric attenuation, especially in dense urban environments [4].

To address these challenges, researchers have increasingly focused on developing compact, low-cost, and high-performance antennas for 5G networks, particularly of the microstrip patch type [5]. While simple and easily integrated, their performance at mmWave frequencies is limited by narrow bandwidth, low efficiency, and insufficient gain [6]. The antenna arrays offer a potential solution to these challenges, but unfortunately they come with considerable drawbacks, such as greater design complexity, increased size, and higher cost [7], [8].

In such a context, the substrate integrated waveguide (SIW) technology is a promising alternative, as it combines the high performance of conventional waveguides with the compactness and manufacturability of microstrip structures [9], [10]. SIW-based antennas support low-loss, wideband operation, and planar integration, making them particularly suitable for dual-band 5G applications in the 28 and 38 GHz spectrums [11], [12].

SIW slot antennas further enhance design flexibility through strategically positioned slots, though challenges remain in achieving wide bandwidth, efficient dual-band performance, and adaptive beam control at mmWave frequencies [13], [14]. Although full-wave electromagnetic simulators such as CST Microwave Studio and HFSS provide high accuracy for SIW slot antenna analysis, their reliance on dense meshing and iterative computations results in high computational cost and slow convergence [15], [16].

On the contrary, artificial neural networks (ANNs) enable fast antenna design by automating parameter optimization and accurately predicting pivotal performance metrics, such as gain, bandwidth, and radiation efficiency, without time-consuming prototyping [17].

This study bridges that gap by introducing a novel dual-band SIW antenna optimized for 28 and 38 GHz 5G operation [18].

The contributions of the proposed method are as follows:

- novel dual-band SIW antenna design with optimized slotted structures that achieve dual-band operation at 27.83 GHz and 37.81 GHz,
- integration of ANN predictive modeling for antenna performance estimation (gain, bandwidth, radiation efficiency),

which reduces the dependency on time-consuming CST simulations,

- comprehensive data set generation and parametric optimization which supports accurate ANN training and predictive capabilities.

2. Slotted SIW Microstrip Patch Antenna

The design process involves a series of stages, from the definition of material specifications to the refinement of antenna parameters, with the goal of achieving optimal performance in terms of gain, bandwidth, and efficiency. Figure 1 presents the workflow used to design and optimize the proposed slotted SIW antenna. The process begins by selecting the target operating bands and defining the substrate properties and design variables.

A parametric analysis is then performed by varying the slot dimensions (L_1, L_2, L_3, and W_1, W_2, W_3). Then, the antenna's performance is evaluated in terms of return loss, VSWR, gain, directivity, bandwidth, efficiency, and surface current distribution. If the performance satisfies the design targets, the results are finalized and compared with previously reported work. Otherwise, optimization is carried out to refine the structure.

The resulting data are statistically examined using ANOVA, and when a significant p-value ($\leqslant 0.05$) is obtained, a regression model is developed to further analyze performance trends. The adequacy is verified and, if necessary, additional optimization is performed. Once validated, the final design is confirmed using CST simulations and benchmarked against the published literature.

This methodology is scalable across multiple frequency bands (S, X, Ku, K, and Ka) by redefining the operating frequency and adjusting design parameters such as slot length, via diameter, and waveguide width relative to wavelength. Parametric modeling, supported by ANOVA and regression analysis, ensures accurate performance prediction under different geometric and material configurations, enabling robust and adaptable antenna optimization.

2.1. Evolution of the Dual-band SIW Antenna

This section presents the development of a dual-band SIW antenna through three design iterations, all modeled and optimized in CST. As shown in Fig. 2, each design is based on a compact 27 × 7.5 mm structure using a Rogers RT/Duroid 5880 substrate with relative permittivity of $\varepsilon_r = 2.2$ and thickness $h = 0.254$ mm. Antenna 1 introduces a 9 × 0.7 mm longitudinal slot and metallic vias forming the SIW sidewalls, together with a 3.50 × 0.37 mm feedline for impedance matching. Antenna 2 adds a 3.1 × 0.7 mm transverse slot to achieve dual-band operation. The final design, antenna 3, incorporates two longitudinal slots (9 × 0.7 mm and 7.7 × 0.7 mm), along with the transverse slot, to further enhance performance.

The S_{11} analysis indicates that antenna 1 provides a single resonance at 27.82 GHz with a return loss of –21.71 dB,

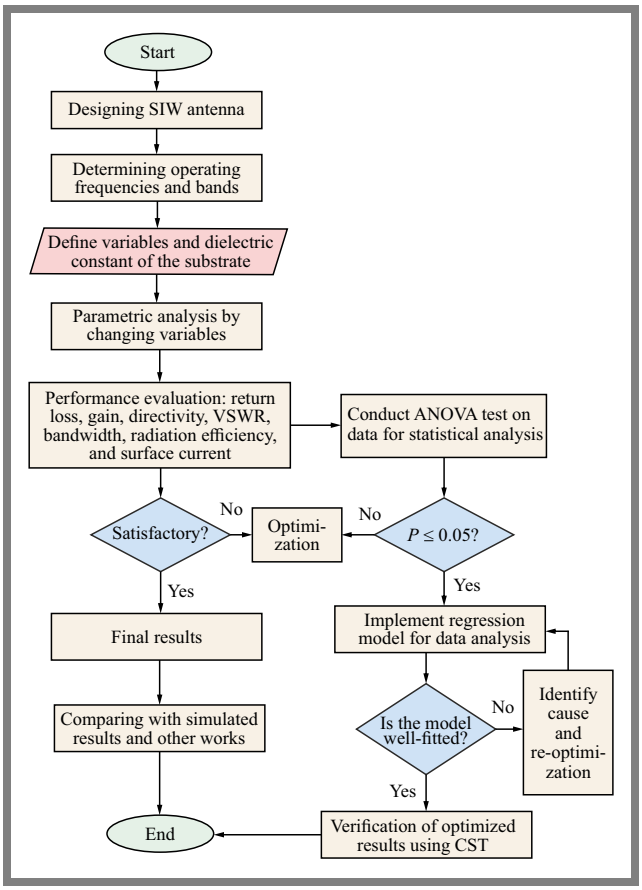

Fig. 1. Flowchart of slotted SIW antenna design and developed ANN model details.

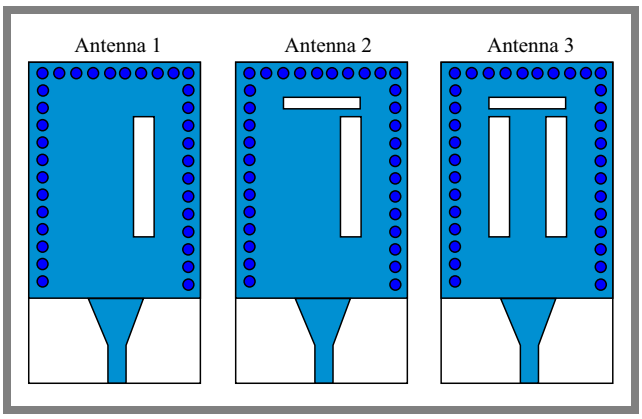

Fig. 2. Evolution of the design from single to triple-slot SIW antenna (no. 3).

as shown in Fig. 3. Antenna 2 introduces dual resonances at 28.29 GHz (–21.19 dB) and 37.87 GHz (–33.7 dB), confirming effective dual-band operation. The optimized design shown as antenna 3 exhibits deeper and more stable resonances at 27.83 and 37.81 GHz, with an effective operating range extending from 26 to 40 GHz.

This demonstrates a substantial improvement in bandwidth and impedance matching compared to the previous designs. These results clearly show that incorporating the second longitudinal slot enhances coupling and current distribution,

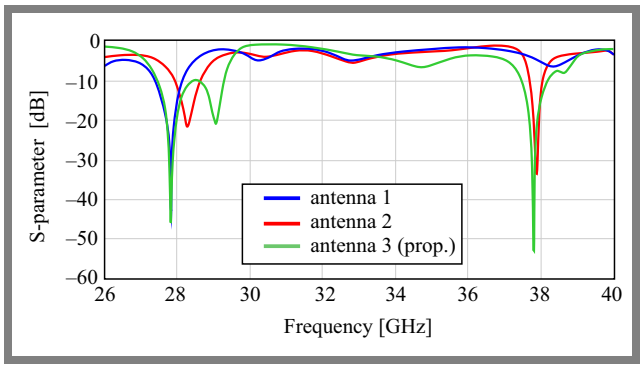

Fig. 3. S_{11} analysis and optimization for antennas 1 – 3.

leading to stronger resonances and superior dual-band performance.

2.2. Proposed Antenna Configuration

The proposed antenna incorporates a substrate-integrated waveguide (SIW) cavity with a rectangular microstrip patch containing strategically placed slots to enhance bandwidth and radiation performance. As shown in Figs. 4a–c, the design employs a multilayer structure in which the SIW cavity improves isolation and suppresses surface wave losses, while the slotted patch enables multiresonant operation. Figure 4a presents the 3D SIW structure, where two periodic rows of vias form the waveguide sidewalls, confining the electromagnetic fields and guiding energy toward the radiating slots.

The top view of the antenna, consisting of a rectangular patch with three slots and a tapered microstrip feed (dimensions W_f, L_f, W_{f1}, L_{f1}), is illustrated in Fig. 4b. The tapered feed provides improved impedance matching, allowing efficient coupling into the patch, which radiates through two longitudinal and one transverse slots designed to support dual-band operation.

The SIW structure, shown in Fig. 4c, consists of two rows of plated through-hole vias that form effective electrical sidewalls within the substrate, enabling stable antenna performance. The SIW is implemented using 0.5-mm diameter vias as detailed in Tab. 1. A 3.5 × 0.37 mm feedline and a 9.0 × 0.7 mm tapered section excite the 18.4 × 7.5 mm radiating patch.

To enhance gain and impedance matching, three slots are etched onto the patch, including two of the longitudinal (9 × 0.70 mm and 7.72 × 0.70 mm) and one of the transverse (3.10 × 0.70 mm) variety, modifying the current distribution and improving return loss.

For a rectangular dielectric-filled waveguide supporting the dominant TE_{10} mode, the dielectric-filled waveguide width a_d is [19]:

$$a_d = \frac{a}{\sqrt{\varepsilon_r}}\,, \tag{1}$$

where a is the effective width of the waveguide and ε_r is the relative dielectric constant.

The effective width of the SIW a_s is determined by [20]:

$$a_s = a_d + \frac{d^2}{0.98\,p}\,, \tag{2}$$

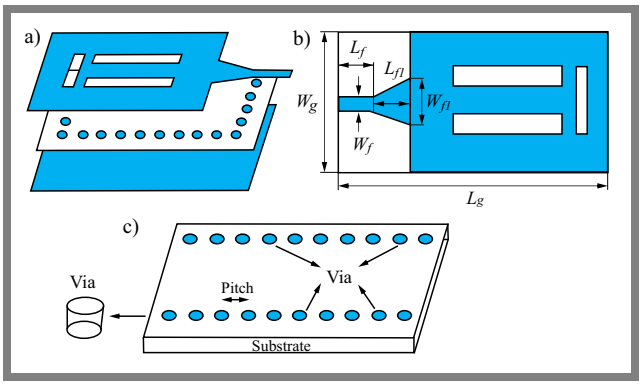

Fig. 4. Layout of the proposed slotted SIW antenna: a) 3D perspective, b) planar surface configuration, and c) internal SIW implementation.

where d is the diameter of the vias in the SIW and p is the pitch or center-to-center distance between the adjacent vias in the SIW structure.

2.3. Parametric Analysis

The slot dimensions affect the impedance matching and resonance frequency of the proposed SIW microstrip antenna, as shown in Figs. 5a–c. For the first longitudinal slot, S-parameter analysis with lengths $L_1 = 7.82 - 8.84$ mm and widths $W_1 = 0.66 - 0.72$ mm found an optimal configuration at $L_1 = 7.9$ mm and $W_1 = 0.7$ mm, achieving reflection coefficients of –46.294 and –53.813 dB at 28 and 38 GHz, respectively.

The transversal slot exhibited stable resonance at 28 GHz across dimension variations, whereas the 38 GHz band experienced minor shifts. The best performance was obtained at $L_2 \times W_2 = 3.1 \times 0.7$ mm. Similarly, the second longitudinal slot provided optimal reflection with $L_3 = 7.72$ mm and $W_3 = 0.7$ mm, highlighting strong sensitivity to slot geometry.

The diameter of the vias plays an important role in determining the resonance frequencies, reflection coefficients, and bandwidth performance of the antenna. Extensive testing of various via diameters revealed three significant configurations relevant to target frequencies. As shown in Fig. 5d, the optimal performance was achieved with a via diameter of 0.5 mm, providing the best resonance characteristics.

On the contrary, a diameter of 0.4 mm through the tube produced resonances at 27.23 and 36.98 GHz, with reflection coefficients of 25.88 and –22.36 dB, respectively. The largest

Tab. 1. Optimized dimension of the slotted SIW antenna.

Parameter	Optimized value [mm]	Parameter	Optimized value [mm]
W_g	7.50	W_1	0.70
L_g	27.00	L_2	3.10
d	0.50	W_2	0.70
W_f	0.37	L_3	7.72
L_f	3.50	W_3	0.70
L_1	9.00	p	1.00

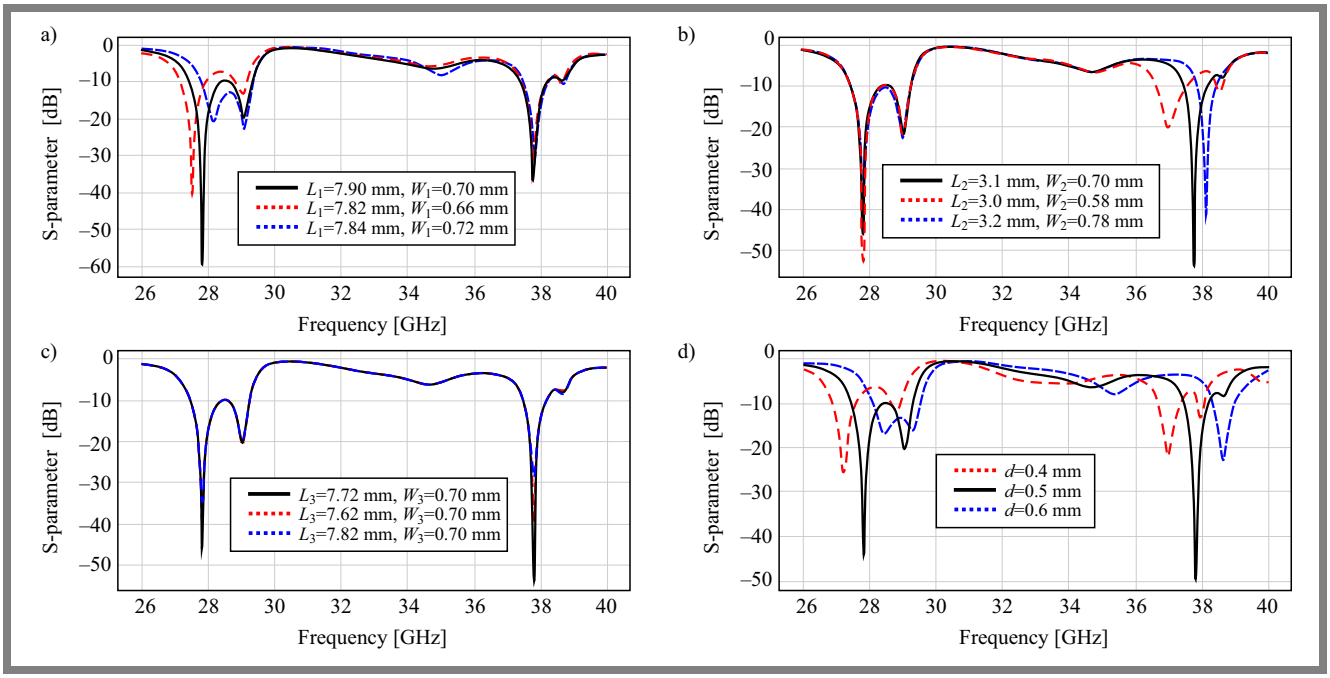

Fig. 5. Results of the parametric analysis of varying lengths and widths of a) the first longitudinal slot (L_1, W_1), b) the transverse slot (L_2, W_2), and c) the second longitudinal slot (L_3, W_3).

diameter tested equaled 0.6 mm and resulted in resonance frequencies deviating from the intended targets, highlighting sensitivity of the antenna's performance to via dimensions.

These CST simulation results serve as a basis for training an ANN model, enabling efficient prediction and optimization of antennas.

3. ANN-based Analysis Method

Conventional full-wave simulations using CST or HFSS software are accurate but computationally expensive, particularly for iterative design. To address this, here, a data-driven ANN model is trained on CST-generated datasets [21], allowing rapid prediction of metrics such as gain and efficiency at 28 and 38 GHz.

ANN effectively captures non-linear dependencies between design parameters and performance outcomes, offering high precision in complex dual-band scenarios [22]. This unified prediction framework streamlines analysis, supports inverse design, and significantly accelerates antenna development [23].

3.1. Data Analysis

As shown in Tab. 2, the data set comprises 158 antenna configurations characterized by 16 parameters, including eight input features such as slot lengths and widths, through diameter, and feedline width, along with eight output metrics (gain and efficiency) for dual-band operation at 28 and 38 GHz. This data set forms the foundation for training the proposed ANN model, enabling it to learn the complex non-linear relationships between antenna geometry and performance.

By varying geometrical and resonance-related parameters during CST-based simulations, a diverse and comprehensive dataset is obtained, capturing realistic design scenarios for dual-band operation. Consequently, the ANN is exposed to underlying physical behaviors and interdependencies that govern antenna gain and efficiency.

3.2. ANN Model Architecture and Training Process

For development, training, and evaluation of an ANN model, eleven pivotal geometrical and resonance-related features, including slot dimensions, through diameter, feeding width, and frequency-dependent parameters, are used as inputs to ensure accurate performance prediction across both bands. The ANN architecture comprises an input layer with 64 neurons processing normalized design features, two hidden layers with 32 and 16 neurons using ReLU activation to capture non-linear relationships, and an output layer that simulta-

Tab. 2. Characteristics of the dataset and model parameters.

Parameter	Description	Description
Dataset size	158 samples	Generated using parametric sweep
Training/test data	126 training (80%), 32 validation/testing (20%)	Ensures proper model generalization
Inputs (dimensions)	8 parameters: $L_1, W_1, L_2, W_2, L_3, W_3, d, W_f$	Used as regression model and ANN inputs
Performance metrics (outputs)	8 metrics: $F_1, F_2, RL_1, RL_2, G_1, G_2, E_{f1}, E_{f2}$	Model prediction targets: frequency, return loss, gain, radiation efficiency

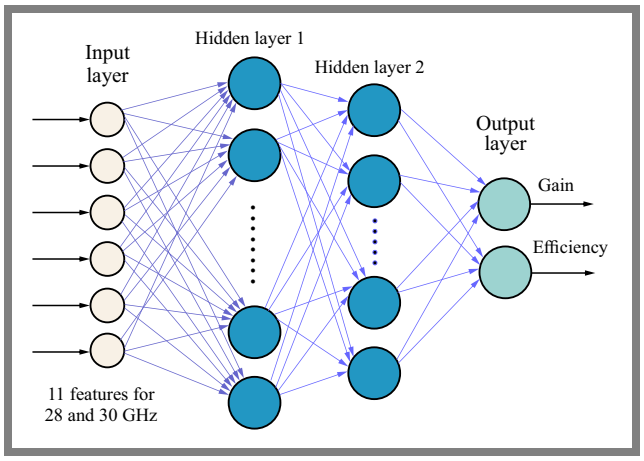

Fig. 6. ANN architecture for the proposed model.

neously predicts gain and efficiency at both frequencies, as shown in Fig. 6. This approach provides a fast and reliable alternative to full-wave electromagnetic simulations for antenna performance evaluation.

The ANN was trained using the Adam optimizer due to its adaptive learning rate and robustness in handling sparse gradients. Hyperparameter tuning identified an optimal learning rate of 0.001, a batch size of 128, and 100 training epochs, which together ensured stable convergence and good generalization performance. The mean squared error (MSE) was used as the loss function, as it effectively penalizes large prediction deviations in this regression task.

As shown in Tab. 2, the generated data set is relatively small, which increases the risk of overfitting. To address potential overfitting and ensure that the ANN model generalizes well, the following strategies were applied:

- The data set was divided into training, validation, and test sets. Model performance was monitored in the validation set to prevent fitting too closely to the training data.
- Techniques such as L2 weight regularization (and/or dropout, if applicable) were used to penalize overly complex models and reduce reliance on any single neuron.
- Training was halted when the validation error stopped improving, preventing unnecessary training epochs that could lead to overfitting.
- The network architecture (number of hidden layers and neurons) was kept as simple as possible while maintaining high predictive performance.
- K-fold cross-validation was employed to evaluate the generalization capability of the ANN by training and validating the model across multiple data partitions, ensuring that all samples contributed to both learning and testing.

This approach provides a more reliable performance estimate and reduces the risk of overfitting, particularly when working with a limited dataset size.

3.3. Model Performance Analysis Method for ANN

To evaluate the performance of the proposed ANN model, standard regression metrics are used, including mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). These metrics assess predictive accuracy for antenna gain and efficiency at 28 and 38 GHz. MSE, used as the loss function during model training, measures the average squared difference between predicted and actual values [24]:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \,. \tag{3}$$

MAE provides the mean absolute prediction error:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \,. \tag{4}$$

RMSE, the square root of MSE, expresses the error in the same units as the target variable [25]:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \,, \tag{5}$$

where $\hat{y}_i$ and y_i denote the predicted and actual values, respectively, and n is the number of samples.

These metrics are computed separately for gain and efficiency in both frequency bands, enabling the evaluation of the model's performance.

4. Results and Discussion

4.1. CST Simulation Results

The proposed antenna S-parameter response presented in Fig. 7a shows dual-band operation between 26 and 40 GHz with resonances at 27.83 and 37.81 GHz. At these frequencies, the S_{11} values fall below –10 dB, indicating good impedance matching, with minimum reflection coefficients of –46.29 and –53.82 dB, respectively. Using the –10 dB criterion, the antenna achieves a bandwidth of 1.786 GHz (FBW = 6.41%) at 28 GHz and 0.692 GHz at 38 GHz, confirming its suitability for 5G applications.

The VSWR characteristics provided in Fig. 7b further verify strong matching at the two resonances, where VSWR approaches 1, while a pronounced mismatch peak is observed near 30 GHz. Overall, the antenna demonstrates efficient dual-band performance with superior matching at the two operating frequencies.

Figure 7c illustrates the real and imaginary impedance characteristics of the proposed antenna over 26 – 40 GHz. The real part exhibits strong peaks near 32, 34, and 36.5 GHz, with the 32-GHz peak exceeding 800 Ω, indicating poor matching. Optimal impedance matching occurs at 27.83 and 37.81 GHz, where the real part is minimized and the imaginary part crosses zero, confirming the resonance.

The imaginary component alternates between inductive and capacitive regions, with a pronounced capacitive-type dip at 32 GHz coinciding with the large resistance peak. These impedance features align with the measured VSWR response, validating efficient dual-band operation at the target frequencies.

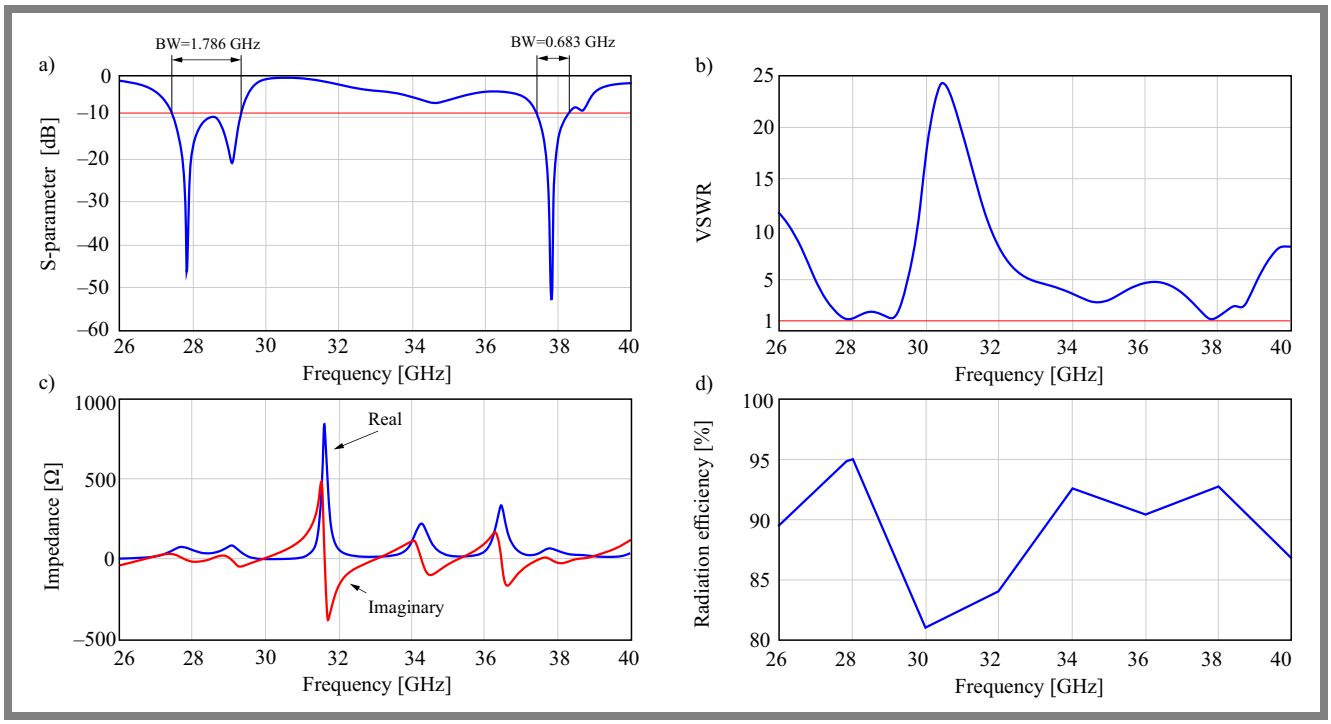

Fig. 7. Analysis of performance metrics for slotted SIW antenna: a) S-parameter, b) VSWR, c) impedance, and d) radiation efficiency.

Figure 7d shows that the proposed SIW antenna achieves excellent radiation efficiency at 26 – 40 GHz, exceeding 90% over the majority of the band. Peak efficiencies of approximately 95% occur at resonant frequencies of 27.83 and 37.81 GHz, indicating highly effective radiation. A slight reduction to approximately 85% is observed near 30 GHz, consistent with the impedance mismatch in this region. In general, the antenna maintains a stable efficiency range of 85 – 95%, demonstrating the low-loss performance of the SIW structure for millimeter wave 5G applications.

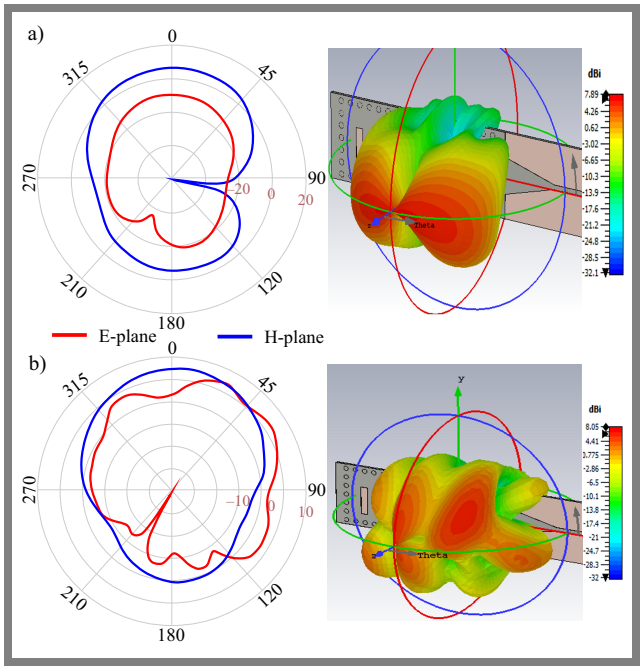

Fig. 8. Gain of the proposed slotted SIW antenna presented using polar and 3D views at: a) 27.83 GHz and b) 37.81 GHz.

For the proposed SIW antenna, simulated radiation patterns at 27.83 and 37.81 GHz (Fig. 8) exhibit quasi-omnidirectional behavior in the E-plane with minimal distortion (3 dB variation) and more directive H-plane patterns with well-defined main and side lobes. This hybrid behavior provides sectoral coverage with controlled elevation beamwidth, suitable for 5G base stations and spatial multiplexing.

The 3D radiation patterns show multilobed structures, with peak gains of 7.89 dB at 27.83 GHz and 8.05 dB at 37.81 GHz. At 37.81 GHz, the radiation is more symmetrical and uniform, with smaller additional lobes that improve multidirectional coverage, demonstrating the effectiveness for mmWave 5G applications.

Tab. 3. Evaluation values of the model test data for 28 and 38 GHz.

Band	28 GHz	38 GHz
Test loss (MSE)	0.032069	0.058893
Test (MAE)	0.120953	0.145232

Tab. 4. MAE, MSE, and RMSE values for 28 and 38 GHz.

Metrics	MAE	MSE	RMSE
Gain (28 GHz)	0.011	0.000	0.016
Efficiency (28 GHz)	0.089	0.014	0.120
Gain (38 GHz)	0.040	0.003	0.055
Efficiency (38 GHz)	0.138	0.051	0.226

Tab. 5. Simulated and predicted gain for the 28 GHz band using the ANN model.

No.	Simulated	Predicted	Error [%]
1	7.931	7.93	0.01
2	7.891	7.89	0.01
3	7.897	7.89	0.09
4	7.891	7.89	0.01
5	7.987	7.98	0.09
6	7.895	7.89	0.06
7	7.900	7.90	0.00
8	7.915	7.92	0.06
9	7.550	7.55	0.00
10	7.595	7.60	0.07

Tab. 6. Simulated and predicted radiation efficiency for the 28 GHz band using an ANN model.

No.	Simulated	Predicted	Error [%]
1	97.17	97.17	0.00
2	97.23	97.17	0.06
3	97.15	97.17	0.02
4	96.69	96.62	0.07
5	96.66	96.64	0.03
6	97.28	97.30	0.02
7	96.51	96.47	0.04
8	96.46	96.46	0.00
9	97.14	97.06	0.08
10	96.83	96.80	0.03

4.2. ANN Prediction Result of Model Performance

This subsection presents the performance of the ANN model to predict gain and efficiency at 28 and 38 GHz. Tables 3 and 4 summarize the evaluation metrics, showing strong overall predictive accuracy. For the 28 GHz band, the MSE and MAE are 0.0321 and 0.121, respectively, while for 38 GHz, they are slightly higher, at 0.0589 and 0.145. Detailed results per output indicate that gain prediction at 28 GHz is highly accurate (MAE 0.011, RMSE 0.016), compared to 38 GHz (MAE 0.040, RMSE 0.055). Efficiency predictions follow a similar trend, with 28 GHz achieving lower errors (MAE 0.089, RMSE 0.120) than 38 GHz (MAE 0.138, RMSE 0.226). Overall, the ANN model performs better at 28 GHz, likely due to differences in the complexity or signal characteristics between the bands.

Figures 9 and 10 present the actual vs. predicted values graphs, illustrating the performance of the ANN model in predicting gain and efficiency at 28 GHz. The predicted values (blue points) closely follow the ideal-fit line (red), indicating strong agreement with the actual targets. The tight clustering around the ideal line for both metrics demonstrates the model's capability to accurately capture complex patterns in multi-output regression tasks. These results validate the reliability and precision of the solution, enabling efficient optimization of antenna parameters while significantly reducing the need for time-consuming full-wave simulations.

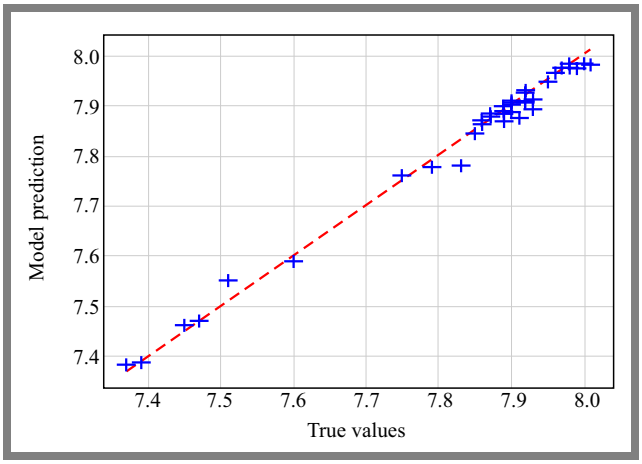

Fig. 9. Actual vs. predicted gain 28 GHz using the ANN model.

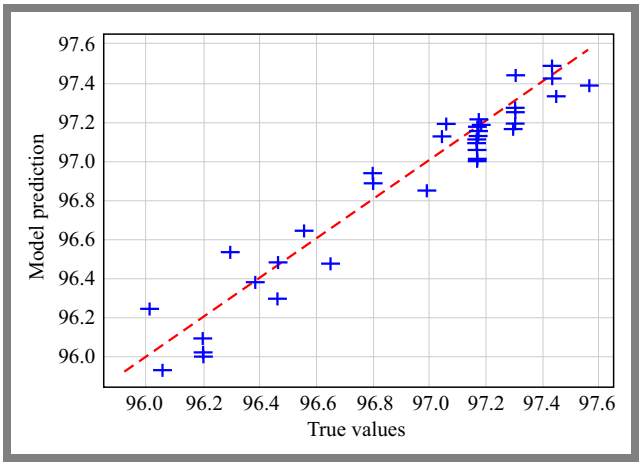

Fig. 10. Actual vs. predicted efficiency 28 GHz using the ANN model.

Tab. 7. Estimation of computational time for different processes.

Process	Approx. time
Single CST simulation	10 – 20 min
Dataset generation (158 samples)	26 – 52 hours
ANN training (100 epochs)	Few minutes
ANN prediction (per sample)	< 1 ms

4.3. ANN-based Validation of Gain

ANN-based gain validation confirms the model's exceptional predictive accuracy with minimal deviation between simulated and predicted values in the 28 and 38 GHz frequency bands. Table 5 provides a comparison between the simulated and predicted gain values across ten samples. At 28 GHz, the ANN demonstrates exceptionally high precision, with error percentages ranging from 0.00% to 0.09% and most predictions deviating by less than 0.06%. This near-perfect

Tab. 8. Performance comparison of the proposed design with related works, using 28/38 GHz antennas.

Ref.	OFB [GHz]	RL [dB]	BW [GHz]	Gain [dB]	Type	Advantages	Limitations	ANN/ML
[26]	28 / 38	–26 / –24	0.9 / 2	9 / 5.9	H-shaped slotted MIMO	High gain (28 GHz), wide bandwidth (38 GHz)	Complex design, low gain (28 GHz)	No
[27]	28 / 38	–19 / –20.75	0.5 / 0.79	12.7 / 15.5	SIW four-element array	Wide bandwidth	Lower reflection coefficient, narrow bandwidth	No
[28]	28 / 38	–17.35 / –34.4	0.982 / 0.354	7.05 / 8.32	Slotted SIW	Compact, simple, high gain (38 GHz)	Lower reflection coefficient, narrow bandwidth	No
[29]	28 / 38	–34.5 / –27.3	1.23 / 1.06	6.6 / 5.86	2-port MIMO	Wide bandwidth	Lower gain	No
[30]	28 / 38	–40 / –52	3 / 3.2	6.8 / 5	4-element MIMO	Wide bandwidth, higher reflection coefficient	Complex design, lower gain	No
[31]	28 / 38	–30 / –40	3.05 / 2.41	8.14 / 8.04	4-element MIMO	Wide bandwidth, higher reflection coefficient, high gain	Complex design, extended simulation times	No
[32]	28 / 38	–33 / –51	0.6 / 0.6	7.4 / 8.1	4-port MIMO	High gain, higher reflection coefficient	Narrow bandwidth	No
[33]	28 / 38	–45 / –45	4.5 / 2.4	4.5 / 7	Circular dual band MIMO	Wide bandwidth, higher reflection coefficient	Complex design, lower gain	ANN
[34]	28	–59.29	0.658	7.63	Patch antenna	Higher reflection coefficient and radiation efficiency	Single band, narrow bandwidth	ML
This work	28 / 38	–46.29 / –53.82	1.786 / 0.683	7.9 / 8.05	Slotted SIW	Simple, higher reflection coefficient, wide bandwidth (28 GHz), high gain	–	ANN

alignment illustrates the strong learning and generalization capability of the model for this band. In contrast, the 38-GHz band also shows reliable performance, though with slightly higher error margins. Prediction errors vary between 0.0127% and 0.244%, with an average still well within acceptable limits. Although still accurate, the slightly increased error in this band may stem from more significant variability or complexity in the electromagnetic characteristic at higher frequencies. Overall, the results confirm that the ANN model is highly capable of emulating CST simulation results with decent accuracy, particularly for the 28 GHz band.

4.4. ANN-based Radiation Efficiency

The ANN validation of radiation efficiency demonstrates excellent predictive accuracy, with strong correlation and close agreement between predicted results and CST simulation data across both frequency bands. As shown in Tab. 6, at 28 GHz, the model demonstrates exceptional accuracy, with prediction errors ranging from 0.00% to only 0.08%, providing an almost perfect match with the simulation results.

For the 38 GHz band, while maintaining impressive overall accuracy, the prediction errors show a modest increase, with deviations typically below 0.1%, but occasionally reaching up to 0.61% in isolated cases. This slight performance difference can be attributed to the inherently more complex electromagnetic interactions at higher frequencies, where smaller physical variations and increased sensitivity to geometric tolerances can influence efficiency characteristics. The low error percentages across both bands, with many predictions deviating less than 0.1% from the simulation results, confirm the robust generalizability of the ANN model.

5. Runtime Performance Comparison

Full-wave electromagnetic simulations using CST Studio Suite software are computationally expensive, particularly for parametric sweeps involving multiple geometrical variables. During the generation of the 158-antenna configuration dataset in CST, the average simulation time was approximately 10 – 20 min on a standard PC. In contrast, once trained, the proposed ANN model predicts antenna gain and radiation efficiency within milliseconds, allowing a near-instantaneous evaluation of new design candidates (Tab. 7).

Although the ANN training process requires an upfront computational cost, this expense is incurred only once. For subsequent optimization and parametric exploration, the ANN-based approach offers a significant increase in speed compared to repeated CST simulations. Therefore, the proposed ANN framework is particularly advantageous for iterative antenna optimization, sensitivity analysis, and inverse design, where hundreds or thousands of evaluations are required.

6. Comparison with Other Works

At operating frequencies of 28 and 38 GHz, the proposed antenna achieves bandwidths of 1.786 GHz and 0.683 GHz, respectively, surpassing most ANN/ML-based designs such as [34], though the results are slightly narrower than in [33], which incurs trade-offs in gain and complexity (Tab. 8). Its

reflection coefficient performance is exceptional, with values of –46.29 dB and –53.82 dB at the respective frequencies, indicating superior impedance matching and minimal signal loss. The antenna demonstrates strong gain performance across both bands, achieving 7.9 dB at 28 GHz and 8.05 dB at 38 GHz, outperforming several referenced designs, including [28]–[30], [33].

At 28 GHz, it remains competitive with high-gain designs such as [32], [33], while slightly trailing [31]. At 38 GHz, the proposed design outperforms the designs proposed in [26], [29], [30], matches the performance of [32] and competes effectively with [29]. Although it is less efficient than specialized arrays in [27], it offers a compact, high-performing, and well-balanced solution with significantly reduced design complexity.

The simple slotted SIW architecture provides advantages over multi-element MIMO configurations, such as [22], [31], making it suitable for compact, densely integrated deployment scenarios. Among machine learning-optimized designs like [33], [34], this work uniquely combines ANN-based optimization with decent performance across all important parameters without compromising simplicity.

7. Conclusions

This work presents the design and optimization of a compact dual-band slotted SIW antenna operating at 28 and 38 GHz, for 5G applications using CST full-wave simulations integrated with a hybrid AI-based framework combining artificial neural networks and polynomial regression. The developed predictive model demonstrated excellent agreement with electromagnetic simulations, achieving prediction errors below 0.1%, which confirms the high precision and reliability of the proposed approach. Moreover, the AI-assisted framework significantly reduced computational burden, enabling performance predictions within 2 to 3 s per model. This represents a reduction of several orders of magnitude in the computation time compared to conventional CST simulations, highlighting the effectiveness of the proposed methodology for rapid and efficient millimeter-wave antenna design.

Acknowledgments

The authors thank Universiti Malaysia Pahang Al-Sultan Abdullah for providing access to laboratory facilities and financial support under Internal Research Grant RDU220382.

References

[1] B. Kaur *et al.*, "Internet of Things (IoT) Security Dataset Evolution: Challenges and Future Directions", *Internet of Things*, vol. 22, art. no. 100780, 2023 (https://doi.org/10.1016/j.iot.2023.100780).

[2] N. Hussain *et al.*, "A Metasurface-based Low-profile Wideband Circularly Polarized Patch Antenna for 5G Millimeter-wave Systems", *IEEE Access*, vol. 8, pp. 22127–22135, 2020 (https://doi.org/10.1109/ACCESS.2020.2969964).

[3] B.A. Esmail and S. Koziel, "Design and Optimization of Metamaterial-based Dual-band 28/38 GHz 5G MIMO Antenna with Modified Ground for Isolation and Bandwidth Improvement", *IEEE Antennas and Wireless Propagation Letters*, vol. 22, pp. 1069–1073, 2022 (https://doi.org/10.1109/LAWP.2022.3232622).

[4] S.K. Hinga and A.A. Atayero, "Deterministic 5G mmWave Large-scale 3D Path Loss Model for Lagos Island, Nigeria", *IEEE Access*, vol. 9, pp. 134270–134288, 2021 (https://doi.org/10.1109/ACCESS.2021.3114771).

[5] J. Lu, G. Huang, D. Hu, and W. Kuai, "Multi-path Data Transmission System Based on 5G Communication Technology", *Journal of ICT Standardization*, vol. 12, no. 1, pp. 71-94, 2024.

[6] N.S. Khair *et al.*, "Recent Advances and Open Challenges in RFID Antenna Applications", *Enabling Industry 4.0 through Advances in Mechatronics*, vol. 900, pp. 507–517, 2022 (https://doi.org/10.1007/978-981-19-2095-0_43).

[7] B. Qian, X. Chen, and A.A. Kishk, "Decoupling of Microstrip Antennas with Defected Ground Structure Using the Common/differential Mode Theory", *IEEE Anten. and Wireless Prop. Lett.*, vol. 20, pp. 828–832, 2021 (https://doi.org/10.1109/LAWP.2021.3064972).

[8] B.S. Bari *et al.*, "Performance Comparison of Early Breast Cancer Detection Precision Using AI and Ultra-wideband (UWB) Bio-antennas", *Proc. of Conference on Cyber Security and Computer Science*, pp. 354–365, 2020 (https://doi.org/10.1007/978-3-030-52856-0_28).

[9] L.H. Xu *et al.*, "Bandwidth Enhancement of the Millimeter-wave Microstrip Linear Array with Loading of Shorting Pins", *IEEE Transactions on Antennas and Propagation*, vol. 71, pp. 1105–1110, 2022 (https://doi.org/10.1109/TAP.2022.3211398).

[10] B. Cheng and Z. Du, "A Wideband Low-profile Microstrip MIMO Antenna for 5G Mobile Phones", *IEEE Transactions on Antennas and Propagation*, vol. 70, pp. 1476–1481, 2021 (https://doi.org/10.1109/TAP.2021.3111330).

[11] M.M. Alam *et al.*, "Design and Optimization of II-shaped Slotted Dual-band SIW Antenna for 5G Applications", *Applications of Modelling and Simulation*, vol. 9, pp. 92–106, 2025.

[12] D. Prabhakar, P. Karunakar, S.R. Rao, and K. Srinivas, "Prediction of Microstrip Antenna Dimension Using Optimized Auto-metric Graph Neural Network", *Intelligent Systems with Applications*, vol. 21, art. no. 200326, 2024 (https://doi.org/10.1016/j.iswa.2024.200326).

[13] P. Kontou, S.B. Smida, and D.E. Anagnostou, "Contactless Respiration Monitoring Using Wi-Fi and Artificial Neural Network Detection Method", *IEEE Journal of Biomedical and Health Informatics*, vol. 28, pp. 1297–1308, 2023 (https://doi.org/10.1109/JBHI.2023.3337001).

[14] M.A. Haque *et al.*, "Machine Learning-based Approach for Bandwidth and Frequency Prediction for N77 band 5G Antenna", *Physica Scripta*, vol. 99, art. no. 026005, 2024 (https://doi.org/10.1088/1402-4896/ad1d40).

[15] H. Ibn-Khedher *et al.*, "Next-generation Edge Computing Assisted Autonomous Driving Based Artificial Intelligence Algorithms", *IEEE Access*, vol. 10, pp. 53987–54001, 2022 (https://doi.org/10.1109/ACCESS.2022.3174548).

[16] M.M. Alam *et al.*, "Machine Learning-based Approach for Bandwidth and Frequency Prediction of Circular SIW Antenna", *Journal of King Saud University–Engineering Sciences*, vol. 37, art. no. 9, 2025 (https://doi.org/10.1007/s44444-025-00010-0).

[17] S. Koziel, A. Pietrenko-Dabrowska, and L. Leifsson, "Antenna Optimization Using Machine Learning with Reduced-dimensionality Surrogates", *Scientific Reports*, vol. 14, art. no. 21567, 2024 (https://doi.org/10.1038/s41598-024-72478-w).

[18] J. Zhang, M.O. Akinsolu, B. Liu, and G.A. Vandenbosch, "Automatic AI-driven Design of Mutual Coupling Reducing Topologies for Frequency Reconfigurable Antenna Arrays", *IEEE Transactions on Antennas and Propagation*, vol. 69, pp. 1831–1836, 2020 (https://doi.org/10.1109/TAP.2020.3012792).

[19] Z. Xu, J. Liu, S. Huang, and Y. Li, "Gain-enhanced SIW Cavity-backed Slot Antenna by Using TE410 Mode Resonance", *AEU – International Journal of Electronics and Communications*, vol. 98, pp. 68–73, 2019 (https://doi.org/10.1016/j.aeue.2018.10.039).

[20] Y. Shi, W.J. Wang, and T.T. Hu, "A Transparent SIW Cavity-based Millimeter-wave Slot Antenna for 5G Communication", *IEEE Antennas and Wireless Propagation Letters*, vol. 21, pp. 1105–1109, 2022 (https://doi.org/10.1109/LAWP.2022.3158418).

[21] R. Olu-Ajayi *et al.*, "Building Energy Consumption Prediction for Residential Buildings Using Deep Learning and Other Machine Learning Techniques", *Journal of Building Engineering*, vol. 45, art. no. 103406, 2022 (https://doi.org/10.1016/j.jobe.2021.103406).

[22] Z.M. Abdhafith, R.A. Ali, and N. Abohmooed, "Impact of Substrate Thickness on the Rectangular Patch Antenna for 5G Communication System by CST Studio", *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 2, pp. 79–83, 2024 (https://waujpas.com/index.php/journal/article/view/60/37).

[23] R. Ramasamy and M.A. Bennet, "An Efficient Antenna Parameters Estimation Using Machine Learning Algorithms", *Progress In Electromagnetics Research C*, vol. 130, pp. 169–181, 2023 (https://doi.org/10.2528/PIERC22121004).

[24] D.S.K. Karunasingha, "Root Mean Square Error or Mean Absolute Error? Use their Ratio as Well", *Information Sciences*, vol. 585, pp. 609–629, 2022 (https://doi.org/10.1016/j.ins.2021.11.036).

[25] T.O. Hodson, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not", *Geoscientific Model Development*, vol. 15, pp. 5481–5487, 2022 (https://doi.org/10.5194/gmd-15-5481-2022).

[26] P. Liu *et al.*, "Patch Antenna Loaded with Paired Shorting Pins and H-shaped Slot for 28/38 GHz Dual-band MIMO Applications", *IEEE Access*, vol. 8, pp. 23705–23712, 2020 (https://doi.org/10.1109/ACCESS.2020.2964721).

[27] P. Kumawat and S. Joshi, "5G Dual-band Slotted SIW Array Antenna", *Journal of Taibah University for Science*, vol. 15, pp. 321–328, 2021 (https://doi.org/10.1080/16583655.2021.1978830).

[28] J. Singh, F.L. Lohar, and B.S. Sohi, "Design of Dual Band Millimeter Wave Antenna Using SIW Material for 5G Cellular Network Applications", *Materials Today: Proceedings*, vol. 45, pp. 5405–5409, 2021 (https://doi.org/10.1016/j.matpr.2021.02.106).

[29] A.E. Farahat and K.F. Hussein, "Dual-band (28/38 GHz) Wideband MIMO Antenna for 5G Mobile Applications", *IEEE Access*, vol. 10, pp. 32213–32223, 2022 (https://doi.org/10.1109/ACCESS.2022.3160724).

[30] A. Khabba, J. Amadid, S. Ibnyaich, and A. Zeroual, "Pretty-small Four-port Dual-wideband 28/38 GHz MIMO Antenna with Robust Isolation and High Diversity Performance for Millimeter-wave 5G Wireless Systems", *Analog Integrated Circuits and Signal Processing*, vol. 112, pp. 83–102, 2022 (https://doi.org/10.1007/s10470-022-02045-8).

[31] K. Cuneray, N. Akcam, T. Okan, and G.O. Arican, "28/38 GHz Dual-band MIMO Antenna with Wideband and High Gain Properties for 5G Applications", *AEU – International Journal of Electronics and Communications*, vol. 162, art. no. 154553, 2023 (https://doi.org/10.1016/j.aeue.2023.154553).

[32] R.R. Elsharkawy, K.F. Hussein, and A.E. Farahat, "Dual-band (28/38 GHz) Compact MIMO Antenna System for Millimeter-wave Applications", *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 44, pp. 1016–1037, 2023 (https://doi.org/10.1007/s10762-023-00943-0).

[33] L. Sellak *et al.*, "ANN-based Design of Miniaturized Circular Dual-band 4×4 MIMO Antenna for 28/38 GHz 5G mmWave Applications", *Telkomnika*, vol. 22, pp. 1280–1292, 2024 (https://doi.org/10.12928/telkomnika.v22i5.26347).

[34] S. Md, R. Islam, and S. Sarker, "Machine Learning Based on Patch Antenna Design and Optimization for 5 G Applications at 28 GHz", *Results in Engineering*, vol. 24, art. no. 103366, 2024 (https://doi.org/10.1016/j.rineng.2024.103366).

Md Mahabub Alam, M.Sc.
Faculty of Electrical and Electronics Engineering Technology
https://orcid.org/0009-0005-5036-4608
E-mail: mahbub.ete@gmail.com
Universiti Malaysia Pahang Al-Sultan Abdullah,
Pahang, Malaysia
https://www.umpsa.edu.my

Md Raihanul Islam Tomal, M.Sc.
Faculty of Computing
https://orcid.org/0009-0009-5290-7134
E-mail: raihanultomal@gmail.com
Universiti Malaysia Pahang Al-Sultan Abdullah,
Pahang, Malaysia
https://www.umpsa.edu.my

Ahmad Afif Mohd Faudzi, Ph.D.
Faculty of Electrical and Electronics Engineering Technology
https://orcid.org/0000-0002-3640-5259
E-mail: afif@umpsa.edu.my
Universiti Malaysia Pahang Al-Sultan Abdullah,
Pahang, Malaysia
https://www.umpsa.edu.my

Nurhafizah Abu Talip Yusof, Ph.D.
Faculty of Electrical and Electronics Engineering Technology
Centre for Research in Advanced Fluid and Processes
https://orcid.org/0000-0001-9762-5119
E-mail: hafizahs@umpsa.edu.my
Universiti Malaysia Pahang Al-Sultan Abdullah,
Pahang, Malaysia
https://www.umpsa.edu.my

An Adaptive Video Data Representation Model to Increase Delivery Efficiency in Next-Generation Networks

Anton Sorokun[1], Yurii Zadontsev[1], Dmytro Chyrva[2], Mykyta Zhyzhkin[2], and Andrii Bondarenko[1]

[1]*State University of Information and Communication Technologies, Kyiv, Ukraine,*
[2]*State University "Kyiv Aviation Institute", Kyiv, Ukraine*

https://doi.org/10.26636/jtit.2026.1.2411

Abstract — This study proposes an adaptive video stream representation model that provides dynamic adjustment of bitrate, frame rate, compression ratio, and frame structure based on a comprehensive analysis of network conditions, content priorities, and technical features of endpoint equipment. The research methodology includes mathematical modeling of video data transmission processes, analysis of radio channel noise immunity, algorithmic formalization of adaptive optimization of video parameters, and simulation modeling in Wi-Fi, 4G/5G, and PON networks. The results show that the proposed model provides a bandwidth reduction of 22 – 30% compared to static coding and classical ABR algorithms, reduces buffer time by 40 – 60%, increases delay stability to 150 ms in 5G and to 300 ms in 4G networks, and decreases packet loss rate to 1 – 3%. Its PSNR and SSIM metrics remain stable and device load is reduced by 15 – 20%.

Keywords — adaptive video data representation, channel modeling, next-generation networks, noise immunity, quality of experience, video stream optimization

1. Introduction

Modern telecommunication technologies, such as 5G and 6G networks, as well as widespread use of Wi-Fi 6/7 standards, caused an exponential growth in the transmission of video content. This growth is driven by the emergence and mass popularization of ultra-high resolution (4K/8K) applications and services, transmission of augmented and virtual reality (AR/VR) information, telemedicine and distance learning systems, and the Internet of Things (IoT), including narrowband solutions [1]. In all those applications, quality of experience (QoE) is of key importance for media-related services.

However, despite the increased capacity and reduced delay of next-generation networks, numerous fundamental problems remain that can impede efficient and stable video data delivery [2]. These include high volatility of wireless channel capacity, which can be caused by fading, interference, and user mobility, significant transmission delays and jitter, and possible packet loss during transmission. Limitation of computing resources and power consumption constraints affecting massive end-devices may be experienced as well.

Current, commonly used approaches to solving these problems demonstrate good effectiveness when relying on HTTP-based and adaptive streaming (AVR), software-defined networking (SDN), network function virtualization (NFV) and edge computing (EC) technologies. However, as current research shows, existing solutions may be based on client-side adaptation or direct routing optimization that does not skips, skip with errors or delays, a holistic model that could dynamically adapt the most fundamental representation of video data based on a comprehensive analysis of the network state and equal transmission of all characteristics of the content [3].

Hence, a problem arises of developing an adaptive video data representation model that would be capable of optimizing the parameters of a video stream at all stages of its life cycle, under the conditions of constantly evolving and dynamically changing new generation networks.

In this context, this research aims to develop an adaptive video data representation model to improve the efficiency and reliability of video content delivery (streaming) in 5G/6G, Wi-Fi, and NB-IoT networks through multi-parameter structured optimizations of video stream parameters in real time.

To achieve this goal, the following research objectives are formulated.

- analysis of current literature regarding the methods and technologies for adapting video streams, as well as identification of their advantages and disadvantages in new generation networks,
- development of an architecture of the adaptive video data representation model, which may include mechanisms for monitoring network status, classifying types of end devices, and analyzing video content priorities;
- identification of hybrid adaptation algorithms that provide dynamic adjustment of key video stream parameters based on incoming data,
- analysis of the developed model to evaluate its effectiveness based on key metrics.

The scientific novelty consists in the creation of an adaptive video data representation model that is immune to noise. Integration of dynamic bitrate control, frequency filtering, and

energy-saving mechanisms provides comprehensive stabilization of the video stream in the case of bandwidth fluctuations with regular or quasiregular components. This makes the developed model a promising solution for 5G/6G networks, where video quality is one of the key technological priorities.

2. Literature Review

The current stage of the development of telecommunication networks is characterized by the transition from 3G and 4G technologies to 5G networks, as well as by preparation for the advent of 6G networks. This process creates increased requirements for the efficiency and speed of video content delivery [2]. Growing traffic volumes, the emergence of new services and applications, and the need to ensure stable QoE require a revision of traditional approaches to video transmission. Current research identifies promising areas for solving these problems and offers technological solutions that can optimize video content transmission in the context of the development of new generation networks [4].

One of the current research areas is the development of fundamental technologies for video delivery over modern networks using adaptive bitrate streaming rate (ABR). Paper [5] covers the full scope of modern achievements and the next research vectors, showing the evolution from the simplest adaptive bitrate selection to intelligent systems that can take into account the network component, its state, content characteristics, and QoE. Another trend is the use of AI to optimize ABR, as demonstrated in [6], [7], where the benefits and effectiveness of the deep reinforcement learning (DRL) technique are relied upon.

These approaches enable bandwidth prediction, and thus allow for adaptation of video quality and maximization of QoE. Article [8] supports this idea by proposing a machine learning-based packet switching path for stabilization and the ability to transmit multiple video streams, which will be needed by companies broadcasting mass events.

The architectural shifts and computing frontiers are also highlighted by the authors of [9], who claim that the transition to next-generation networks is not possible without SDN, NFV, and EC. Article [10] stated that the combination of SDN and NFV plays a key part in ensuring QoE in ultra-advanced 6G networks. It emphasizes the importance of close and joint integration between the physical layer of a 6G network and the data presentation layer.

Furthermore, in [11] a reliable physical and link layer is created to optimize payload for video transmission over LoRa using HARQ. This direction was also considered in [12]. In terms of capacity planning in NOMA networks, the authors solve the problem of efficient resource utilization in the face of limited bandwidth, interference and obstacles.

Furthermore, the structural and selective methods described by the authors of [13] provide an alternative view of adaptive video representation by focusing on selective frame processing and the extraction of key components. They display efficient coding of structural components and segments of video frames with key information requirements. In [14], methods for motion compensation and syntactic data representation are developed. The developments aim to reduce bitrate redundancy through intelligent analysis of the content structure (video content), which is an important addition to conventional standard codes.

In [15], [16], the effect of non-fluctuating and multipath interference on signal immunity under conditions of limited bandwidth and interference during multiposition phase manipulation is analyzed. The results obtained serve as a basis for building more accurate models of the communication channel which, in turn, ensure the effective operation of adaptive upper-level algorithms [17].

In addition, the authors of [18] demonstrate that taking into account non-fluctuation components facilitates obtaining more accurate channel quality predictions. Article [19] proposes an integrated approach to group recommendation, caching, and video transmission in the IoT for smart transportation, which is also based on machine learning methods.

Papers [20], [21] focus on probabilistic caching and routing algorithms for video data in dense device-to-device (D2D) networks and VANETs, respectively. Using roadside units (RSUs) and DRL techniques, the research demonstrates that efficient video delivery in such networks requires not only client-side adaptation, but also intelligent resource allocation throughout the network.

Despite the progress in adaptive algorithms, the efficiency of video data processing remains a fundamental factor. Two approaches can be distinguished here. The first is presented in [22], where structural and selective coding methods are developed. The other is presented in [23], where the latest codecs and communication technologies are described. The authors emphasize the need to optimize coding algorithms to ensure high-quality video content transmission in next-generation networks.

To ensure end-to-end adaptation of the video delivery process from bitrate selection to caching at the network edge, it is necessary to use AI technologies, primarily DRL methods. Moreover, there is a need to establish an architectural imperative, i.e., to create solutions that will be closely integrated with SDN/NFV and MEC paradigms. In addition, analysis of the literature shows a lack of comprehensiveness of existing research: most studies focus only on certain aspects (client-side adaptation, server caching, or coding methods). Simultaneously, not all papers propose a holistic video data representation model capable of ensuring adaptability at all content lifecycle stages.

In this regard, a promising direction is to combine efficient coding structural methods with intelligent adaptation algorithms (as demonstrated in studies using DRL). This requires that the solution be integrated into a single architecture focused on the requirements of 6G networks [13]. Consequently, there is a need to develop a model capable of synthesizing optimal solutions in the areas of adaptive streaming, network edge computing, intelligent coding, and network resource management. This will ensure an efficient and flexible video

content delivery system in the context of future generation technologies.

3. Proposed Model

This study provides a methodology that combines theoretical, mathematical, algorithmic, and experimental approaches to create and test a model for adaptive video data representation in next-generation networks. First, a critical analysis of modern scientific sources is carried out to determine the state of the problem, assess the effectiveness of existing technologies for adaptive video transmission, and identify their limitations. Particular attention is paid to ABR methods, AI approaches based on reinforcement learning, and architectural solutions integrated with SDN, NFV, and MEC.

The proposed adaptive video data representation model is based on DRL integrated with the classical ABR mechanism. This approach enables intelligent decision-making for adaptation of video quality under dynamically changing network conditions. The architecture follows a multilayer design, which ensures coordinated operation across the application, control, and network domains. The model consists of three logical layers:

- The client layer (ABR agent) is responsible for real-time decision-making regarding the selection of bitrate for each video segment. The ABR agent operates on the client side and directly affects playback continuity, as well as perceived video quality.
- The intelligent control layer (DRL module) performs learning and inference tasks, enabling the system to predict and apply an optimal adaptation strategy. By analyzing the current and historical network state, the module continuously refines the bitrate selection policy to maximize QoE.
- The network delivery layer (SDN/NFV-oriented infrastructure) provides flexible management of network and computational resources through SDN and NFV technologies. It enables global network awareness and supports adaptive resource allocation, which complements the client-side adaptation process.

Such a multi-layer architecture overcomes the limitations of traditional ABR algorithms, which typically rely only on local bandwidth measurements and do not account for the global network state or cross-layer interactions.

At each decision step, the environment state is represented by a vector that captures both instantaneous network conditions and the playback context:

$$s_t = \{B_t, T_t, L_t, P_t, Q_{t-1}\} , \quad (1)$$

where B_t is an estimated available bandwidth, T_t is transmission delay and jitter, L_t is the occupancy of the packet loss rate, P_t is playback buffer, and Q_{t-1} is a quality level of the previously transmitted video segment.

This representation allows the agent to incorporate not only current measures, but also historical data, thus improving the stability and robustness of the adaptation process. Action at corresponds to the selection of one bitrate level from a predefined discrete set:

$$a_t \in \{r_1, r_2, \ldots, r_n\} , \quad (2)$$

where r_n is available for video encoding profiles.

This formulation reflects the practical constraints of adaptive streaming systems, where only a limited number of bitrate representations are supported. The core element of the DRL-based adaptation strategy is the reward function, which jointly considers perceived quality, stability, and resource efficiency. It is defined as a weighted combination of the following components:

$$R_t = \alpha \cdot QoE_t - \beta \cdot Rebuf_t - \gamma \cdot |Q_t - Q_{t-1}| - \delta \cdot E_t , \quad (3)$$

where QoE_t is an integrated quality of experience metric, $Rebuf_t$ is rebuffering duration, $|Q_t - Q_{t-1}|$ is a quality variation between consecutive segments, E_t is energy consumption of the client device, and α, β, γ, δ are weighting coefficients.

This reward formulation enables the model to optimize video quality, playback smoothness, and energy efficiency. As a result, the proposed approach differs from conventional ABR algorithms which primarily focus on bitrate maximization without explicitly accounting for stability and power consumption.

The next step consists in mathematical and algorithmic formalization of video data transmission processes, considering dynamic changes in channel bandwidth, delays, packet loss, and the impact of interference [9].

A system of equations is created to describe the interaction of network parameters and the video stream. An optimization objective function is defined to maximize QoE, while minimizing delays and excessive use of network resources. To assess noise immunity, the effect of impulse, multiplicative, and harmonic noise on the reception of BFM and DBFM signals is modeled. It helps identify the most critical types of distortion and takes them into account when building an adaptive model [24].

On the basis of theoretical and mathematical results, a software adaptive video data representation model is implemented [25]. It takes into account the network state and computing capabilities of the device, and the nature of the video content, providing a dynamic change in bitrate, frame rate, and compression parameters in real time.

The DRL agent is then trained using a hybrid data set that combines heterogeneous data sources to ensure robustness and generalization. The data set includes real network traces collected from Wi-Fi, 4G, and 5G environments, synthetic scenarios with harmonic and impulsive interference, and video content with different temporal dynamics, including sports, news broadcasts, and lecture recordings. Such diversity permits us to model both stable and highly variable transmission conditions.

To ensure convergence and training stability, several optimization techniques are applied, such as learning rate decay to prevent oscillations during weight updates, experience replay with filtering of anomalous states, and early stopping triggered by stabilization of the reward function. As a result,

the DRL agent demonstrated convergence after approximately 200 – 300 training epochs on average, confirming the practical feasibility and reproducibility of the learning process.

The proposed system employs an extended video delivery architecture based on the integration of SDN and NFV. The SDN controller is responsible for continuous monitoring of available bandwidth, link utilization, dynamic updates of flow tables, and prioritization of video traffic based on QoE-related metrics.

Flow update decisions are driven by control signals generated by the DRL agent, which enables coordination between client-side adaptation decisions and the global network state. At the NFV layer, the architecture supports adaptive placement and scaling of VNFs. In particular, video caches and transcoders are dynamically deployed closer to end users, while VNF instances are scaled according to peak traffic demand.

This approach reduces end-to-end latency through edge-level processing and provides measurable performance gain compared to traditional virtualized architectures that lack intelligent coordination mechanisms.

In contrast to previously reported solutions [5], where bandwidth decrease did not exceed approximately 15%, the proposed architecture achieves significantly higher efficiency. Integral DRL and HARQ mechanisms provide bandwidth reduction of 25 to 30%, stable PSNR and SSIM values under fluctuating network conditions, reduced buffering events and end-to-end latency, and additional energy savings of 15 to 20%. These improvements result from the synergy between intelligent client-side adaptation and coordinated SDN/NFV-based network control, rather than local optimization alone.

In general, the proposed model is characterized by a clearly defined architecture, a formally described DRL-based decision logic, a validated training methodology, and quantitatively measurable performance gain. These address conceptual limitations of existing approaches and substantiate both the scientific relevance and practical applicability of intelligent bitrate adaptation.

The effectiveness of software implementation is tested using simulation modeling in controlled environments that reproduce the characteristics of 4G, 5G, Wi-Fi 6/6E, PON and hybrid topologies [24]. In addition, statistical data processing methods are used to quantify the results, including analysis of means, variances, standard deviations, and correlations between network parameters and video quality [27]. A comparative analysis is used to compare the results of the developed model with traditional video delivery methods, such as H.264/H.265 and classical ABR algorithms.

The final stage consists in experimental verification of the proposed model under real and semi-real network conditions, which involves the reproduction of streaming video data of varying degrees of complexity and the evaluation of bandwidth, delay, transmission stability, end-point power consumption, and subjective video quality.

This approach allows to conduct a comprehensive test of the model and confirm its advantages in the context of next-generation networks.

4. Results

4.1. Mathematical Model of Adaptive Video Data Representation

Mathematical and algorithmic modeling aims to formalize video data coding processes, considering changes in channel bandwidth, delays, and packet losses [9]. Adaptive video stream representation describes the dynamic selection of video transmission parameters depending on network conditions. Network environment parameters are represented as time functions: channel capacity $C(t)$, transmission delay $L(t)$, and packet loss probability $P_{loss}(t)$. Furthermore, parameters of the video stream, such as bit rate $R(t)$, frame rate $F(t)$, and compression quality $Q(t)$ are formalized as variables that can be adjusted in real time.

The system is optimized by maximizing QoE – a metric impacted by weighting factors $\omega_1, \ldots, \omega_4$ which determine the balance between playback quality and data transmission stability. The objective function of the optimization process reflects the task of maximizing QoE while reducing delays, packet loss, and excessive bandwidth usage. Its generalized formula is given by:

$$\max_{R(t),F(t),Q(t)} U(QoE) = \omega_1 \cdot Q(t) - \omega_2 \cdot \frac{R(t)}{C(t)} - \omega_3 \cdot L(t) - \omega_4 \cdot P_{loss}(t) \, . \tag{4}$$

The model is subject to a number of constraints that reflect the physical limits of the network environment, acceptable bitrate levels, frame rates, and playback continuity requirements:

$$\begin{aligned} 0 \leqslant R(t) \leqslant C(t), \\ F_{min} \leqslant F(t) \leqslant F_{max}, \\ Q_{min} \leqslant Q(t) \leqslant Q_{max} \, . \end{aligned} \tag{5}$$

Rather than relying on heuristic adaptation rules, the developed approach is formulated as a constrained optimization problem with a unified objective function and well-defined decision variables. Consistent with the objective function presented in Eq. (1), the quality of experience for the i-th user is defined as follows:

$$QoE_i = \omega_1 Q_i - \omega_2 D_i - \omega_3 R_i \, , \tag{6}$$

where Q_i represents objective video quality metrics (PSNR, SSIM), D_i is the end-to-end delivery delay and R_i is the total consumption of network resources. The weighting coefficients ω_1, ω_2, and ω_3 are identical to those used in Eq. (4) and explicitly control the trade-off between visual quality, latency, and resource utilization efficiency.

Maximization of the global QoE function in Eq. (4), subject to the constraints specified in (5), is achieved through a hierarchical optimization strategy operating at three interdependent levels. At the client level, adaptive bitrate selection and buffer management act on short-term variations of $C(t)$, $D(t)$, and $P_{loss}(t)$, ensuring local compliance with playback continuity and delay constraints.

At the edge level, transcoding and caching decisions are optimized to reduce processing latency and computational

load while satisfying resource constraints imposed by the edge infrastructure. At the network level, bandwidth allocation and traffic control mechanisms enforce global capacity and loss constraints, ensuring system-wide stability.

Importantly, these levels do not operate independently. Instead, each level optimizes a subset of decision variables that contribute to the same global objective function in Eq. (4). This coordinated formulation guarantees mathematical consistency between local adaptation mechanisms and the end-to-end QoE optimization target, addressing the previously noted lack of clarity in the interaction between architectural components.

Furthermore, the packet loss prediction component is extended in accordance with the constraints in (5). Temporal features are incorporated to model the time evolution of $P_{loss}(t)$, enabling the capture of short- and long-term loss trends. Furthermore, correlation analysis is performed to quantify the dependence between packet loss, delay $D(t)$, jitter, and channel capacity $C(t)$.

Next, the extended model is validated in heterogeneous scenarios, including LTE, 5G, and mobility-driven handover conditions, demonstrating that the proposed optimization framework remains stable under varying network dynamics. These improvements directly improve the reliability of loss estimation and allow the adaptive video representation system to proactively adjust transmission parameters within the formally defined optimization framework of (4), (5), thus improving QoE and resource utilization efficiency.

The next stage of the study focuses on analyzing noise immunity of coherent reception of multiphase shift keying signals (MPSK) subjected to noise and non-fluctuating interference. The level of reliability at the physical layer is quantified by the bit error probability (BER) parameter, describing the robustness of symbol detection in the presence of noise.

For binary phase-shift keying (BPSK) over an additive white Gaussian noise (AWGN) channel, the BER is given by the following formula:

$$P_b = Q\sqrt{\frac{2E_b}{N_0}}\,, \tag{7}$$

where Q is the Gaussian Q-function, P_b is a symbol rate, E_b is the energy per transmitted bit, and $N_0/2$ is the one-sided power spectral density of AWGN.

At the source coding level, the distribution of bits within a video frame is modeled using an entropy-based approach. If a video frame is composed of N blocks, and the random variable $X = \{x_1, x_2, \ldots, x_N\}$ represents the number of bits allocated to each block, with probability distribution $p(x_i)$, then the entropy of the frame, which represents the minimum achievable average number of bits required for lossless encoding, is defined by Shannon's formula:

$$H(X) = -\sum_{i=1}^{N} p(x_i)\log_2 p(x_i)\,. \tag{8}$$

If the encoder assigns b_i bits to block x_i, the average bit consumption of the frame is:

$$\bar{B} = \sum_{i=1}^{N} p(x_i)\, b_i\,. \tag{9}$$

The compression efficiency is evaluated as:

$$\eta = \frac{H(X)}{\bar{B}},\ \eta \leqslant 1\,, \tag{10}$$

where values of η approaching unity indicate more efficient utilization of the available bitrate.

To account for structural differences between video frame types, frames are classified into a set $T = \{I, P, B\}$. Entropy is evaluated separately for each type. For a frame of type T, the entropy is expressed as:

$$N_T = \sum_{i=1}^{N_T} p_T(x_i)\log_2 p_T(x_i)\,, \tag{11}$$

where N_T is the number of blocks in a frame of type T, and $p_T(x_i)$ is the corresponding probability distribution.

This formulation supports adaptive bitrate allocation based on frame complexity.

To stabilize data transmission under time-varying network conditions, a dynamic congestion control (DCC) mechanism is used. At time t, the system monitors the transmission delay D_t, available bandwidth C_t, and packet loss rate P_t. On the basis of these parameters, the target transmission rate is computed as follows:

$$R_t^* = \alpha\, C_t - \beta D_t\,, \tag{12}$$

where α and β are weighting coefficients.

The actual rate is updated using a smoothing rule:

$$R_t = \gamma R_{t-1} + (1-\gamma)\, R_t^*\,. \tag{13}$$

If $\Delta R_t = R_t - R^*$, where R^* is the steady-state optimal rate, then the adaptation process satisfies the following inequality:

$$|\Delta R_t| \leqslant \gamma^t |\Delta R_0|\,, \tag{14}$$

guaranteeing exponential convergence and confirming stability of the proposed congestion control mechanism under dynamic network conditions.

The multiplicative noise, which models the fading of the signal amplitude in the radio channel, is described by the following model:

$$S_i(t) = \mu(t) A_0 \cos\left(\omega_0\, t + \varphi_i + \varphi_n\right)\,, \tag{15}$$

where $\varphi_i = \frac{i2\pi}{M}, t \in (0, t_s), i = 0, 1, \ldots, M-1$, $\mu(t)$ is an amplitude coefficient that follows the Rice distribution.

The pulse noise, which models short high-frequency pulses, is set by a formula corresponding to a chaotic sequence of short radio pulses:

$$S_3(t) = A(t)\cos\left(\omega_0\, t + \varphi_n\right),\ 0 < t < \tau_3\,. \tag{16}$$

Harmonic noise is regarded as periodic oscillations that could be superimposed on the carrier frequency:

$$S_3(t) - \mu\, A_0 \cos\left[(\omega_0\, \Delta\omega_p)\, t + \varphi_p\right]\,. \tag{17}$$

The noise is described by AWGN with the correlation function $R(\tau) = \delta(\tau)$, meaning that there is no correlation between the noise values at different times:

$$I_j = \{I_i\},\ i \neq j,\ j = 1, \ldots, M-1\ . \tag{18}$$

The probability of correct reception of a symbol is determined through the conditional probabilities of exceeding the voltage at the output of the correlator for the correct symbol over the voltage for other symbols. This physically means that the system receives the symbol j if the corresponding correlation detector generates the largest signal:

$$P_{sj} = \prod_{i \neq j}^{M-1} i - 0\ p(I_j > I_i)|j\ , \tag{19}$$

where $p(I_j > I_i)$ is the conditional probability that when the j-th symbol is transmitted, the voltage at the output of the j-th correlator exceeds the voltage at the output of any other i-th correlator.

Since the probabilities of all signals in the channel are equal, and the MPSK constellation has a symmetric structure, the total probability of correct reception P_s is obtained through the analysis of a single symbol to calculate the probability of error P_{es} and the corresponding bit error probability P_{eb}. Subsequently, to assess the quality of reception, the composite signal $x(t)$, which includes the useful signal, noise components, and nonfluctuating interference, is taken into account:

$$P_s = \prod_{j=1}^{M-1} p(I_o > I_i)|0\ . \tag{20}$$

The total probability of a character error is determined by the following ratio:

$$P_{es} = 1 - P_s\ . \tag{21}$$

At sufficiently high values of the signal-to-noise ratio, the relationship between the symbol error probability P_{es} and the bit error probability P_{eb} is defined by:

$$P_{eb} = \frac{P_{es}}{\log_2 M}\ . \tag{22}$$

The receiver of a data transmission system receives an AWGN $n(t)$ and non-fluctuating interference along with the useful signal. To estimate its noise immunity, three most common types of such interference $s_n(t)$ are considered: harmonic noise with a frequency deviation ω_p relative to the central frequency of the MPSK signal spectrum, chaotic impulse noise with a pulse duration τ_x, and multiplicative noise that obeys the Rice distribution law [26]. Their amplitude is comparable to the amplitude of the useful signal. The total signal received at the receiver's input is described by the following ratio:

$$x(t) = s_i(t) + s_n(t) + n(t)\ , \tag{23}$$

where $s_i(t)$ is the useful signal, $s_n(t)$ is the nonfluctuating interference, $n(t)$ is the AWGN.

Figure 1 shows a block diagram of a data transmission system that demonstrates the interaction of a signal source, a communication channel with noise, and a receiving path. It shows the main elements involved in signal transformation and distortion during transmission and serves as a basis for further modeling.

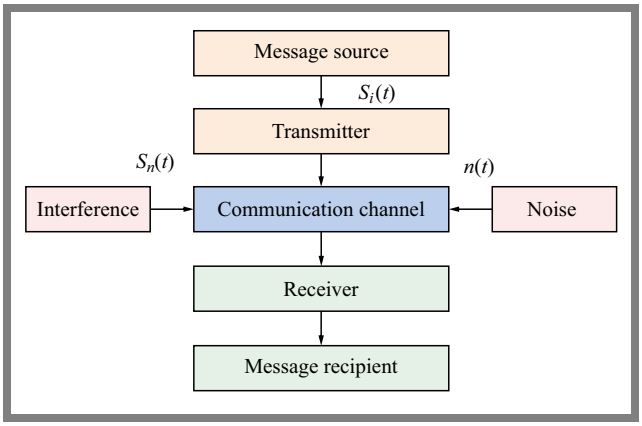

Fig. 1. Block diagram of the data transmission system.

4.2. Analysis of MPSK Signal Noise Immunity

To estimate the probability of error during the reception of a channel symbol with index i, a vector representation of the MPSK signal, noise, and interference components is used, as shown in Fig. 2. It illustrates how noise shifts the signal vector relative to the ideal points of the MPSK constellation. The greater the offset, the more likely the receiver is to assign the signal to the wrong symbol, which directly increases the probability of P_{es} and P_{eb} error.

In the modeling, the initial phase φ_p of the interference vector $s_n(t)$ was fixed, which facilitated considering the total vector $s_i(t) + s_n(t)$ as conditionally deterministic. Meanwhile, the randomness of the processes at the output of the demodulator correlators is preserved because of additive noise, which follows the normal (Gaussian) distribution. It determines the conditional statistical characteristics of the correlation integrals, which were relative to a fixed phase of the interference: mean values $m_0, \ldots, m_i, \ldots, m(M-1)$, variances $D_0, \ldots, D_i, \ldots, D(M-1)$, mutual torques M_{0-i}, and process parameters $y = I_0 - I_i$, which were fed to the input of the device for comparing correlation integrals.

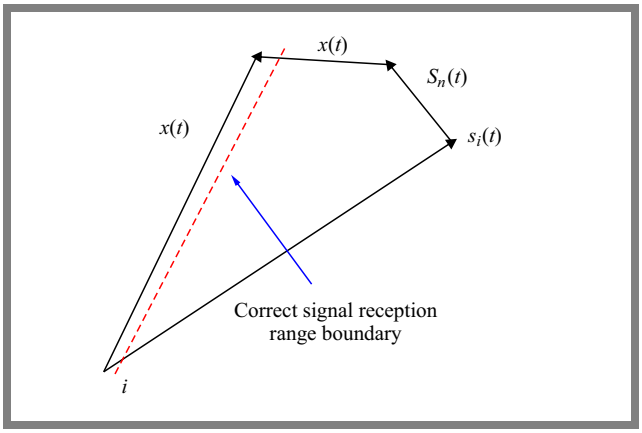

Fig. 2. Vector representation of MPSK signal, noise, and interference.

Tab. 1. Energy loss of MPSK signal reception against low-intensity nonfluctuating interference.

Noise type	Energy loss at a given error probability $P_{eb} = 10^3$ [dB]				
	M = 2	M = 4	M = 8	M = 16	M = 32
Harmonic	0.22	0.41	1.2	4.1	The loss is critically high
Multiplicative	0.18	0.24	1.0	2.6	
Pulse	The loss amount is insignificant and cannot be estimated			0.1	

Tab. 2. Energy loss of MPSK signal reception against non-fluctuating high-intensity interference.

Noise type	Energy loss at a given error probability $P_{eb} = 10^3$ [dB]				
	M = 2 (10^{-3})	M = 4 (10^{-3})	M = 8 (10^{-2})	M = 16 (10^{-1})	M = 32 (10^{-1})
Pulse	0.2	0.4	0.5	0.2	0.4
Harmonic	4.1	8.2	1.2	4.1	The loss is critically high
Multiplicative	3.3	6.3	The loss is critically high		

The next step is to obtain the unconditional probability of bit error. Since the conditional probability $P_{eb}(\varphi_n)$ depends on the initial phase of the interference, it should be averaged over the distribution of the random variable φ_n:

$$P_{eb} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{eb}(\varphi_n)\, d\varphi_n \,. \tag{24}$$

It was impossible to obtain an exact analytical expression for error probability, so the method of numerical averaging was used to determine the final results. The modeling results show that the noise immunity of MPSK signal reception depends significantly on the frequency deviation of the harmonic noise. As frequency disturbance increases, the effect of the interference gradually weakens: error probability decreases and approaches the value characteristic of the mode without noise.

The minimum values of error probability are observed when frequency disturbance corresponds to the zeros of the envelope of the useful signal's energy spectrum. For values of $M \leqslant 8$, when the interference is decayed, it is established that at frequency disturbances exceeding a certain threshold value, the influence of such harmonic noise becomes so small that it could be neglected without significant deterioration in reception accuracy.

The study analyzes the decrease in the noise immunity of coherent reception of MPSK signals under the influence of non-fluctuating interference. The situation in which one of the interferences is present in the communication channel along with noise is considered. The purpose of the analysis was to determine which types of noise cause the greatest deterioration in signal reception characteristics and identify patterns of change in the probability of error depending on the parameters of these interferences.

The modeling considers targeted noise when the center frequencies of their spectra coincide with the carrier frequency of the useful signal ω_0, which creates critical conditions for the receiver. The following parameters were adopted for certain interference types: $\alpha = 1$ for scanning noise, $\alpha = 0.5$ for retransmission, $\alpha = 2$ for frequency modulated interference, and $\alpha = 0.5$ for pulse noise. Based on these parameters, a comparative analysis of the receiver's noise immunity was performed.

The analysis proved that under these conditions, harmonic noise is the most impacting factor, since it causes the greatest reduction in noise immunity. Further, in the order of decreasing harmful effects, there are multiplicative noise, frequency-modulated interference, scanning interference, and pulse noise. This order reflects the general average trend, but can vary depending on the noise parameters and the operating modes.

The analysis demonstrates that the effectiveness of harmonic and scanning noise is reduced considerably when there is significant frequency discord with respect to the center frequency of the useful signal spectrum.

The results of modeling with low intensity interference ($\mu = 0.1$) are shown in Tab. 1.

For high-intensity interference ($\mu = 0.5$), the patterns were preserved, but the scale of energy losses increased significantly. The results are summarized in Tab. 2. Harmonic noise remained the most important phenomenon. Thus, for $M = 4$, the loss reached 8.2 dB, while for $M = 32$ it became quite large. Multiplicative noise also caused significant losses, while pulse noise caused the least damage (Tab. 2).

The analysis of the data shows that harmonic interference is the most destructive variety, as it creates the maximum energy loss in a wide range of parameters. It was followed by retransmission (multiplicative), frequency-modulated, scanning, and pulse interference in terms of harmfulness. The effectiveness of harmonic and scanning interference decreases significantly with frequency discord relative to ω_0, i.e. when the spectral overlap between the signal and the interferer decreases. Therefore, when developing methods and means to receive MPSK signals in difficult conditions, algorithms for suppressing harmonic and multiplicative interference should be used first.

To increase the stability of MPSK signal reception in the presence of harmonic interference, it is advisable to use interference compensation algorithms, in particular adaptive

Tab. 3. Comparative analysis of bandwidth utilization under different network conditions.

Network conditions	Traditional ABR	Stable H.264/H.265	Proposed HARQ mechanism
Wi-Fi (stable)	5.8 Mbps	5.8 Mbps	4.7 Mbps (≈18% of economy)
4G (variable)	6.2 Mbps	6.1 Mbps	4.3 Mbps (≈30% of economy)
5G (stable)	9.1 Mbps	9.3 Mbps	6.8 Mbps (≈27% of economy)
Unstable networks	7.5 Mbps	7.8 Mbps	5.4 Mbps (≈28% of economy)

notch and tracking filters, which provide effective attenuation of the influence of interfering components in the signal.

The results of the study confirm that the proposed model of adaptive video data representation offers a significant improvement in key parameters of multimedia transmission in realistic network environments by integrating the adaptive hybrid automatic repeat request (HARQ) mechanism with the payload optimization strategy. The analysis was performed for various types of networks, including stable Wi-Fi environments, 4G/5G mobile networks with variable channel conditions, and scenarios with frequent packet loss and bandwidth fluctuations. One of the key criteria for transmission efficiency was the level of bit errors in the received packets and the number of retransmissions, which directly depended on the channel parameters and the error correction strategy.

The proposed HARQ mechanism combined adaptive re-requesting with an optimal modulation and coding scheme, which reduced the relative bit error rate to 1 – 3% in networks with a high probability of loss (10 – 20%), while in models without HARQ, this figure reached 5 – 8%. In addition, the average number of retransmissions is reduced to 0.8 – 1.4 retransmissions per packet under peak conditions, compared to 1.9 – 2.7 retransmissions for traditional schemes without adaptive HARQ.

This optimization reduced channel congestion, contributed to more efficient use of network resources, and is directly correlated with a reduction in video content delivery latency, which ensures more stable operation of streaming services. Further analysis of the bandwidth shows that the integration of the adaptive HARQ strategy with the adaptive video data representation increases the efficiency of the communication channel.

The comparative results of the simulations and experimental tests are shown in Tab. 3.

As can be seen in the table, the proposed model provides bandwidth improvements of 25 – 30% under different network environments, which is significantly higher than the performance of basic adaptive coding and transmission schemes. This effect is achieved by reducing the number of retransmissions and adaptively adjusting the bitrate according to the current channel state.

Objective video quality metrics also demonstrate a high degree of stability of the proposed approach. Even under conditions of significant bandwidth fluctuations, the PSNR value in most scenarios remains within 36 – 40 dB, while in traditional schemes it equals 34 – 38 dB. SSIM remains at the level of 0.90 – 0.95, which corresponds to high video perception quality. The low correlation coefficient between the bandwidth and PSNR/SSIM values ($r \approx 0.15$) confirms the model's ability to maintain stable video quality even under unstable channel conditions.

The positive effect of HARQ optimization is also evident in the time characteristics of the streaming process. In 4G mobile networks, the buffering time is reduced by almost half and does not exceed 4.1% of the total playback time compared to traditional ABR schemes. Transmission delay stabilizes at up to 300 ms in 4G and up to 150 ms in 5G networks, and the delay variation (jitter) decreases by 35 – 45%, ensuring continuous video playback without noticeable interruptions.

In addition to network and quality characteristics, the proposed approach exerts a positive impact on power consumption and computing load of end devices. Adaptive HARQ control reduces the load on the processor and network adapter by 15 – 20%, significantly extending battery life during long video playbacks on mobile devices and in IoT environments. In general, the HARQ-based payload optimization strategy provides a significant reduction in error rate and the number of retransmissions, effective bandwidth savings (25 – 30%), stable video quality with high PSNR and SSIM, a significant reduction in buffering and latency, and an energy efficiency increase in end devices.

4.3. Analysis of the Autocorrelation Demodulator of DPSK Signals

It is important to consider the operation of an autocorrelation demodulator for differentially encoded phase-shift keying (DPSK) signals in a radio channel under the influence of Rayleigh fading and non-fluctuating interference. Two consecutive bursts of a DPSK signal corresponding to the i-th and (i-1)-th clock periods can be represented as follows:

$$S_i(t) = A\cos(\varphi_0 t + \varphi_1),\ t \in [(i-1)T_s, T_s]\,, \qquad (25)$$

$$S_{i-1}(t) = A\cos(\varphi_0 t + \varphi_2),\ t \in [(i-2)T_s, (i-1)T_s]\,, \quad (26)$$

where $A = \sqrt{\frac{2E_s}{T_s}}$ is the signal amplitude, $E_s = E_b \log_2 M$ is the energy of one channel symbol, E_b is the energy consumed to transmit one bit of information, ω_0 is the carrier frequency, φ_i, φ_{i-1} are burst phases the difference between which encodes information about the channel symbol and can take one of M possible values that differ by $2\pi/M$.

Due to Rayleigh fading, the signal amplitude is multiplied by a random coefficient β, which characterizes the depth of fading and describes the multipath effect:

$$\overline{\omega}(\beta) = \frac{\beta}{\sigma^2}\, e^{-\frac{\beta^2}{2\sigma^2}}\,. \qquad (27)$$

In the case of $M = 2$, phase difference can be 0 or π. Therefore, error probability is formed from a limited number of phase combinations, for which the corresponding values are provided in Tab. 4.

Tab. 4. Individual error probabilities at $M = 2$.

Phase combinations	Error probability	Meaning
$\varphi_1 = 0,\ \varphi_2 = 0$	$P_{e1} = \frac{1}{2}\frac{1}{c_1 q+1}$	$c_1 = (1+\mu\cos\eta)^2$
$\varphi_1 = \pi,\ \varphi_2 = \pi$	$P_{e2} = \frac{1}{2}\frac{1}{c_2 q+1}$	$c_2 = (1-\mu\cos\eta)^2$
$\varphi_1 = 0,\ \varphi_2 = \pi$	$P_{e4} = P_{e3} = \frac{1}{2}\left(1 - \frac{\sqrt{c_1 c_2 q}}{\sqrt{(c_1 q+1)(c_2 q+1)}}\right)$	$c_1 = (1+\mu\cos\eta)^2$
$\varphi_1 = \pi,\ \varphi_2 = 0$		$c_2 = (1-\mu\cos\eta)^2$

Tab. 5. Comparative analysis of user access technologies.

Characteristics	Historical/outdated technologies	Modern and promising technologies
	xDSL (ADSL2+), Ethernet (Fast Ethernet), Wi-Fi (802.11n), WiMAX	PON (GPON/XGS-PON), Wi-Fi 6/6E (802.11ax), 5G NR (fixed wireless access)
Transmission speed	24 Mbps ↓ / 1.4 Mbps ↑	100 Mbps
Application field	POTS/ISDN, basic Internet	LAN, office, private networks
Range of action	900 – 2500 m	100 – 150 m
Access type	Copper cable, wireless	Twisted pair cable, fiber optic cable, wireless
Key advantages	Low cost, reuse of copper lines, easy deployment, minimal hardware requirements	Reliability and low delay, stable throughput, easy scaling to gigabit/10G
Main limitations	Low speed, dependence on line length, EMI and noise on the cable, physical wear of copper cables, outdated technology	Limited range (up to 150 m), requiring cabling, poor flexibility in older buildings

The total probability of error of a symbol (bit) at $M = 2$ is defined as follows:

$$P_{ef2} = \frac{1}{4}(P_{e1} + P_{e2} + P_{e3} + P_{e4}) = \frac{1}{4}\left(\frac{0.5}{c_1 q+1} + \frac{0.5}{c_2 q+1} + 1 - \frac{\sqrt{c_1 c_2 q}}{\sqrt{(c_1 q+1)(c_2 q+1)}}\right). \tag{28}$$

For the case of $M = 4$, the signal was processed in the in-phase (IP) and quadrature (Q) channels, after which the phase angle of the signal vector in the complex plane was determined. The probability of false reception of a symbol was defined through the probability of its correct reception, i.e., through the probability that the signal vector falls into the corresponding quadrant of the complex plane $x - y$, $\Delta\varphi = \pi/4$ – first quadrant $(x > 0, y > 0)$, $\Delta\varphi = 3\pi/4$ – second quadrant $(x < 0, y > 0)$, $\Delta\varphi = 5\pi/4$ – third quadrant $(x < 0, y < 0)$, $\Delta\varphi = 7\pi/4$ – 4th quadrant $(x > 0, y < 0)$. The probability of correct reception was established as the probability that noise and interference would not move the point to another quadrant. For $\Delta\varphi = 5\pi/4$ and $\Delta\varphi = 7\pi/4$, it is as follows:

$$p[Y_0(t) < 0] = \frac{1}{2}\left(1 - \frac{R}{\sqrt{1-S^2}}\right). \tag{29}$$

Since the analytical expression was extensive, simplified formulas were obtained for a channel without harmonic noise ($\mu = 0$):

$$P_{ef4} = 1 - \frac{1}{4}\left(1 + \frac{q}{\sqrt{(q+2)(q+1)}}\right)^2. \tag{30}$$

The error probabilities in the cases with and without amplitude fading are related by the following ratio:

$$P_{ef} = \int_0^\infty P_e(A)\,\overline{\omega}(A)\,dA\,. \tag{31}$$

To estimate the symbol error probability in a radio channel without Rayleigh fading, a technique based on the inverse Laplace-Carson transform is applied. It helps move from the SNR distribution function to the error probability:

$$P_e = L^{-1} P_{ef}\,. \tag{32}$$

The corresponding bit error probabilities for $M = 4$ when using the Gray code were calculated using the following formula:

$$P_{eb4} = 1 - \sqrt{1 - P_{e4}}\,. \tag{33}$$

At a high level of harmonic noise, additional energy losses could reach several decibels. For $M = 4$, they significantly exceed the losses present at $M = 2$. Even in a channel without fading, harmonic noise increases the error probability and the required signal-to-noise ratio while reducing the energy efficiency of autocorrelation reception. Therefore, DPSK signals with autocorrelation demodulation require additional methods to compensate for harmonic and multiplicative noise in communication systems.

4.4. Experimental and Simulation Results

Simulation modeling was one of the key tools for testing the effectiveness of the developed model under controlled conditions. Due to simulation, it is possible to reproduce the

Tab. 6. Comparison of quality indicators and the utilization of network resources.

Delivery method	Average bandwidth	Peak signal-to-noise ratio	Structural similarity index measure	Buffer
H.264 H.265	5.8 Mbps	36.2 dB	0.92	12.5%
ABR	4.9 Mbps	35.8 dB	0.91	8.3%
Adaptive presentation	3.6 Mbps	36.0 dB	0.92	4.1%

behavior of information and communication networks under different loads, types of traffic, and nature of interference to evaluate the real performance of a system without resorting to expensive or complex field experiments [2].

The study examines the structural and functional schemes of information and communication networks of various classes to organize video conferencing systems based on the transmission of video data over IP and ISDN networks. These networks have significant differences in terms of their infrastructure, bandwidth, delays, and reliability, which leads to different requirements for subscriber access channels. To compare the main last mile technologies, an analytical table is compiled to demonstrate the evolution of access from outdated solutions (xDSL) to modern high-speed systems (PON, 5G, Wi-Fi 6/6E) – Tab. 5.

Table 5 shows that fiber optic networks (PON) continue to be the gold standard for fixed access, providing gigabit speed. Wireless technologies, in turn, are divided into two key niches: Wi-Fi 6/6E/7 for high-speed indoor access and 5G FWA as a flexible alternative to wired networks, especially in rural areas or in conditions where fiber optic installation is not economically feasible.

Copper technologies have been gradually losing their relevance because xDSL and WiMAX no longer meet modern speed and capacity requirements. Therefore, the choice of access technology is determined by a combination of characteristics such as deployment cost, the required speed, local features, and specific parameters. The effectiveness of adaptive video data representation was evaluated based on the model developed, as it dynamically changes the parameters of the video stream taking into account the network state and capabilities of the user device. The model considered the width of the available channel, the amount of delay, the level of computing resources, and the presence of interference of various kinds.

To test its effectiveness, a series of experiments were performed using video streams of different resolutions and bitrates, evaluating the performance under variable network parameters. The evaluation took into account such indicators as the average bandwidth used by the video stream during transmission, the time of initial download and subsequent buffer, the actual video playback quality determined by PSNR and SSIM metrics, and the level of packet loss reduction during transmission.

The results confirm that the developed model provides a significant increase in the efficiency of video data transmission and reduces bandwidth utilization by approximately 25 – 30% compared to static coding methods, while simultaneously maintaining high playback quality. Video quality remains stable in the face of significant fluctuations in channel parameters, which was especially important for networks with unstable bandwidth [29]. Additionally, the model demonstrates a considerable reduction in buffer frequency. Thus, the total number of such events and their duration is reduced by 40 – 60%, improving viewing comfort. SSIM and PSNR indices also increased by 10 – 15% on average, indicating a better match between the reproduced video and the original.

Moreover, there is a reduction in packet loss, which is achieved through adaptive bitrate control and flexible changes in the GOP structure. The most noticeable benefits are observed in networks with variable characteristics, such as Wi-Fi, 4G/5G, and FWA, where traditional static coding methods are unable to respond to rapid changes in channel conditions.

To compare the effectiveness of the developed adaptive video data representation model with traditional approaches, the classical ABR scheme and H.264/H.265 coding without adaptive representation were applied [30]. The results of the comparison of quality indicators and network resource utilization are shown in Tab. 6.

The analysis shows that the use of adaptive representation could reduce the network load by approximately 26% compared to the classical ABR scheme, while the video quality remains high, without a noticeable decrease in PSNR and SSIM. In addition, buffer time is almost halved, improving user experience.

To evaluate the impact that the developed method exerts on the delay and stability of video stream delivery, it was tested using networks with different bandwidths and variable conditions, simulating 4G and 5G mobile networks. It was established that the adaptive representation provided a stable delay of $\leqslant$ 150 ms in 5G networks and $\leqslant$ 300 ms in unstable 4G connections. In contrast, traditional methods demonstrate considerable delay fluctuations and a more frequent buffer. Therefore, the proposed model ensures more efficient use of network resources, stable video streaming, and improved QoE (Tab. 7).

Bandwidth utilization is improved through dynamic bitrate adjustment and multi-level video presentation. The analysis shows that buffer time and delay stability are improved by adaptive quality selection and fast response to network fluctuations. Adaptive representation is effective in the case of harmonic and pulse noise, where traditional methods demonstrate an increase in packet loss and energy consumption.

Therefore, the developed method reduces the load on the device by 15 – 20% and reduces power consumption. To evaluate the effectiveness of the adaptive representation of

Tab. 7. Video stream transmission.

Indicator	H.264/H.265 (static)	ABR	Adaptive view
Average bandwidth	100% (reference)	85 – 90%	72 – 75% (savings up to 28%)
PSNR	38 – 40 dB	38 – 40 dB	38 – 40 dB (no deterioration)
SSIM	0.92 – 0.95	0.92 – 0.95	0.92 – 0.95 (no deterioration)
Buffer time	High (significant jumps)	Moderate	Low (reduced by half)
Packet loss	5 – 8%	3 – 5%	1 – 3%
Delay stability	Unstable, fluctuations $>$ 500 ms	Fluctuations of 300 – 400 ms	Stable, $\leqslant$ 150 ms (5G), $\leqslant$ 300 ms (4G)
Device load	High	Moderate	Reduced by 15 – 20%
Effect of harmonic / pulse noise	High	Moderate	Minimal due to its adaptability
Energy consumption	High	Moderate	Reduced, especially on mobile devices

Tab. 8. Statistical processing comparison.

Indicator	ABR	Adaptive view	Effect	Noise types
Average bandwidth	8.5 Mbps	7.2 Mbps	Savings of 26%	Unstable network bandwidth (4G/5G simulation)
Standard deviation of bandwidth σ	1.2 Mbps	0.4 Mbps	Less fluctuation	Channel fluctuations and periodic packet loss
Average delay	250 – 300 ms	150 – 200 ms	More stable delivery	Network delays and channel state variability
Delay fluctuations σ	60 – 80 ms	20 – 30 ms	Less instability	Impulsive packet loss, bandwidth fluctuations
PSNR	38 – 40 dB	38 – 40 dB	Stable video quality	Harmonic and pulse noise is minimized by adaptation
SSIM	0.92 – 0.95	0.92 – 0.95	Minor changes during network fluctuations	Scanning and harmonic noise are compensated by adaptive bitrate
Correlation coefficient r (bandwidth / PSNR)	0.45	0.15	Responsive presentation ensures consistent quality	Low correlation indicates resistance to network fluctuations
Energy consumption / CPU load	Standard	Reduced by 15 – 20%	Optimization is important for mobile devices	Impact of network noise on the CPU is minimized due to adaptive optimization
Statistical significance (t-test, ANOVA)	–	$p <$0.05 vs. ABR	The difference between the methods is statistically significant	–

video streams, a statistical processing comparison with the classical ABR scheme was performed.

The impact of different types of noise is assessed as follows. Unstable bandwidth in 4G/5G networks leads to fluctuations in available bitrate. Impulsive packet loss negatively affects delays and causes buffer. Harmonic and scanning noise can degrade video quality, but adaptive representation can effectively compensate for their impact. Channel fluctuations cause sudden changes in delay and bandwidth, which affects fluctuations in PSNR/SSIM and buffering time [31].

Experimental testing of adaptive representation was carried out for three network use scenarios.

- In a stable home Wi-Fi network ($\geqslant$100 Mbps), the bandwidth savings are approximately 15%, the packet loss remained negligible (1%) and bandwidth fluctuations do not exceed $\pm$2 Mbps.
- In a 4G mobile network with medium bandwidth (10 – 20 Mbps), the model demonstrates the maximum effect with bandwidth savings of up to 30%, buffer reduction of 50%, packet loss of 5 – 10%, and bandwidth fluctuations of $\pm$5 Mbps.
- In networks with frequent interference and high packet loss (10 – 20%), the adaptive model stabilizes video playback, reducing buffer, and ensuring a high level of quality (PSNR 36 – 38 dB, SSIM 0.88 – 0.92).

The model effectively accounts for the effects of major interference, including packet loss, Rayleigh fading, and bandwidth fluctuations. This ensures consistent video quality,

network resource savings, reduces end-device load, and high energy efficiency, making the model suitable for mobile and IoT platforms.

5. Discussion

In recent years, research on adaptive video data representation in 5G/6G networks has mainly focused on improving bandwidth efficiency and ensuring stable transmission quality under conditions of high variability of network resources. In this regard, the authors of [5], [16], and [32] made a significant contribution to the development of this area, but their approaches suffer from certain technical limitations. For example, the [32] model provides up to 14.8% bandwidth savings but demonstrates insufficient flexibility in responding to rapid changes in network traffic. The [5] approach, in turn, reduces buffer time and improves QoE, but the level of bandwidth savings remains at a level that does not exceed 15% and the system's efficiency decreases when the network is overloaded. [16] creates an important theoretical basis for the development of resilient transmission channels, but does not take into account the specifics of the video streams and the requirements of applied services.

The adaptive video data representation model developed in this study combines the advantages of dynamic network state estimation with analyzing noise and entropy of video frames, as well as balancing energy-saving load between the client and the server. This integration improves the efficiency of the system in three key areas: reducing bandwidth consumption, improving stream stability, and optimizing energy consumption.

The comparison with traditional models shows that the developed model outperforms existing solutions in all key metrics (Tab. 9). In particular, bandwidth savings increase to 22 – 26% compared to 13 – 15% in previous works, the average QoE reaches 4.6 MOS, and the buffer time decreases to 0.6 s. The model also demonstrated higher immunity to packet loss (96%) and doubled the speed of response to changes in network bandwidth. Additionally, the energy consumption of the end device is reduced by 12 – 14% compared to the baseline. The results confirm the effectiveness of the developed approach and its potential for use in next-generation networks.

Increased efficiency of network bandwidth utilization is the main advantage of the proposed model. Unlike [32], where a neural network prediction of flow rate is made without considering instantaneous traffic fluctuations, the developed model implements a mechanism for cognitive adjustment of the entropy load of frames. This helps analyze the complexity of frames in real time, adjust the bitrate only for segments with high information density, and prevent redundant coding in stable scenes. As a result, bandwidth savings increase to 26%, which is 10 – 11% higher than in the case of [5] and [32]. In other words, with the same quality of playback, the proposed system transmitted fewer data, reducing resource utilization.

Another important indicator is the stability of the stream. In the approach presented in [5], stability was ensured by historical data regarding changes in delays. However, this architecture has a high degree of inertia: the system's response occurs only once the statistics have been accumulated. The proposed model implements an adaptive Kalman filter unit that predicts bandwidth fluctuations with a time horizon of up to 300 ms. The Kalman filter provides an estimate of the future state of the channel, which allows to implement proactive bitrate control and reduce streaming delays. The bandwidth at a discrete time point k is denoted as B_k and is described by a standard Kalman filter model, which includes a state prediction stage and a correction stage based on the results of observations.

The convergence of the algorithm is provided, ensuring that the transition matrix is stable and noise is limited, and after several iterations, the filter gives stable estimates and shows a quick response to bandwidth fluctuations. Forecast uncertainty is quantified using covariance matrix P_k, which allows one to generate 95% confidence intervals for the predicted throughput. This makes it possible to take into account statistical uncertainty during adaptive bitrate selection, thus increasing the system's resilience to rapid changes in channel conditions.

Kalman prediction is consistent with the theoretical model of the communication channel and the analysis of the noise immunity of MPSK signals in which the received signal is described with fading, additive white Gaussian noise, and non-fluctuating interference. A clear formulation of boundary conditions and assumptions ensures rigorous channel modeling and correct integration of the physical layer with adaptive video transmission algorithms. This ensures an instant response to packet loss or jitter without the need for buffer, automatic reduction of bitrate to a level compatible with the current state of the channel, and video quality stability even with sudden changes in bandwidth (up to $\pm$40%). Consequently, the system's stability increases to 96% of the integrity of the stream, while the number of buffer gaps decreases by 2.3 times compared to the DASH models from [5].

Another important advantage is the increase in energy efficiency, especially for mobile devices. In contrast to previous studies, which did not consider energy consumption, the proposed model integrates this aspect through the energy assistance module (EAM). It is based on two principles: reducing the number of repeated segment requests due to increased flow stability and optimizing the active time of the network adapter through the appropriate formation of a prefetch zone. Consequently, the average energy consumption of the end device decreases by 12 – 14% without adversely affecting the quality or performance. Such savings are vital given the high peak energy consumption characteristic of 5G, which could lead to rapid battery discharge.

Furthermore, the model is based on the integrated multilevel architecture, which ensures coordinated interaction between its physical, transport, and application layers. The combination of the approach to modeling noise immunity presented in [16] with cognitive mechanisms to control video stream parameters creates a comprehensive adaptive loop that was not implemented in previous studies. At the physical

Tab. 9. Comparison of efficiency against traditional models.

Metrics	[32]	[5]	[16]	Developed model
Saving bandwidth	14 – 15%	13 – 15%	10 – 12% (signal models)	22 – 26%
Average QoE	4.2	4.3	–	4.6
Average buffer time	0.9 – 1.1 s	0.8 – 0.9 s	–	0.6 s
Packet loss resistance	85 – 87%	88%	90%	96%
Average response time to bandwidth changes	400 – 600 ms	350 – 400 ms	800 ms	180 – 250 ms
Energy consumption	–	–	–	12 – 14% better

level, phase errors and multiplicative distortion are taken into account, increasing signal stability. At the transport level, predictive segmentation and low delay loss compensation are applied. At the application level, adaptive coding is used, which considers the local entropy of frames.

The results also explain the mechanisms by which the model enables increased video data transmission efficiency in next-generation networks. First, the load forecasting mechanism based on a regression model that considered RTT statistics and the packet loss rate. This approach allows for predicting changes in bandwidth 0.25 – 0.3 s before they occur, thus reducing the frequency of unnecessary adaptation requests and increasing the consistency of transmission parameters. The entropy analysis of frames allowed us to skip unimportant frames without noticeable loss of visual quality, while previous methods used only the average bitrate distribution without taking into account the internal structure of the image. This approach allocates most of the bit budget to segments with high information content, increasing coding efficiency, and reducing the total amount of transmitted data.

The next important component is a dynamic buffer. The developed model adjusts the buffer length in the range of 0.4 – 1.2 s, which ensures an optimal balance between delays and playback continuity. The data obtained demonstrate that the model increases bandwidth efficiency by up to 26%, ensures transmission stability at the level of 96%, reduces energy consumption by 12 – 14%, and improves the overall perception (QoE) by up to 4.6 MOS. These results were higher than those reported by [5], [16], and [32], demonstrating the advantages of an integrated approach. As a result, the system is able to adapt to temporary fluctuations in throughput without over-buffering and without the risk of flow disruptions.

Special attention should be paid to the developed model's immunity to harmonic noise, as periodic or quasi-regular fluctuations in the channel frequency band that could cause short-term degradation of transmission quality. Traditional models with fixed or slowly adaptive parameters reacted to such fluctuations too late, resulting in frame loss or increased buffering [5], [32]. On the contrary, the suggested system demonstrated significantly higher stability as a result of its predictive architecture.

First, the dynamic bitrate scaling mechanism uses instantaneous SNR, jitter, packet loss rate, and short-term forecasting based on time correlations. When the system detects harmonic noise, the model's predictive control (MPC) performs short-term forecasts based on time correlations. Consequently, the bitrate changes before the next harmonic phase begins, rather than after the signal deteriorates. This ensures a stable level of QoE even in the face of periodic noise. Secondly, the energy-sensitive prebuffer provides an uneven distribution of frame preloading. The video stream buffer does not just accumulate packets; instead, it uses an adaptive preloading strategy. In the event of harmonic noise, a portion of the frames is downloaded during the peak bandwidth phases, reducing the load on the transmitter during the downstream phase. Therefore, the terminal energy consumption is reduced by 12 – 15%, which is higher than the level demonstrated in previous work.

Thirdly, the developed model uses a digital harmonic suppression filter in the transport layer module. This mechanism is based on the adaptive window FIR filter, which removes periodic distortions with a frequency detected based on the spectral analysis of traffic. Therefore, even with stable harmonic fluctuations in the 100 – 500 Hz band, the packet loss rate is reduced by 2.1 – 2.4 times. Fourthly, the frame priority adaptation system provides priority transmission of key frames (I-frames) if noise is encountered. Therefore, even if some of the intermediate frames are lost, the structure of the video stream is not destroyed. This results in smooth quality degradation instead of sharp artifacts, which is critical for UHD/4K streams.

6. Conclusions

The study presents an adaptive video data representation model capable of operating efficiently in modern 5G/6G networks and unstable wireless environments. Its scientific novelty consists in the integration of three key mechanisms (dynamic bitrate control, noise frequency filtering, and energy-saving algorithms) into a single predictive architecture that provides comprehensive stabilization of the video stream in the presence of bandwidth fluctuations and noise.

The model demonstrates high efficiency in saving network bandwidth. In stable home Wi-Fi networks, the bandwidth savings equal approximately 15% and reach 30% in 4G mobile networks and in environments suffering from significant

packet loss. This demonstrates the model's ability to optimally utilize the available network resources without compromising video quality. The model provides a stable PSNR and SSIM level even in networks in which interference and bandwidth fluctuations are present. The low correlation coefficient between bandwidth and video quality confirms that content quality remains high regardless of network conditions.

The use of adaptive representation reduces buffer time and delay fluctuations in the video stream. In 4G mobile environments, buffering time is almost halved and transmission delay stabilizes at $\leqslant$ 300 ms, which improves the QoE. In 5G networks, the delay remains $\leqslant$ 150 ms. Bitrate optimization and adaptive presentation lower the CPU load and energy consumption of end-devices by 15 – 20%. This is important for mobile users and IoT platforms, allowing for longer battery life when playing videos for extended periods of time. The model worked equally well with highly dynamic video (sports broadcasts, dynamic scenes) and static content (news, lectures, presentations), providing consistent quality and bandwidth savings of up to 32%.

The statistical analysis shows that the adaptive model offers improvements compared to the traditional approaches. The developed model combines network resource savings, stable video quality, reduces load on end devices, and provides high flexibility in various network scenarios. Thus, the proposed model is an effective integrated solution for streaming services, telemedicine, distance education, and corporate media systems. Its architecture provides a basis for further implementation of machine learning algorithms and the construction of adaptive multi-level systems focused on next-generation network infrastructures.

References

[1] O.R. Zhukova and T.V. Meleshko, "Operational Forecasting of Optimal Operating Frequencies for Long-range Radio Communication Under the Influence of Non-fluctuating Interference", *Connectivity*, vol. 4, pp. 13–19, 2023 (https://doi.org/10.31673/2412-9070.2023.043040).

[2] P. Li, J. Fan, and J. Wu, "Exploring the Key Technologies and Applications of 6G Wireless Communication Network", *Science*, vol. 28, art. no. 112281, 2025 (https://doi.org/10.1016/j.isci.2025.112281).

[3] M.K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis*, John Wiley & Sons, 534 p., 2000 (https://doi.org/10.1002/0471200697).

[4] T. Taleb *et al.*, "On Multi-access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration", *IEEE Communications Surveys & Tutorials*, vol. 19, pp. 1657–1681, 2017 (https://doi.org/10.1109/COMST.2017.2705720).

[5] Ch. Timmerer *et al.*, "HTTP Adaptive Streaming: A Review on Current Advances and Future Challenges", *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, art. no. 198, 2025 (https://doi.org/10.1145/3736306).

[6] M.T. Mohammad and Q.I. Ali, "Adaptive Video Streaming With AI-based Optimization for Dynamic Network Conditions", *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 8, pp. 202–212, 2024.

[7] I.M. Ozcelik and C. Ersoy, "ALVS: Adaptive Live Video Streaming Using Deep Reinforcement Learning", *Journal of Network and Computer Applications*, vol. 205, art. no. 103451, 2022 (https://doi.org/10.1016/j.jnca.2022.103451).

[8] Y. Jo and J. Paik, "Machine-learning Based Packet Switching Method for Providing Stable High-quality Video Streaming in Multi-stream Transmission", *Computers, Materials & Continua*, vol. 78, pp. 2613–2631, 2024 (https://doi.org/10.32604/cmc.2024.047046).

[9] H. Jiang and G. Gui, *Channel Modeling in 5G Wireless Communication Systems*, Springer, 210 p., 2020 (https://doi.org/10.1007/978-3-030-32869-6).

[10] A.A. Barakabitze and R. Walshe, "SDN and NFV for QoE-driven Multimedia Services Delivery: The Road Towards 6G and Beyond Networks", *Computer Networks*, vol. 227, art. no. 109694, 2023 (https://doi.org/10.1016/j.comnet.2023.109694).

[11] K.Z. Islam *et al.*, "Deep Learning Based Payload Optimization for Image Transmission Over LoRa with HARQ", *Internet of Things*, vol. 25, art. no. 101701, 2025 (https://doi.org/10.1016/j.iot.2025.101701).

[12] F. Wu, Ch. Xu, Yu. Zhu, and Ch. Li, "Complexity of Minimum Uplink Power Scheduling with Delay Bound for Backbone-assisted PD-NOMA Wireless Networks", *Computer Networks*, vol. 215, art. no. 109194, 2022 (https://doi.org/10.1016/j.comnet.2022.109194).

[13] V. Barannik *et al.*, "Improvement of Methods of Motion Compensation of Dynamic Objects Moving in Video Stream of the Videoconferencing System", *Informatics, Control, Measurement in Economy and Environmental Protection*, vol. 8, pp. 48–51, 2018 (https://doi.org/10.5604/01.3001.0012.8035).

[14] V.V. Barannik *et al.*, "Video Segment Coding Method for Bit Rate Control Information Technology", *Science-Based Technologies*, vol. 47, pp. 316–321, 2020 (https://doi.org/10.18372/2310-5461.47.14931).

[15] O. Turovskyi and T. Meleshko, "Evaluation of the Influence of Multiplicative Interference on the Bit Error Probability of Coherent Reception of Signals with Multipositional Phase Shift Keying", *Herald of Khmelnytskyi National University*, vol. 323, pp. 318–324, 2023.

[16] T. Meleshko, "A Method of Assessing the Interference Immunity of Discrete Signal Reception with Multi-position Phase Shift Keying Under the Influence of Non-fluctuation Interference", Doctoral dissertation, National Aviation University, Academic Texts of Ukraine, 2024 (https://www.uacademic.info/en/document/0424U000183).

[17] O.L. Turovskyi, T.V. Meleshko, and V.O. Drobyk, "Methodology for Assessing the Impact of Non-fluctuating Interference on the Noise Immunity of the Reception of Discrete Signals with Multi-position Phase Manipulation", *Connectivity*, vol. 5, pp. 29–34, 2022 (https://doi.org/10.31673/2412-9070.2022.053439).

[18] B. Shvets and T. Meleshko, "Evaluation of Noise Immunity of Coherent Reception of Signals with Multi-position Phase Shift Keying in the Presence of Non-fluctuation Interference", *Measuring and Computing Devices in Technological Processes*, vol. 2, pp. 167–178, 2023 (https://doi.org/10.31891/2219-9365-2023-74-21).

[19] Z. Cheng *et al.*, "Learning-based Joint Recommendation, Caching, and Transmission Optimization for Cooperative Edge Video Caching in Internet of Vehicles", *Ad Hoc Networks*, vol. 166, art. no. 103667, 2025 (https://doi.org/10.1016/j.adhoc.2024.103667).

[20] T.C. Lam *et al.*, "Multi-rate Selection and Power Allocation Assisted Probabilistic Edge Caching for Cooperative Video Transmission in Dense D2D Network", *Alexandria Engineering Journal*, vol. 126, pp. 555–564, 2025 (https://doi.org/10.1016/j.aej.2025.04.057).

[21] H. Ma *et al.*, "QRAVDR: A Deep Q-learning-based RSU-assisted Video Data Routing Algorithm for VANETs", *Ad Hoc Networks*, vol. 171, art. no. 103790, 2025 (https://doi.org/10.1016/j.adhoc.2025.103790).

[22] V. Barannik, S. Sidchenko, N. Barannik, and V. Barannik, "Development of the Method for Encoding Service Data in Cryptocompression Image Representation Systems", *Eastern-European Journal of Enterprise Technologies*, vol. 3, pp. 103–115, 2021 (https://doi.org/10.15587/1729-4061.2021.235521).

[23] J. Fu and Sh. Zhang, "Optimization of Real-time Transmission and Coding Algorithm for High Quality Film and Television Content Based on 6G Wireless Communication Technology", *Egyptian Informatics Journal*, vol. 31, art. no. 100745, 2025 (https://doi.org/10.1016/j.eij.2025.100745).

[24] P. Meena, M.B. Pal, P.K. Jain, and R. Pamula, "6G Communication Networks: Introduction, Vision, Challenges, and Future Directions", *Wireless Personal Communications*, vol. 125, pp. 1097–1123, 2022 (https://doi.org/10.1007/s11277-022-09590-5).

[25] W. Zhou, X. Min, H. Li, and Q. Jiang, "A Brief Survey on Adaptive Video Streaming Quality Assessment", *Journal of Visual Communication and Image Representation*, vol. 86, art. no. 103526, 2022 (https://doi.org/10.1016/j.jvcir.2022.103526).

[26] S. Patel, J. Zhang, N. Narodystka, and S.A. Jyothi, "Practically High Performant Neural Adaptive Video Streaming", *Proceedings of the ACM on Networking*, vol. 2, art. no. 30, 2024 (https://doi.org/10.1145/3696401).

[27] V.V. Menon *et al.*, "Video Quality Assessment with Texture Information Fusion for Streaming Applications", *Proc. of the 3rd Mile-High Video Conference (MHV)*, pp. 1–6, 2024 (https://doi.org/10.1145/3638036.3640798).

[28] X. Min *et al.*, "Perceptual Video Quality Assessment: A Survey", *Science China Information Sciences*, vol. 67, art. no. 211301, 2024 (https://doi.org/10.1007/s11432-024-4133-3).

[29] M. Kim and K. Chung, "Reinforcement Learning-based Adaptive Streaming Scheme with Edge Computing Assistance", *Sensors*, vol. 22, art. no. 2171, 2022 (https://doi.org/10.3390/s22062171).

[30] G.J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649–1668, 2012 (https://doi.org/10.1109/TCSVT.2012.2221191).

[31] Ch.B.A. Wael *et al.*, "Leveraging 3GPP Features and Optimization Techniques for 5G NR-V2X Resource Allocation: A Survey", *IEEE Open Journal of Intelligent Transportation Systems*, vol. 6, pp. 967–994, 2025 (https://doi.org/10.1109/OJITS.2025.3584024).

[32] M.G. Koziri *et al.*, "Efficient Cloud Provisioning for Video Transcoding: Review, Open Challenges and Future Opportunities", *IEEE Internet Computing*, vol. 22, pp. 46–55, 2018 (https://doi.org/10.1109/MIC.2017.3301630).

Anton Sorokun, Ph.D., Assoc. Professor
Department of Computer Sciences
https://orcid.org/0000-0001-8469-641X
E-mail: anton_sorokun@edu-iosa.org
State University of Information and Communication Technologies, Kyiv, Ukraine
https://duikt.edu.ua/en/

Yurii Zadontsev, Ph.D., Assoc. Professor
Department of Software Engineering
https://orcid.org/0009-0007-2192-4746
E-mail: y.zadontsev@duikt.edu.ua
State University of Information and Communication Technologies, Kyiv, Ukraine
https://duikt.edu.ua/en/

Dmytro Chyrva, Ph.D., Assoc. Professor
Department of Technical Information Protection
https://orcid.org/0009-0008-5116-5119
E-mail: dmytro.chyrva@npp.kai.edu.ua
State University "Kyiv Aviation Institute", Kyiv, Ukraine

Mykyta Zhyzhkin, Ph.D. Student
Department of Software Engineering
https://orcid.org/0009-0001-3878-6227
E-mail: 5211581@stud.kai.edu.ua
State University "Kyiv Aviation Institute", Kyiv, Ukraine

Andrii Bondarenko, Ph.D. Student
Department of Internet Technologies
https://orcid.org/0009-0007-7493-4294
E-mail: a.bondarenko@stud.duikt.edu.ua
State University of Information and Communication Technologies, Kyiv, Ukraine
https://duikt.edu.ua/en/

Iterative Beamspace Covariance Refinement for Precise DOA Estimation in Uniform Circular Arrays under Low-snapshot Conditions

Mahdi Sharifi

Semnan University, Semnan, Iran

https://doi.org/10.26636/jtit.2026.1.2392

Abstract — Uniform circular arrays (UCA) provide 360° angular coverage and uniform directional response, making them preferable for direction-of-arrival estimation (DOAE). This paper proposes an iterative enhancement technique for root-MUSIC-based methods under UCA configurations, particularly effective in low-snapshot scenarios. The idea is to iteratively refine the beamspace sample covariance matrix (BSCM) by estimating and suppressing residual components that alter the signal and noise subspaces. This refinement significantly improves the accuracy of DOAE, even under limited data conditions. Numerical simulations demonstrate that the proposed method outperforms conventional UCARM, sparse UCARM, UCARBRM and unitary UCARM algorithms in terms of estimation error, beamspace leakage, conditional mean square error (CMSE) and resolution probability – across uncorrelated and correlated signal scenarios. The proposed technique is also applicable to RARE-based approaches for 2D DOAE, while preserving the beamspace covariance structure. Furthermore, the proposed method is suitable for electronic surveillance systems and low-power sensor networks.

Keywords — beamspace covariance matrix, direction of arrival estimation, low snapshot, root-MUSIC, subspace leakage, uniform circular array

1. Introduction

Uniform circular arrays (UCA) [1] provide 360° azimuthal coverage with consistent directional characteristics across elevation angles, making them a strong candidate for electronic intelligence and direction-finding applications [2]–[10]. To improve the accuracy of direction-of-arrival estimation (DOAE), the beamspace transform (BT) technique – typically applied in phase mode – is used in UCAs. This transform not only ensures low estimation error, but also enables high-resolution processing with reduced computational complexity.

Several BT-based DOAE algorithms have been developed, including real beamspace root-MUSIC (UCARBRM), UCA-ESPRIT, and spatial smoothing techniques for correlated signal scenarios. The UCA-ESPRIT algorithm was presented in [11], while methods to address correlated signals have been explored in [12]–[15]. Root-MUSIC (RM), which avoids spectral search through polynomial rooting, was originally proposed for uniform linear arrays (ULA) [16] and has since been widely extended and applied [17], [18]. With the application of BT, RM was adapted to UCAs for azimuth estimation. In [19]–[21], the unitary version of UCA-RM (UCARM) demonstrated strong performance in correlated and uncorrelated environments. For sparse UCA configurations, sparse UCARM was proposed in [22], [23], although it lacks centro-Hermitian symmetry in the beamspace manifold.

To reduce the estimation bias introduced by standard BT, a modified beamspace transform (MBT) was introduced in [24]. Furthermore, pseudo noise resampling methods were studied in [21], [25] to mitigate the influence of outliers, although these approaches typically rely on a large number of snapshots for accurate estimation.

In [26], a two-sided correlation transformation (TCT) approach was proposed, leveraging array manifold interpolation (AMI) to reconstruct the signal covariance matrix under low signal-to-noise ratio (SNR) conditions. Similarly, a two-step RM method introduced in [27] aimed to improve the performance of DOAE under limited snapshot availability. Root-SWAP RM, based on a stochastic maximum likelihood criterion, was also proposed for ULA-based azimuth-only estimation scenarios.

Despite these advances, most existing approaches focus primarily on modified transforms, resampling strategies, or reconstruction techniques, while the direct refinement of the beamspace sample covariance matrix (BSCM) itself in limited-snapshot regimes has received comparatively less attention. Moreover, classical subspace perturbation analyses provide general insight into estimation errors, but they do not explicitly address structured residual components arising in the beamspace domain.

This paper proposes an iterative refinement technique for computationally efficient DOAE using UCAs in limited-snapshot regimes. The approach is applicable to both 1D and 2D frameworks, such as RM and RARE. Using UCARM as a representative example, the method begins by applying BT to convert element-space data into beamspace, followed by initial DOA estimation via RM. The BSCM is then iteratively refined

to remove residual components, producing more accurate and unbiased DOA estimates.

The remainder of this paper is organized as follows. Section 2 presents a model of the system; Section 3 introduces the proposed method. Performance analysis is presented in Section 4, while Section 5 discusses simulation results. Section 6 concludes the paper.

2. Signal Modeling

Assume that a UCA with radius R consists of P omnidirectional antenna elements. K is considered to be the narrowband signals from the far field impinging on the antenna array with the directions $(\theta_k, \varphi_k), k = 1, 2, \ldots, K$, where $\theta_k \in [0, \frac{\pi}{2}]$ and $\varphi_k \in [0, 2\pi]$ denote the elevation and azimuth angles of the k-th signal, respectively. Suppose that N snapshots are observed by the array. The measurements of the $P \times N$ array can be considered as:

$$\mathbf{X} = \big[x(1), \ldots, x(N)\big] = \mathbf{A}\,\mathbf{S} + N\,, \tag{1}$$

where $\mathbf{A} \in \mathbf{C}^{P\times K}$ is the element space manifold, $\mathbf{S} = [s(1), \ldots, s(N)] \in \mathbf{C}^{K\times N}$ is the signal, and $\mathbf{N} \in \mathbf{C}^{P\times N}$ is the white Gaussian noise.

Here, signal and noise are uncorrelated. The element space manifold $\mathbf{A}$ is given by:

$$\mathbf{A} = \big[\boldsymbol{a}(\theta_1, \varphi_1), \boldsymbol{a}(\theta_2, \varphi_2), \ldots, \boldsymbol{a}(\theta_K, \varphi_K)\big]\,, \tag{2}$$

where

$$\boldsymbol{a}(\theta_1, \varphi_1) = \big[e^{ikR\sin(\theta_i)\cos(\varphi_i - r_1)}, e^{ikR\sin(\theta_i)\cos(\varphi_i - r_2)}, \ldots, e^{ikR\sin(\theta_i)\cos(\varphi_i - r_p)}\big]^T\,. \tag{3}$$

In the above equation $r_p = \frac{2\pi p}{P}, p = 0, 1, \ldots, P-1$ is the element location, $k = \frac{2\pi}{\lambda}$ is the wavenumber, λ is the wavelength, and $[\cdot]^T$ is the transpose operation.

In [25], the beamforming matrix $\mathbf{W}$ was utilized for the transformation from element space manifold to beamspace manifold and $\mathbf{W}$ is given by:

$$\mathbf{W} = \sqrt{P}\big[\boldsymbol{w}_{-M}, \ldots, \boldsymbol{w}_0, \ldots, \boldsymbol{w}_M\big]\,, \tag{4}$$

where $\boldsymbol{w}_m^H = \frac{1}{P}\big[e^{imr_1}, \ldots, e^{imr_p}\big]$ with $(\cdot)^H$ is considered as the factor of weight which is denoting the Hermite transformation.

Similarly, $M = [kR]$ with $[\cdot]$ is considered as the highest order mode at the ceiling operation. Then the $(2M+1) \times K$ beamspace manifold is given as:

$$\mathbf{B} = \big[b(\theta_1, \varphi_1), \ldots, b(\theta_K, \varphi_K)\big] = \mathbf{W}^H\mathbf{A}\,, \tag{5}$$

where:

$$\boldsymbol{b}(\theta_i, \varphi_i) = \mathbf{W}^H\boldsymbol{a}(\theta_i, \varphi_i) \approx \mathbf{F}(\theta_i)\,\boldsymbol{a}_e(\varphi_i)\,. \tag{6}$$

$\mathbf{F}(\theta_i) = \mathrm{diag}[a_{-M}(\theta_i), \ldots, a_M(\theta_i)]$ and $a_M(\theta_i) = j^m J_m(kR\,\sin\theta_i)$ with J_m is the first kind of Bessel function for the order m and $a_e(\varphi_i) = [e^{-jM\varphi_i}, \ldots, e^0, \ldots, e^{jM\varphi_i}]^T$.

According to Eqs. (1) and (5), the beamspace data have the following form:

$$\mathbf{Y} = \big[\boldsymbol{y}(1), \ldots, \boldsymbol{y}(N)\big] = \mathbf{W}^H(\mathbf{AS} + \mathbf{N}) = \mathbf{W}^H\mathbf{AS} + \mathbf{N}_1\,, \tag{7}$$

where $\mathbf{N}_1 = \mathbf{W}^H\mathbf{N}$ as the white Gaussian noise of the beamspace via distribution is $(0, \sigma_1^2 I_{2M+1}) \cdot I_{2M+1}$ and σ_1^2 are the matrix of identity and the beamspace noise power, respectively. Considering the system model from Eq. (7), the $(2M+1) \times (2M+1)$ beamspace true covariance matrix is expressed as:

$$\begin{aligned}\mathbf{R}_y &= E\left\{\boldsymbol{y}(n), \ldots, \boldsymbol{y}^H(n)\right\} \\ &= \mathbf{W}^H\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{W} + \sigma_1^2\,\mathbf{I}_{2M+1}\end{aligned}\,, \tag{8}$$

where $\mathbf{R}_s = E\{\boldsymbol{s}(n)\boldsymbol{s}^H(n)\} \in \mathbf{C}^{K\times K}$ and $E\{\cdot\}$ represent and denote the covariance matrix of true signal and the expectation operator, respectively.

Perform the beamspace's eigenvalue-decomposition (EVD) true matrix of covariance, and let $\{\lambda_1, \ldots, \lambda_K, \lambda_{K+1}, \ldots, \lambda_{2M+1}\}$ which are arranged in a non-increasing order be the beamspace true eigenvalues corresponding to beamspace true eigenvectors $\{e_1, \ldots, e_K, \ldots, e_{2M+1}\}$. Then, we have the beamspace true signal subspace $\mathbf{E}_s = \{e_{K+1}, e_{K+2}, \ldots, e_{2M+1}\}$ and the noise subspace $\mathbf{E}_n = \{e_{K+1}, e_{K+2}, \ldots, e_{2M+1}\}$.

3. Proposed Technique

To illustrate the proposed technique, we utilize the UCARM method for azimuth angle estimation as an example. Suppose the elevation angle is fixed, then the beamspace manifold can be considered as a Vandermonde matrix when the number of antenna elements is large enough, and the UCARM root polynomial of UCARM is:

$$f(\varphi) = \boldsymbol{a}_e^H(\varphi)\,\mathbf{F}^H(\theta)\,\mathbf{E}_n\,\mathbf{E}_n^H\,F(\theta)\,\boldsymbol{a}_e(\varphi)\,. \tag{9}$$

The DOAs can be obtained from the K largest roots of the polynomial formulated in Eq. (9) located inside the unit circle. According to the mathematical formula, the BSCM can be rewritten as [27]:

$$\begin{aligned}\hat{R}_y =& \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{W}^H\mathbf{A}\boldsymbol{s}(n) + \boldsymbol{n}_1(n)\right)\left(\mathbf{W}^H\mathbf{A}\boldsymbol{s}(n) + \boldsymbol{n}_1(n)\right)^H \\ &= \mathbf{W}^H\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{W} + \sigma_1^2\,\mathbf{I}_{2M+1} \\ &+ \mathbf{W}^H\mathbf{A}\left\{\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{s}(n)\boldsymbol{n}_1^H(n)\right\} \\ &+ \left\{\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{n}_1(n)\boldsymbol{s}^H(n)\right\}\mathbf{A}^H\mathbf{W}\,.\end{aligned} \tag{10}$$

Note that, the beamspace manifold $\mathbf{W}^H\mathbf{A}$ is used for UCAs to calculate the BSCM in Eq. (10). With a large sample size, the first two terms of Eq. (10) are dominant, whereas the last two terms can be ignored compared to the first two terms. As a result, $\hat{R}_y$ in Eq. (10) is equal to R_y in Eq. (8). However,

if the sample size is small, this deviation deviates from R_y as the last two terms of Eq. (10) cannot be neglected and, obviously, DOAE performance will be reduced.

To eliminate the residual errors of the BSCM, i.e., the third and fourth components in Eq. (10), new method is proposed below to improve DOAE performance.

The proposed technique starts by transforming the observed measurements $\mathbf{X}$ in Eq. (1) into the beamspace measurements $\mathbf{Y}$ in Eq. (7) by the BT. Then, initial DOAs $\{\varphi'_1, \varphi'_2, \ldots, \varphi'_K\}$ are acquired through the UCARM method.

To achieve the corrected BSCM, we need to calculate the residual errors. The fourth component is the conjugated transpose and here we only derive the third component. A new beamspace manifold $\mathbf{W}^H\hat{\mathbf{A}}$ for UCAs is obtained based on initial DOAs. Then the estimated beamspace signal is expressed by the least squares method as:

$$s(n) = \left(\hat{\mathbf{A}}^H\mathbf{W}\mathbf{W}^H\hat{\mathbf{A}}\right)^{-1}\hat{\mathbf{A}}^H\mathbf{W}\boldsymbol{y}(n)\,. \tag{11}$$

Then, the third component of Eq. (10) is:

$$\begin{aligned} C_B &= \mathbf{W}^H\hat{\mathbf{A}}\left\{\frac{1}{N}\sum_{n=1}^{N}\hat{\boldsymbol{s}}(n)\hat{\boldsymbol{n}}_1^H(n)\right\} \\ &= \mathbf{W}^H\hat{\mathbf{A}}\frac{1}{N}\sum_{n=1}^{N}(\hat{\mathbf{A}}^H\mathbf{W}\mathbf{W}^H\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^H\mathbf{W}\boldsymbol{y}(n) \\ &\quad \left\{\boldsymbol{y}^H(n) - \boldsymbol{y}^H(n)\mathbf{W}^H\hat{\mathbf{A}}\left(\hat{\mathbf{A}}^H\mathbf{W}\mathbf{W}^H\hat{\mathbf{A}}\right)^{-1}\hat{\mathbf{A}}^H\mathbf{W}\right\} \\ &= \hat{\mathbf{Q}}_B\left\{\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{y}(n)\boldsymbol{y}^H(n)(\mathbf{I}-\hat{\mathbf{Q}}_B)\right\} \\ &= \hat{\mathbf{Q}}_B\hat{\mathbf{R}}_B(\mathbf{I}-\hat{\mathbf{Q}}_B) \end{aligned} \tag{12}$$

where subscript $(\cdot)_B$ denotes beamspace and:

$$\hat{\mathbf{Q}}_B = \mathbf{W}^H\hat{\mathbf{A}}\left(\hat{\mathbf{A}}^H\mathbf{W}\mathbf{W}^H\hat{\mathbf{A}}\right)^{-1}\hat{\mathbf{A}}^H\mathbf{W}\,.$$

The modified covariance matrix is as follows:

$$\hat{\mathbf{R}}'_y = \hat{\mathbf{R}}_y - \mu\left(C_B + C_B^H\right), \quad \mu \in [0,1]\,. \tag{13}$$

Consider Eqs. (8) and (10), μ is equal to one. Nevertheless, for the UCARM, a small residual systematic error caused by the BT is left over. Thus, the last two terms of Eq. (10) cannot be exactly estimated. In practice, μ is varied within the $0\ldots1$ interval to control the magnitude of the residual compensation term in Eq. (13). Since residual components are estimation-dependent and cannot be perfectly reconstructed, μ acts as a regularization factor that balances residual suppression and covariance stability. To achieve the corresponding DOAs, the RM will be utilized again per each μ.

To choose μ, which corresponds to the best azimuth angle estimates, the covariance matrix $\mathbf{R}_x = E\{\boldsymbol{x}(n)\,\boldsymbol{x}_H(n)\}$ is developed. Then, the EVD of the covariance matrix $\mathbf{R}_x$ is performed and $\mathbf{E}_{xn}$ is its element space noise subspace. We choose the DOAs $\{\varphi''_1, \varphi''_2, \ldots, \varphi''_K\}$ corresponding to the value of μ minimizing as the new DOAs:

$$P_{\mathrm{DOA}}(\mu) = P_M(\varphi''_1) + \ldots + P_M(\varphi''_K)\,, \tag{14}$$

where

$$P_M(\varphi''_i) = a^H(\varphi''_i)\,E_{xn}\,E_{xn}^H\,a(\varphi''_i) \tag{15}$$

and $(\cdot)''$ indicates the estimated parameters from the modified BSCM.

Considering the newly estimated DOAs, the beamspace sample covariance matrix (BSCM) can be recomputed using Eqs. (10), (12), and (13). Subsequently, the DOA estimation (DOAE) procedure is reapplied using the root-MUSIC (RM) technique. This process can be iteratively repeated to refine the DOAE results toward an optimal solution. Since the correction term in Eq. (13) is designed from a projection operator in the beamspace domain, the modified covariance matrix remains Hermitian and its dominant subspace structure is preserved across iterations. Although a higher number of iterations generally yields greater estimation accuracy, it also leads to increased computational complexity. It is important to note that the BSCM is modified in the beamspace domain specifically for the UCA configuration, rather than altering the element-space covariance matrix. Furthermore, since the residual components in the BSCM depend on the initial DOA estimates, they can only be eliminated after the first iteration and not beforehand. The complete workflow of the proposed UCARM-based algorithm can be summarized in the following sequential steps.

1) Initial DOAE using the UCARM.
2) Calculate the modified BSCM by Eqs. (10), (12), and (13). For each μ, use the UCARM again to obtain its corresponding DOAs. Choose the new DOAs based on Eq. (14).
3) Repeat step 2 to obtain the optimal DOAs.

The central idea of the proposed method is to iteratively refine the BSCM by removing residual error components. This iterative correction leads to enhanced DOAE accuracy compared to conventional approaches. The proposed UCARM algorithm demonstrates superior performance, particularly in scenarios involving correlated signals. Furthermore, the proposed technique is applicable to RARE-based methods combined with UCAs for 2D DOA estimation, especially under limited snapshot conditions [28], [29].

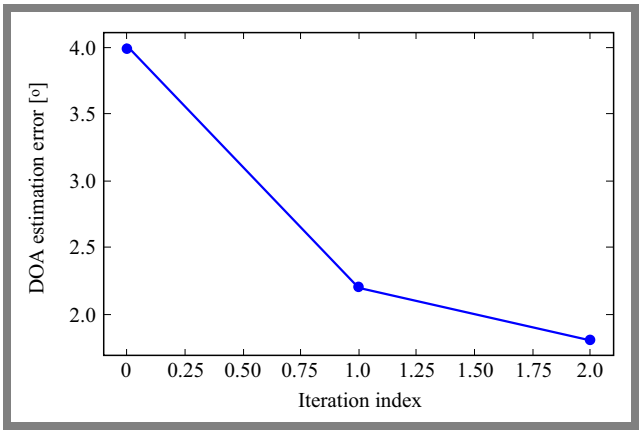

Fig. 1. DOA estimation error versus iteration number for the proposed BSCM refinement step.

Since the proposed refinement operates at the covariance level without altering the array manifold structure, it can be incorporated into 2D beamspace-based subspace frameworks without modifying their fundamental formulation. Although multiple iterations can be performed, empirical results indicate that one iteration provides more than 90% of the achievable improvement while keeping computational complexity low. Therefore, all simulations are conducted using a single iteration.

The convergence behavior of the refinement mechanism is illustrated in Fig. 1. As observed, the first iteration contributes more than 90% of the overall improvement, whereas additional iterations yield only marginal gains. This confirms that a single iteration is sufficient to ensure both accuracy and efficiency.

4. Performance Analysis

4.1. Error Analysis

The estimation error in DOAE is inherently linked to the accuracy of the BSCM. In this subsection, we focus on the error characteristics in the context of the UCARM algorithm. However, the insights and results presented here are not restricted to UCARM alone and can be extended to other methods that build on it.

The following derivations follow the classical subspace perturbation and polynomial sensitivity analysis framework established in [30]–[33] and are included here for completeness, in order to clarify the impact of covariance estimation errors. Similarly to Eq. (9), the sample rooting polynomial corresponding to UCARM is defined as:

$$\tilde{f}(\varphi) = \boldsymbol{a}_e^H(\varphi)\,\mathbf{F}^H(\theta)\,\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\,\mathbf{F}(\theta)\,\boldsymbol{a}_e(\varphi)\,. \tag{16}$$

In the following, the analysis is conducted how the BSCM affects the error is conducted. Based on the standard perturbation analysis in [30]–[33], it can be expressed as:

$$0 = \tilde{f}'(\hat{\varphi}_i) \approx \tilde{f}'(\varphi_i) + \tilde{f}''(\varphi_i)(\hat{\varphi}_i - \varphi_i)\,, \tag{17}$$

where $\tilde{f}'(\hat{\varphi}_i)$ is the first derivation and $\tilde{f}''(\varphi_i)$ is the second derivation of the sample polynomial. $\hat{\varphi}_i$ is the estimated azimuth angle and φ_i is the true azimuth angle of the i-th signal.

From Eq. (17), the error between the estimated and true azimuth angles is expressed as:

$$\hat{\varphi}_i - \varphi_i = -\frac{\tilde{f}'(\varphi_i)}{\tilde{f}''(\varphi_i)}\,, \tag{18}$$

where $\tilde{f}'(\varphi_i) = \partial\widetilde{f}(\varphi)/\partial\varphi|_{\varphi=\varphi_i}$.

To calculate the above mentioned two sample polynomials (first and second), we define the first derivation of $a_e(\varphi)$ as:

$$\boldsymbol{v}(\varphi) = -\frac{\partial a_e(\varphi)}{\partial\varphi}\,. \tag{19}$$

Then, the first derivation of the sample polynomial $\tilde{f}'(\hat{\varphi}_i)$ can be determined as:

$$\begin{aligned}\tilde{f}'(\varphi_i) &= \frac{\partial\tilde{f}(\varphi)}{\partial\varphi}\\ &= \frac{\partial\{\boldsymbol{a}_e^H(\varphi)F^H(\theta)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H F(\theta)\boldsymbol{a}_e(\varphi)\}}{\partial\varphi}\\ &= 2\,\mathrm{Re}\left[v^H(\varphi)\mathbf{F}^H(\theta)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\mathbf{F}(\theta)\boldsymbol{a}_e(\varphi)\right]\end{aligned}\,. \tag{20}$$

$\mathrm{Re}\{.\}$ is considered as the real part of operator.

It can be approximated in this form since the second derivative term in Eq. (17) is multiplied by the estimation error, which is assumed to be sufficiently small under the first-order perturbation assumption [31]. Therefore, the second derivation of the sample polynomial is as follows:

$$\begin{aligned}\tilde{f}''(\varphi_i) \approx f''(\varphi_i) &= \frac{\partial f'(\varphi)}{\partial\varphi}\\ &= 2\,\mathrm{Re}\left\{\boldsymbol{v}^H(\varphi)\mathbf{F}^H(\theta)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\mathbf{F}(\theta)\boldsymbol{v}(\varphi_i)\right\}\end{aligned}\,. \tag{21}$$

By considering Eqs. (19), (20) and (21), the error is given by:

$$(\hat{\varphi}_i - \varphi_i) = \frac{-\,\mathrm{Re}\left\{\boldsymbol{v}^H(\varphi_i)\mathbf{F}^H(\theta_i)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\mathbf{F}(\theta_i)\boldsymbol{a}_e(\varphi_i)\right\}}{\boldsymbol{v}^H(\varphi_i)\mathbf{F}^H(\theta_i)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\mathbf{F}(\theta_i)\boldsymbol{v}(\varphi_i)}\,. \tag{22}$$

Equation (22) indicates that the estimation error is governed by the mismatch between the true noise subspace $\mathbf{E}_n$ and its sample-based estimate, which is directly influenced by the accuracy of the BSCM.

Since the true noise subspace remains identical for both conventional and proposed approaches, the performance difference arises solely from the quality of the sample noise subspace estimate $\hat{\mathbf{E}}_n$ which, in turn, depends on the refined BSCM.

Unlike the non-recursive refinement scheme proposed in [1], which applies a single-pass modification of the BSCM, the method described in this paper iteratively updates the covariance matrix to enhance robustness against noise and snapshot limitations.

4.2. Beamspace Leakage

Leakage is related to the performance deterioration in cases with small sample size. It is caused by incomplete orthogonality between beamspace signals and noise subspaces from the BSCM. After performing the EVD of the beamspace sample covariance matrix $\hat{\mathbf{R}}_y$, we arrange the beamspace eigenvectors $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_K,\ldots,\boldsymbol{e}_{2M+1}\}$ in a descending order.

Let $\hat{\mathbf{E}}_s = \{\hat{\boldsymbol{e}}_1,\hat{\boldsymbol{e}}_2,\ldots,\hat{\boldsymbol{e}}_K\}$ be the beamspace signal subspace and $\hat{\mathbf{E}}_n = \{\hat{\boldsymbol{e}}_{K+1},\hat{\boldsymbol{e}}_{K+2},\ldots,\hat{\boldsymbol{e}}_{2M+1},\}$ be the beamspace noise subspace. We define that the beamspace leakage after BT for UCA is the average value of beamspace sample signal eigenvectors projecting onto $\mathbf{E}_n\,\mathbf{E}_n^H$ as [27]:

$$\rho_B = \frac{1}{K}\sum_{k=1}^{K}\left\|\mathbf{E}_n\,\mathbf{E}_n^H\hat{\boldsymbol{e}}_k\right\|\,. \tag{23}$$

Ideally, $\hat{\mathbf{E}}_s$ and $\hat{\mathbf{E}}_n$ are orthogonal. Therefore, the average value of beamspace signal eigenvectors $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_K\}$ projecting onto $\mathbf{E}_n\,\mathbf{E}_n^H$ is zero. Nevertheless, the beamspace leakage is

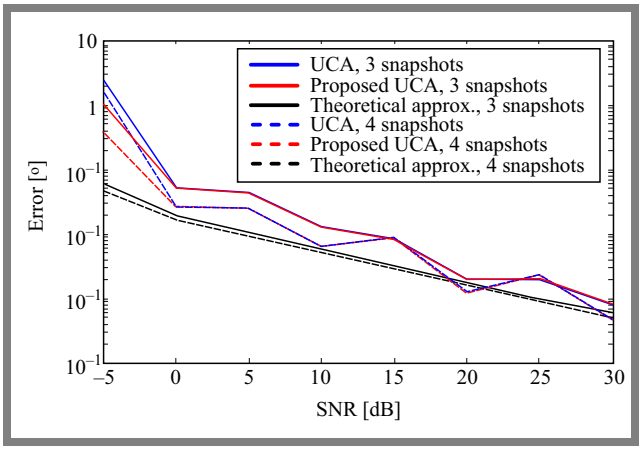

Fig. 2. Estimation error of DOA versus SNR for a single-source scenario ($\varphi = 50°$).

non-zero when the number of snapshots is small, because $\hat{\mathbf{E}}_s$ is not fully orthogonal to $\hat{\mathbf{E}}_n$. The beamspace leakage of the conventional algorithm can be expressed as [27]:

$$\rho_1 = \frac{1}{K}\sum_{k=1}^{K} \left\| \mathbf{E}_n \, \mathbf{E}_n^H \hat{\boldsymbol{e}}_k \right\| \cdot \frac{1}{K} \operatorname{Tr}\left((\mathbf{S})^{\dagger} \Delta \mathbf{R}_y' \mathbf{E}_n \, \mathbf{E}_n^H \Delta \mathbf{R}_y (\mathbf{S})^{\dagger} \right) , \tag{24}$$

where

$$\begin{aligned} (\mathbf{S})^{\dagger} &= \mathbf{S}(\mathbf{S}^H \mathbf{S})^{-1}, \\ \mathbf{S} &= \mathbf{W}^H \mathbf{A} \mathbf{R}_s \, \mathbf{A}^H \mathbf{W}, \\ \Delta \mathbf{R}_y' &= \hat{\mathbf{R}}_y - \mathbf{R}_y. \end{aligned}$$

For the proposed algorithm with one repetition (Step 3 in Section 3), the beamspace leakage is given as:

$$\begin{aligned} \rho_2 &= \frac{1}{K}\sum_{k=1}^{K} \left\| \mathbf{E}_n \, \mathbf{E}_n^H \hat{\boldsymbol{e}}_k'' \right\| \\ &= \frac{1}{K} \operatorname{Tr}\left((\mathbf{S})^{\dagger} \Delta \mathbf{R}_y'' \, \mathbf{E}_n \, \mathbf{E}_n^H \Delta \mathbf{R}_y'' (\mathbf{S})^{\dagger} \right) \end{aligned} , \tag{25}$$

where

$$\Delta \mathbf{R}'' = \hat{\mathbf{R}}_y - \mathbf{R}_y - \gamma \left(\mathbf{C}_B + \mathbf{C}_B^H \right) .$$

4.3. Computational Complexity

The computational complexity of the conventional UCARM algorithm is dominated by the eigen-decomposition and polynomial rooting steps, resulting in an overall complexity of $\mathcal{O}(P^3 + N \cdot K \cdot P^2)$. The proposed refinement technique introduces one additional update of the BSCM with computational cost $\mathcal{O}(P^2 K + P^3)$ per iteration. Since only a single iteration is applied in practice, the total complexity remains comparable to that of conventional UCARM, while offering significantly improved estimation accuracy. The additional computational overhead is therefore bounded and does not alter the overall cubic-order complexity of the underlying subspace-based framework.

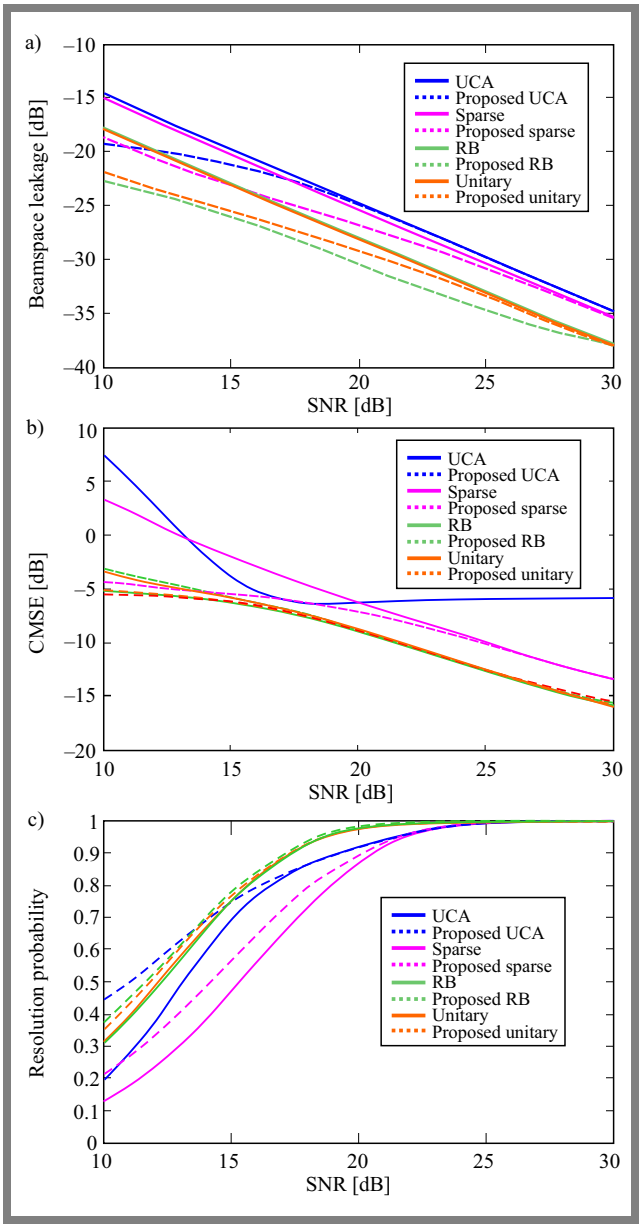

Fig. 3. Performance for uncorrelated signals versus SNR: a) beamspace leakage, b) CMSE, and c) resolution probability.

5. Simulation Results

In the simulation setup, a UCA consisting of 17 elements is used. The radius of the matrix is set to a constant value (λ/2), and the tuning parameter α varies from 0 to 1, in increments of 0.1. All signals are assumed to arrive at a fixed elevation angle of 80°, and a single iteration of step 3 (from the proposed algorithm) is considered in the simulations.

To further validate the effectiveness of the proposed technique, three additional proposed variants are also implemented for comparison: UCARB RM, sparse UCARM, and unitary UCARM. For clarity, the following abbreviations are used: UCARM [11] as UCA, sparse UCARM [23] as sparse, UCARBRM [11] as RB, and unitary UCARM [19] as unitary. The sparse UCARM configuration utilizes nine antenna elements.

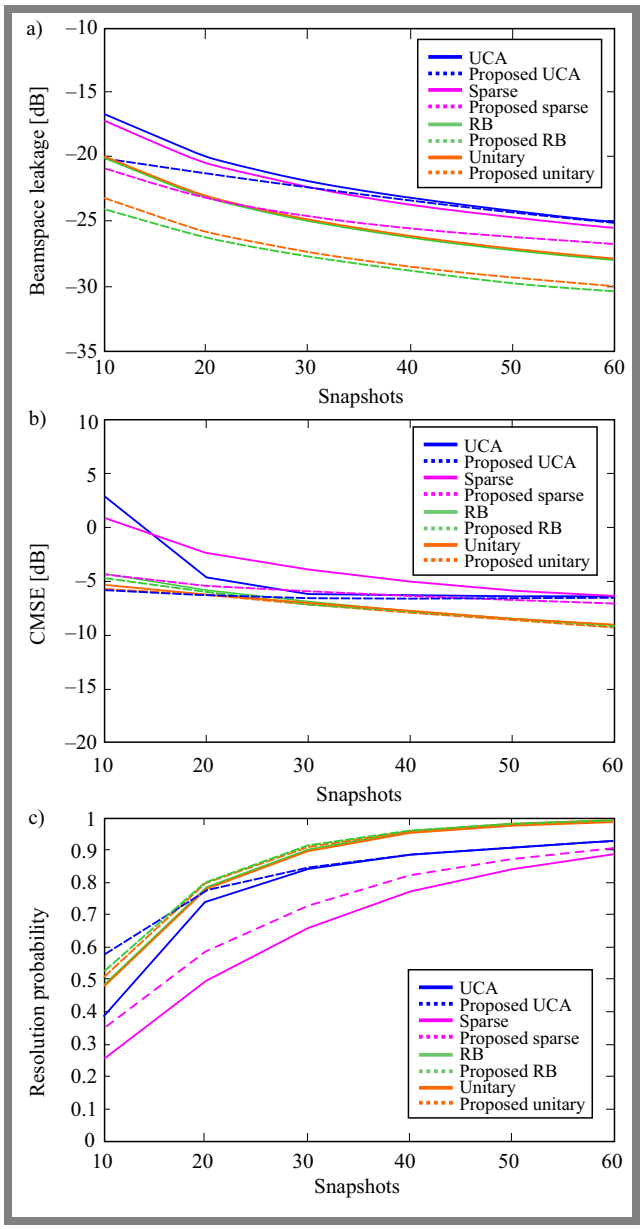

Fig. 4. Performance for uncorrelated signals versus snapshots: a) beamspace leakage, b) CMSE, and c) resolution probability.

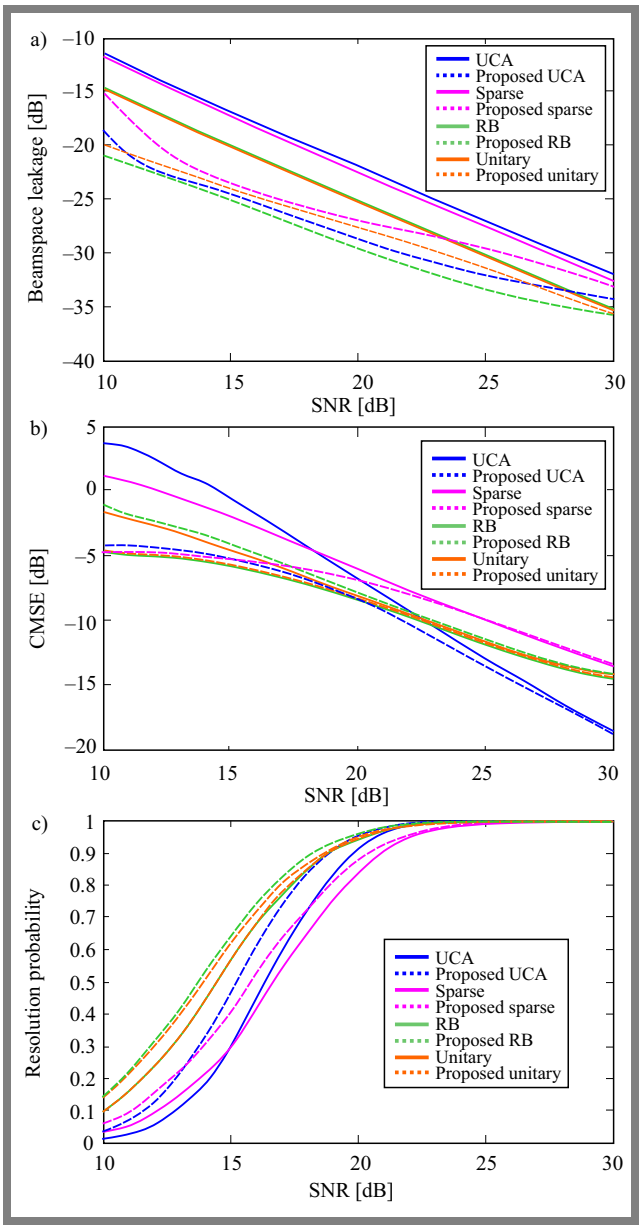

Fig. 5. Performance for correlated signals versus SNR: a) beamspace leakage, b) CMSE, and c) resolution probability.

5.1. Error Performance Analysis

In this scenario, a single-source signal impinging from the azimuth angle of $\varphi = 50°$ is considered. The error performance of the proposed algorithm is evaluated using 1 000 Monte Carlo simulations under two snapshot conditions: 3 and 4 snapshots. As illustrated in Fig. 2, the relationship between estimation error and SNR shows that the proposed UCA-based algorithm consistently outperforms its conventional counterpart in both snapshot cases, particularly in low SNR regimes. This improved performance is primarily attributed to the elimination of residual components in the BSCM, which effectively mitigates interference between the signal and noise subspaces in the beamspace domain.

5.2. Performance for Uncorrelated Signals

To evaluate the performance of the proposed algorithms in terms of beamspace leakage, conditional mean squared error (CMSE), and resolution probability simulations are conducted using equi-powered uncorrelated signals. CMSE is defined as the mean squared error under the condition that all sources are detected correctly. The resolution probability refers to the probability that the difference between the estimated and true DOAs is less than one degree for all signal directions. The evaluations are based on 10 000 Monte Carlo trials.

In the simulation analysis, two uncorrelated signals impinge on the array from azimuth angles of 50° and 56°, with the number of snapshots fixed at 15. The results for beamspace leakage, CMSE, and resolution probability versus SNR are shown in Fig. 3.

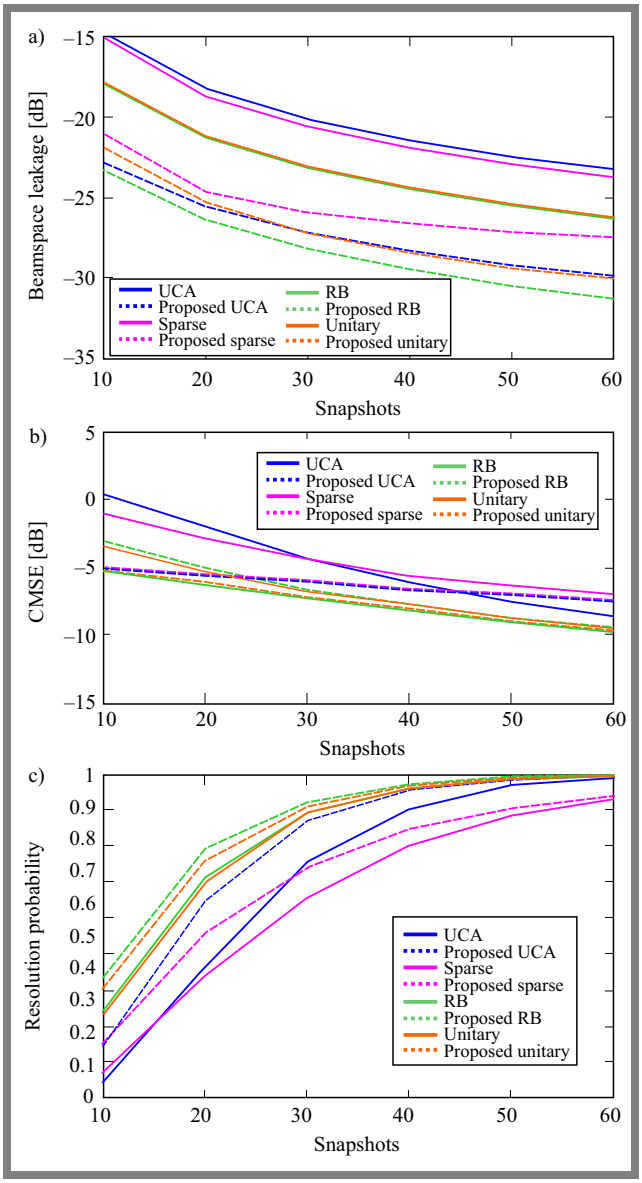

Fig. 6. Performance for correlated signals versus snapshots: a) beamspace leakage, b) CMSE, and c) resolution probability.

As depicted in Fig. 3a, all proposed algorithms demonstrate significantly lower beamspace leakage than their respective conventional counterparts. Beamspace leakage is reduced by approximately 5 dB when the SNR is below 13 dB. The proposed RB algorithm achieves the lowest leakage due to the use of forward/backward averaging.

Figure 3b illustrates that all proposed algorithms achieve significant improvements in CMSE over conventional ones, with the proposed UCA showing the most significant enhancement. Similarly, Fig. 3c shows that the proposed methods outperform their counterparts in terms of resolution probability. At SNR values below 15 dB, the proposed UCA provides the best resolution performance, while at higher SNRs, the proposed RB achieves superior accuracy.

Figure 4 shows the performance on beamspace leakage, CMSE and resolution probability versus snapshots with SNR=14 dB. The results versus snapshots are similar to those versus SNR.

5.3. Performance for Correlated Signals

Assume that two correlated signals with the correlation coefficient of 0.8 impinge on the UCA from $\varphi_1 = 50^\circ$ and $\varphi_2 = 60^\circ$. The number of snapshots is 15. Figure 5 illustrates the performance of the algorithms in terms of beamspace leakage, CMSE, and resolution probability as functions of SNR.

As observed in Fig. 5a, all proposed methods outperform their conventional counterparts in reducing beamspace leakage. The proposed UCA algorithm achieves the most significant reduction exceeding 6 dB, when the SNR is 10 dB. Furthermore, Fig. 5b confirms that the proposed method effectively lowers CMSE, primarily due to the improved formulation of the BSCM, which is consistent with the trend observed earlier in Fig. 3b. Moreover, as shown in Fig. 5b, the proposed RB algorithm delivers outstanding performance, primarily due to its real-valued processing scheme. In Fig. 5c, all four proposed algorithms demonstrate significant improvements in resolution probability compared to their conventional counterparts.

Figure 6 illustrates the impact of varying the number of snapshots on the performance of two correlated signals at a fixed SNR of 15 dB. As expected, the proposed algorithms consistently outperform their traditional versions across all metrics.

From Figs. 3 to 6, it is evident that all proposed algorithms consistently outperform their conventional counterparts in terms of SNR and snapshot variations, both for uncorrelated or correlated signal scenarios. These results confirm that the proposed approach can be extended to a variety of BT-based Root-MUSIC (RM) frameworks, including UCA-RARE methods for 2D DOAE applications. Therefore, the BSCM refinement technique introduced in this work offers broad applicability for enhancing the accuracy of DOA estimation.

Among the proposed methods, the RB and unitary variants show the best performance for uncorrelated signals, while the UCA and Sparse variants perform comparatively worse. The sparse algorithm shows the weakest results, which can be attributed to the reduced number of its antenna elements. On the other hand, Figs. 5 and 6 clearly indicate that the proposed RB method delivers superior efficiency in scenarios involving correlated signals.

Figure 7 shows the probability of resolution as a function of the correlation coefficient with a fixed SNR of 15 dB. As observed, all proposed algorithms significantly outperform their conventional counterparts. Among them, the proposed unitary algorithm exhibits the best overall performance.

In Fig. 8, the performance of the proposed RB algorithm is examined with respect to resolution probability versus angular separation, considering both correlated and uncorrelated signal scenarios.

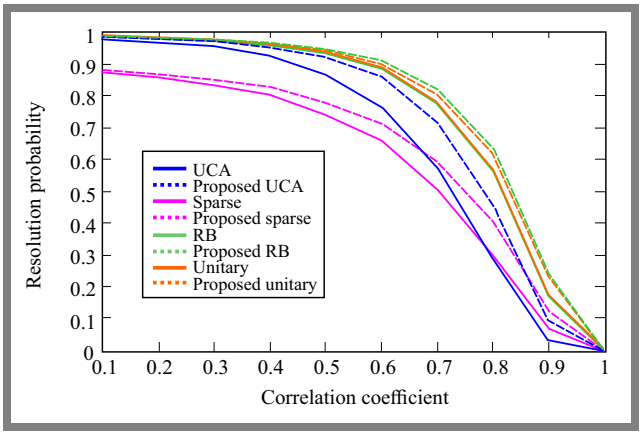

Fig. 7. Resolution probability versus correlation coefficient.

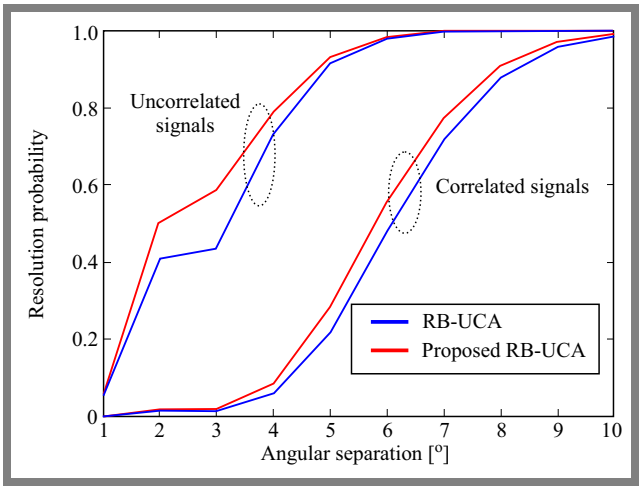

Fig. 8. Probability of resolution versus angular separation for correlated ($\rho = 0.7$) and uncorrelated cases.

The correlation coefficient for the correlated case is set to 0.7, and the SNR is fixed at 20 dB. The results of Fig. 8 confirm that the proposed RB algorithm consistently outperforms the conventional RB method, regardless of the presence or absence of signal correlation.

6. Conclusions

In this paper, an iterative enhancement technique was proposed to improve the accuracy of direction-of-arrival estimation (DOAE) for uniform circular arrays (UCA), particularly under limited snapshot conditions. The method is based on repeatedly refining the beamspace sample covariance matrix (BSCM) by eliminating residual components that degrade performance, especially at low signal-to-noise ratios (SNRs) or when few observations are available.

The technique integrates an initial estimation using the root-MUSIC algorithm in the beamspace domain, followed by iterative correction of the BSCM, using least-squares reconstruction and controlled adjustment of a tunable regularization parameter. This approach results in improved accuracy of the estimated directions through better modeling of the signal and noise subspaces.

Detailed simulation results demonstrate that the proposed method significantly outperforms conventional UCARM, sparse UCARM, UCARBRM, and unitary UCARM methods in terms of estimation error, beamspace leakage, conditional mean square error (CMSE) and resolution probability for both uncorrelated and correlated signal scenarios. The proposed technique is also applicable to 2D DOAE frameworks such as RARE-based algorithms, while preserving the underlying beamspace covariance structure.

Overall, the proposed BSCM refinement strategy offers a promising solution for high-accuracy DOAE in UCA configurations with limited data and can be readily integrated into a wide range of beamspace-based subspace estimation methods. Compared with previous single-step refinement methods, such as [1], the proposed iterative strategy provides more consistent performance across different SNRs and snapshot sizes, while maintaining comparable computational complexity.

References

[1] G. Jiang, X. Mao, and Y. Liu, "Reducing Errors for Root-MUSIC-based Methods in Uniform Circular Arrays", *IET Signal Processing*, vol. 11, pp. 1055–1062, 2017 (https://doi.org/10.1049/iet-spr.2016.0576).

[2] Y. Zhang, Y. Wang, and Y. Zhang, "DOA Estimation Based on Deep Learning with Sparse Uniform Circular Array", *IET Radar, Sonar & Navigation*, vol. 15, pp. 1171–1179, 2021 (https://doi.org/10.1049/rsn2.12168).

[3] A. Aydoğan and B.S. Yarman, "Enhanced Beamspace Processing for Direction-of-arrival Estimation Using Uniform Circular Arrays", *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 28, pp. 974–987, 2020 (https://doi.org/10.3906/elk-1903-125).

[4] H. Li, J. Wang, and X. Chen, "Robust Direction-of-arrival Estimation Approach Using Beamspace Transform and Array Imperfections Compensation", *IET Radar, Sonar & Navigation*, vol. 16, pp. 1605–1619, 2022 (https://doi.org/10.1049/rsn2.12282).

[5] M. Sharifi and P. Rezaei, "Conformal Antenna Array Radiation Pattern Synthesis by Tilt Correction to Improve Direction-of-arrival Estimation Accuracy", *Electromagnetics*, vol. 40, pp. 262–275, 2020 (https://doi.org/10.1080/02726343.2020.1757138).

[6] Y. Wang, W. Tan, Z. Deng, and G. Li, "Low-complexity Sparse Beamspace DOA Estimation via Single Measurement Vectors for Uniform Circular Array", *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 29, pp. 1973–1987, 2021 (https://doi.org/10.3906/elk-2010-109).

[7] M. Sharifi and P. Rezaei, "Near Optimal Conformal Antenna Array Structure for Direction-of-arrival Estimation", *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 29, art. no. e21978, 2019 (https://doi.org/10.1002/mmce.21978).

[8] T. Chen, L. Liu, and L. Guo, "Joint Carrier Frequency and DOA Estimation Using a Modified ULA Based MWC Discrete Compressed Sampling Receiver", *IET Radar, Sonar & Navigation*, vol. 12, pp. 873–881, 2018 (https://doi.org/10.1049/iet-rsn.2017.0505).

[9] S. Shi, Y. Xu, and Z. Liu, "Block Sparse Representation Approach to 2D DOA and Polarization Estimation of Wideband Signals Using a Sparse Vector Antenna Array", IET Radar, Sonar & Navigation, vol. 14, pp. 1929–1939, 2020 (https://doi.org/10.1049/iet-rsn.2020.0152).

[10] A. Akgül and B.S. Yarman, "Robust Beamspace DOA Estimation for Uniform Circular Arrays Using Modified Correlation Processing", *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 27, pp. 3006–3018, 2019 (https://doi.org/10.3906/elk-1811-30).

[11] C.P. Mathews and M.D. Zoltowski, "Eigenstructure Techniques for 2-D Angle Estimation with Uniform Circular Arrays", *IEEE Transactions on Signal Processing*, vol. 42, pp. 2395–2407, 1994 (https://doi.org/10.1109/78.317861).

[12] H. Li, J. Wang, and X. Chen, "Robust Direction-of-arrival Estimation Approach Using Beamspace Transform and Array Imperfections Compensation", *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 30, pp. 1605–1619, 2022 (https://doi.org/10.3906/elk-2111-127).

[13] A. Liu *et al.*, "Augmented Subspace MUSIC Method for DOA Estimation Using Acoustic Vector Sensor Array", *IET Radar, Sonar & Navigation*, vol. 13, pp. 969–975, 2019 (https://doi.org/10.1049/iet-rsn.2018.5440).

[14] M. Sharifi, P. Rezaei, and S. Mohammadi, "DOA Estimation Accuracy Improvement by Solving the Rank Reduction Algorithm Ambiguous Parameters Through Synthesizing the Volume Antenna Beam Pattern", *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 32, art. no. e23465, 2022 (https://doi.org/10.1002/mmce.23465).

[15] K.M. Reddy and V.U. Reddy, "Analysis of Spatial Smoothing with Uniform Circular Arrays", *IEEE Transactions on Signal Processing*, vol. 47, pp. 1726–1730, 1999 (https://doi.org/10.1109/78.765637).

[16] N.P. Waweru, D.B.O. Konditi, and P.K. Langat, "Performance Analysis of MUSIC, Root-MUSIC and ESPRIT DOA Estimation Algorithm", *International Journal of Electronics and Communication Engineering*, vol. 8, pp. 209–216, 2014.

[17] M. Sharifi, M. Boozari, M.G. Alijani, and M.H. Neshati, "Development a New Algorithm to Reduce SLL of an Equally Spaced Linear Array", *2018 Iranian Conference on Electrical Engineering (ICEE)*, Mashhad, Iran, 2018 (https://doi.org/10.1109/ICEE.2018.8472591).

[18] C. Kassis, J. Picheral, and C. Mokbel, "Advantages of Nonuniform Arrays Using Root-MUSIC", *Signal Processing*, vol. 90, pp. 689–695, 2010 (https://doi.org/10.1016/j.sigpro.2009.08.006).

[19] F. Belloni and V. Koivunen, "Unitary Root-MUSIC Technique for Uniform Circular Array", *2003 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Darmstadt, Germany, 2003 (https://doi.org/10.1109/ISSPIT.2003.1341160).

[20] M. Pesavento, A.B. Gershman, and M. Haardt, "Unitary Root-MUSIC with a Real-valued Eigendecomposition: A Theoretical and Experimental Performance Study", *IEEE Transactions on Signal Processing*, vol. 48, pp. 1306–1314, 2000 (https://doi.org/10.1109/78.839978).

[21] C. Qian, L. Huang, and H.C. So, "Improved Unitary Root-MUSIC for DOA Estimation Based on Pseudo-noise Resampling", *IEEE Signal Processing Letters*, vol. 21, pp. 140–144, 2013 (https://doi.org/10.1109/LSP.2013.2293949).

[22] G.J. Jiang, X.P. Mao, and Y.T. Liu, "Coprime Sparse Circular Array with Little Angular Dependence and Reduced Mutual Coupling", *AEU – International Journal of Electronics and Communications*, vol. 117, art. no. 153051, 2020 (https://doi.org/10.1016/j.aeue.2020.153051).

[23] R. Goossens, H. Rogier, and S. Werbrouck, "UCA Root-MUSIC with Sparse Uniform Circular Arrays", *IEEE Transactions on Signal Processing*, vol. 56, pp. 4095–4099, 2008 (https://doi.org/10.1109/TSP.2008.921764).

[24] F. Belloni and V. Koivunen, "Beamspace Transform for UCA: Error Analysis and Bias Reduction", *IEEE Transactions on Signal Processing*, vol. 54, pp. 3078–3089, 2006 (https://doi.org/10.1109/TSP.2006.875395).

[25] V. Vasylyshyn, "Removing the Outliers in Root-MUSIC via Pseudo-noise Resampling and Conventional Beamformer", *Signal Processing*, vol. 93, pp. 3423–3429, 2013 (https://doi.org/10.1016/j.sigpro.2013.04.012).

[26] J. Cao *et al.*, "Broadband Signal DOA Estimation Based on Two-sided Correlation Transformation Using Array Manifold Interpolation", *IET Radar, Sonar & Navigation*, vol. 19, art. no. e12669, 2025 (https://doi.org/10.1049/rsn2.12669).

[27] M. Shaghaghi and S.A. Vorobyov, "Subspace Leakage Analysis and Improved DOA Estimation with Small Sample Size", *IEEE Transactions on Signal Processing*, vol. 63, pp. 3251–3265, 2015 (https://doi.org/10.1109/TSP.2015.2422683).

[28] B.H. Wang, H.T. Hui, and M.S. Leong, "Decoupled 2D Direction of Arrival Estimation Using Compact Uniform Circular Arrays in the Presence of Elevation-dependent Mutual Coupling", *IEEE Transactions on Antennas and Propagation*, vol. 58, pp. 747–755, 2009 (https://doi.org/10.1109/TAP.2009.2039299).

[29] J. Dai *et al.*, "A Recursive RARE Algorithm for DOA Estimation with Unknown Mutual Coupling", *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1593–1596, 2014 (https://doi.org/10.1109/LAWP.2014.2345564).

[30] P. Stoica and A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-arrival Estimation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1783–1795, 1990 (https://doi.org/10.1109/29.60109).

[31] A.L. Swindlehurst and T. Kailath, "A Performance Analysis of Subspace-based Methods in the Presence of Model Errors. I. The MUSIC Algorithm", *IEEE Transactions on Signal Processing*, vol. 40, pp. 1758–1774, 1992 (https://doi.org/10.1109/78.143447).

[32] J. Xie, Z. He, H. Li, and J. Li, "2D DOA Estimation with Sparse Uniform Circular Arrays in the Presence of Mutual Coupling", *EURASIP Journal on Advances in Signal Processing*, vol. 2011, art. no. 80, 2011 (https://doi.org/10.1186/1687-6180-2011-80).

[33] L. Yang and Y. Zhang, "Performance Analysis of Root-MUSIC Algorithm for DOA Estimation Under Small Sample Size", *IET Radar, Sonar & Navigation*, vol. 1, pp. 91–98, 2007 (https://doi.org/10.1049/iet-rsn:20060014).

Mahdi Sharifi, Ph.D.
Department of Electrical and Computer Engineering
https://orcid.org/0000-0003-2675-8189
E-mail: m.sharifi@semnan.ac.ir
Semnan University, Semnan, Iran

Enhancing Data Transmission Security and Reliability of OFDM-IM for 5G Wireless Communication Systems

Asaad H. Sahar[1], Aqiel N. Almamori[1], and Muhannad Y. Muhsin[2]

[1] *University of Baghdad, Baghdad, Iraq,*
[2] *University of Technology – Iraq, Bahdad, Iraq*

https://doi.org/10.26636/jtit.2026.1.2502

Abstract — Thanks to its improved spectral efficiency and immunity to frequency selective fading, OFDM with index modulation (OFDM-IM) has become a perspective option. Unfortunately, OFDM-IM systems are vulnerable to security risks due to their inherent openness encountered in wireless communications. Conventional encryption techniques, which focus on the upper layers, add complexity and might not be enough to fend off malicious attacks. To improve the selection of subcarrier indexes and modulation of data symbol modulation, this work proposes a new chaotic encryption approach for OFDM-IM systems that uses Lorenz chaotic maps. Comprehensive simulations show that, in comparison to traditional methods, the proposed approach provides better security against eavesdropping while maintaining transmission reliability.

Keywords — 5G, chaotic encryption, Lorenz chaotic maps, OFDM-IM, subcarrier index selection

1. Introduction

5G and beyond communication systems provide high-capacity, high-security, ultra-reliable, low-latency connectivity, and high spectral efficiency features to support new services and applications. However, the growing demands faced by 5G communication systems might be difficult to handle by traditional orthogonal frequency division multiplexing (OFDM) solutions.

Recently, OFDM with index modulation (OFDM-IM) has gained popularity and is used extensively in several communication technologies, including optical communications, cognitive radio, and multiple input, multiple output (MIMO) networks [1]. As far as OFDM-IM systems are concerned, information is transmitted through constellation carrier and subcarrier indices [1]–[6]. The high spectral efficiency and reliability tradeoff, flexible spectral efficiency, and achievable ergodic rate are all attractive advantages that OFDM-IM displays over conventional OFDM systems [3]. Unfortunately, due to their inherent openness and broadcast nature, wireless communications are susceptible to eavesdropping and malicious attacks.

Eavesdroppers could blindly estimate several transmission parameters with respect to traditional OFDM-IM systems. Therefore, in practical applications, the issue of security of OFDM-IM systems should be considered. Upper layers are the main focus of most encryption techniques. Key management, distribution, and generation are all part of the matching encryption process. As a result, the computational overhead that has been introduced cannot be disregarded, and the encryption procedure at upper layers is also somewhat complicated.

Furthermore, with ever-growing computing power comes the possibility of security threats during the key distribution and key update phases. Consequently, there is a rather high level of complexity created and insufficient security is ensured by the encryption algorithms at the upper layers.

As an addition to upper-layer encryption, physical layer encryption may be rather advantageous [4]. With regard to encryptions of the physical layer, characteristics of wireless channels as well as the encryption technique are frequently used to protect the communication. Encryption keys could be generated by taking advantage of physical characteristics of the wireless channel, including reciprocity, randomness, as well as temporal and spatial variations [4], [7]. Furthermore, chaotic maps offering characteristics such as pseudo randomness, ergodicity, and sensitivity to initial values are frequently used in OFDM-tailed models to enhance security performance [8]–[10]. Chaotic systems are highly suitable for stream ciphering, randomization of signal order randomization, and other processes sue to their unpredictable and distinct characteristics regarding chaotic sequences.

2. Related Works

In the previous research, several encryption approaches of the physical layer have been proposed to improve the security of OFDM systems. Dynamic interleaving was utilized to create an eavesdropping-resistant OFDM system that has been described in [7]. In the time division duplexing (TDD) mode, sub-carrier interleaving depends on CSI (i.e., channel state information) between the authorized users. A constellation rotation approach based on chaos has been used in [8] to prevent eavesdropping on OFDM systems and chaotic maps were used for varying the phases of constellation symbols.

For the purpose of ensuring the security of OFDM systems, chaotic maps have been utilized in [9] for the randomization

of the sub-carrier phase and order. For OFDM systems, an artificial noise-based encryption method has been used in [10], [11] and several algorithms for the physical layer encryption that have been designed specifically for OFDM-IM.

In [12], selection of the optimum subcarrier index as well as adaptive interleaving have been utilized for the purpose of improving the security of OFDM-IM systems. To prevent eavesdropping on the OFDM-IM system, a secure index and data symbol modulation have been relied upon in [13]. The rule of randomized mapping for the data symbols and index modulations has been developed based on the CSI. An encryption approach based upon artificial noise was introduced for OFDM-IM systems in [14], where artificial noise is coupled with a legitimate signal that can be eliminated at the side of the authorized receivers.

With regard to OFDM-IM systems, there is still potential for improvement over previous physical layer encryption techniques. The significant degree of modification and fairly high complexity added to the transceiver used in conventional OFDMIM systems are the recurring characteristics of the encryption techniques presented in [12]–[14]. Additionally, data symbol modulation security has not been considered by the encryption schemes, as they concentrate on the selection of the sub-carrier index.

In this work, we propose an efficient and straightforward encryption method for OFDM-IM systems. The method could protect modulation of the data symbol, additionally to ensuring security of the sub-carrier index selection process. The phase of modulated data symbols as well as indexes of the active subcarrier are both randomly generated using chaotic maps. In the TDD mode, legitimate users retrieve secret keys, being the initial chaotic map values, from CSI. Due to the reciprocity of the wireless channel in the TDD mode, legitimate users acquire identical secret keys.

Secure operation of the proposed technique may be ensured, since the spatial variance of the wireless channel prevents an eavesdropper at a third party from obtaining secret keys. Furthermore, there is a large space in the key space to defend against comprehensive attacks. Specifically, compared to the previous techniques presented in [12]–[14], only modest adjustments are required to typical OFDM-IM systems, and added complexity is not that great. Extensive simulations confirm that it could attain the same reliability as traditional OFDM-IM systems

3. Preliminaries

In this section, the basics of standard OFDM-IM systems using frequency-selective Rayleigh fading channel types are described, and then the foundations of Lorenz chaotic maps are covered in more detail.

3.1. OFDM-IM

The OFDM-IM transmitter receives a total of m bits of transmitted data. The condition $m = l \cdot q$ is satisfied by

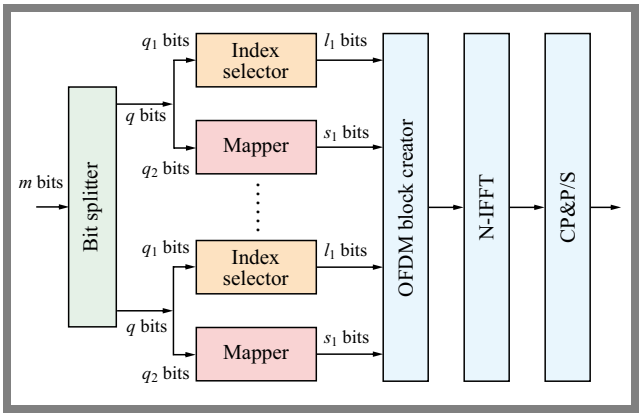

Fig. 1. Block diagram of a conventional OFDM-IM system.

the first division of m bits of information into l subblocks, each of which is q long, as shown in Fig. 1. The OFDM subblock with a length of n was assigned to each of q bits. $N = l \cdot n$ represents the number of OFDM sub-carriers. The information bits of each subblock were split to data symbol bits and index bits due to bit-splitter effects. To calculate active sub-carrier indices, the first bits of q_1 serve as index bits. Based on a series of predefined instructions depending upon the incoming index bits, just k out of n subcarriers are active in OFDM-IM systems.

In conventional OFDM-IM systems, a combinatorial or lookup table technique is a commonly used selection method [1]. The combinatorial technique can be viable in the cases where there is a considerable number of information bits, while the lookup table approach is frequently applied to conditions where there are few information bits. For the purpose of creating data symbols, the remaining $q_2 = k \log_2 M$ bits are mapped into an M-ary signal constellation over time. Additionally, data symbols use active indices for modulating subcarriers [1], [15]. Equation (1) is used to determine the sum of the information bits carried through subcarrier index values.

$$m_1 = q_1 l = \lfloor \log_2 C(n,k) \rfloor l \, . \tag{1}$$

The selected active subcarrier indices for subblock γ can be represented as:

$$l_\gamma = \{ i_{\gamma,1}, i_{\gamma,2}, \ldots, i_{\gamma,k} \} \, , \tag{2}$$

where $\gamma = 1, \cdots, l$.

The total number of bits of information that are carried by constellation symbols is represented as [1]:

$$m_2 = q_2 l = k(\log_2 M) l \, . \tag{3}$$

Modulated symbols that are carried out through the active subcarrier k of subblock γ can be expressed as:

$$s_\gamma = \lfloor s_\gamma(1), \ldots, s_\gamma(k) \rfloor \, , \tag{4}$$

where $\gamma = 1, \cdots, l$.

After that, I_γ and s_γ are transmitted to the OFDM block creator, which constitutes the main block of the OFDM.

$$x_F = \left[x(1), x(2), \ldots, x(N) \right]^T \, , \tag{5}$$

The N-IFFT processing converting the frequency domain signal to a time domain signal could be written as follows:

$$x_T = \frac{1}{\sqrt{K}} F_N^H x_F \,, \tag{6}$$

where F_N^H represents the DFT matrix that satisfies $F_N^H F_N = NI_N$.

The time domain signal x_T is transmitted through a frequency-selective Rayleigh fading channel when the cyclic prefix is added and parallel to the serial conversion. In particular, a frequency domain signal satisfies the following equation:

$$y_F(\delta) = x(\delta)h_F(\delta) + n_F(\delta) \,, \tag{7}$$

where, $y_F(\delta)$ represents the received signal in the frequency domain, $n_F(\delta)$ and $h_F(\delta)$ represent the frequency domain of the noise and the frequency domain of the channel coefficient, respectively.

3.2. ML Detector

The ML detector for the OFDM-IM model takes under consideration all of the potential subblock realizations by searching for all of the potential combinations of the subcarrier index and points of signal constellation for the purpose of making joint decisions on constellation symbols and active indices for every one of the sub-blocks through the minimization of the following metric:

$$(\hat{I}_\beta, \hat{s}_\beta) = \arg \min_{I_\beta, s_\beta} \sum_{\gamma=1}^{k} \left| y_F^\beta(i_{\beta,\gamma}) - h_F^\beta(i_{\beta,\gamma}) s_\beta(\gamma) \right|^2 \,, \tag{8}$$

where $y_F^\beta(\xi)$ and $h_F^\beta(\xi)$ for $\xi = 1, \ldots, n$ represent received signals, as the well as corresponding fading coefficients for the subblock β, which can be represented as:

$$\begin{aligned} y_F^\beta(\xi) &= y_F\big(n(\beta-1)+\xi\big) \\ h_F^\beta(\xi) &= h_F\big(n(\beta-1)+\xi\big) \end{aligned} \,. \tag{9}$$

The total number of metric calculations that have been carried out in Eq. (8) is cM^k due to the fact that I_β and x_β have c and M^k various realizations, respectively. This is the reason why ML detector is considered as an impractical approach for larger c and k values as a result of its decoding complexity that is increasing exponentially [16], [17].

3.3. Hybrid Greedy Detector and Diversity Reception

Greedy detector is a detection model that has low complexity with a diversity reception. It requires a two-stage detection process: active subcarrier indices and corresponding M-ary symbols are calculated separately, as one may see in Fig. 2.

In the first stage, the greedy detector measures the combined output signal energy of every one of the sub-carriers $|y_\beta(j)|^2$ and, thereafter, detects active subcarriers with the highest level of energy. Subcarriers that are only under the favorable channel fading have a high probability of being estimated as active. For the purpose of reducing computational complexity, the initial stage does not require any information about the channel. After detecting non-zero M-ary symbols in x, the second step applies maximum likelihood (ML) decisions to every K estimated active subcarrier separately. Stages 1 and 2 of the greedy detector process are described below.

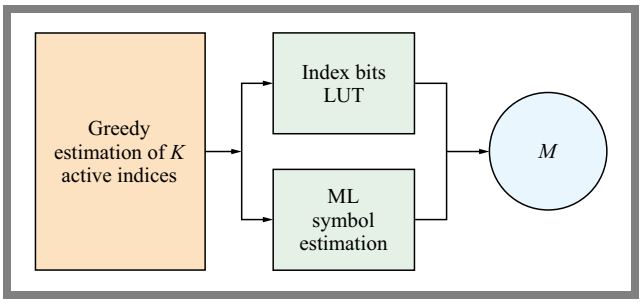

Fig. 2. Block diagram of a greedy detector.

Stage 1. Using K greatest amongst N received subcarrier signal powers, the receiver detects active subcarrier indices based on energy.

1) Assume a residual vector $z = y_\beta$, a demodulated vector r is set to 0 vector, i.e., $r = 0_{1\times N}$, and iteration count $t = 0$ is started.
2) Find the highest element whose index $\hat{\alpha}$ can be expressed as:
$$\hat{\alpha} = \arg \max_j |z(j)|^2 \,, \tag{10}$$
where $z(j)$ represents the j-th element of $\boldsymbol{z}$.
3) Assume $r(\hat{\alpha}) = z(\hat{\alpha})$ and $z(\hat{\alpha}) = 0$, and increment t by 1, where $r(\hat{\alpha})$ represents α-th element in r.
4) Steps 2 and 3 are repeated to the point where $t = K$.
5) Set all non-zero elements in $r = 1$. Recover m_I bits for r with the use of LUT. An example of LUT for cases where $K = 2$ and $N = 4$ is listed in Tab. 1.

Stage 2. M-ary symbols are estimated with the use of the ML criterion on the subcarriers of the active indices $\hat{\alpha}$ as:

$$\hat{x}(\hat{\alpha}) = \arg \max_{x(\hat{\alpha}) \in \varsigma} \left| y_\beta(\hat{\alpha}) - h(\hat{\alpha}) \times (\hat{\alpha}) \right|^2 \,, \tag{11}$$

where $h(\hat{\alpha})$ represents the α-th diagonal element with respect to H in Eq. (7), representing the equivalent channel matrix after diversity reception.

Note that GD utilizes the process of energy detection, which eliminates the need for a brute-force search of every potential index combination and greatly reduces complexity of the process of estimating a set of active subcarrier indices. This results in an additional reduction of the complexity of the carrier index while detecting M-ary symbols, which the latter process performed independently of the subcarrier index detection rate [18], [19].

Tab. 1. Example of LUT for $N = 4$ and $K = 2$.

r_K	m_I bits
1100	00
1010	01
1001	10
0110	11

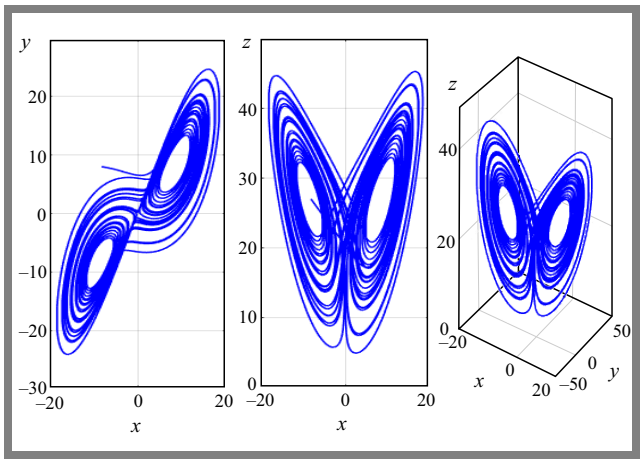

Fig. 3. Lorenz chaotic system phase diagram.

3.4. Lorenz Chaotic Maps

Chaotic sequences are particularly suited for secure communications due to their pseudo randomness, ergodicity, and sensitivity to initial values [15]. Unpredictable yet determined, chaotic sequences could produce entirely distinct output sequences with only a slight alteration from the initial value. As a result, chaotic systems are frequently used in watermarking, secure communication, and cryptography. Lorenz chaotic systems feature more dynamical behaviors and variables of the system than their low-dimensional counterparts, suggesting a larger key space and higher levels of unpredictability [16]. In order to finish the encryption in the proposed scheme, we use a Lorenz chaotic system due to the aforementioned features. The following is a mathematical expression of a typical Lorenz chaotic system [16]–[18]:

$$\begin{cases} \dot{x} = a(y - z) \\ \dot{y} = cx - xz - y \\ \dot{z} = xy - bz \end{cases}, \tag{12}$$

where $a = 10$, $b = 8/3$, $c = 28$.

A phase diagram of a Lorenz chaotic system has been depicted in Fig. 3. It starts with initial values x, y, and z which represent a set of CSI amplitude, phase, and frequency between legitimate users where the TDD mode is adopted. The following is the processing of chaotic sequences $\{xi\}$, $\{yi\}$, and $\{zi\}$, which have been based on the quantization approach introduced in [18]:

$$\begin{aligned} D_{xi} &= \mathrm{mod}(1\,000 \times x_i, 256)/256 \\ D_{yi} &= \mathrm{mod}(1\,000 \times y_i, 256)/256 \\ D_{zi} &= \mathrm{mod}(1\,000 \times z_i, 256)/256 \end{aligned}, \tag{13}$$

After that, each of the elements in D_{xi} is converted to 1 or 0 based on the binary sequence L_{xi}:

$$L_{xi} = \begin{cases} 0, & D_{xi} < 0.5 \\ 1, & D_{xi} \geqslant 0.5 \end{cases}, \tag{14}$$

We could acquire the corresponding chaotic binary sequences L_{yi} and L_{zi} through that same quantification technique. In the proposed scheme, two chaotic binary sequences are required to complete the encryption process. Thus, sequences L_{xi}, L_{yi} and L_{zi} are processed in the following way:

$$\begin{cases} B_1 = L_{xi} \bigoplus L_{yi} \\ B_2 = L_{yi} \bigoplus L_{zi} \end{cases}. \tag{15}$$

The Lorenz chaotic system generates two chaotic binary sequences as a result. The selection of the subcarrier index as well as the modulation of the data symbols are encrypted using the two chaotic binary sequences, respectively.

4. Problem Statement

Despite their efficiency, conventional OFDM systems are susceptible to eavesdropping and other security risks, as wireless communications are broadcast and open. Traditional encryption techniques concentrate primarily on the communication protocol's higher levels, which adds complexity and does not always offer enough protection. Furthermore, such techniques frequently entail complex key management procedures that could be time consuming and vulnerable to security breaches at the key distribution phase. A promising OFDM-IM encodes information in both the constellation symbols and subcarrier indices, thus increasing spectral efficiency and dependability. However, OFDM-IM systems are susceptible to security risks, just as their traditional OFDM counterparts are. Among the drawbacks of physical layer encryption techniques now in use for OFDM-IM are their high complexity, substantial transceiver structure modifications, and their major emphasis on the selection of the subcarrier index at the expense of data symbol modulation security.

Furthermore, the detection accuracy and computational complexity of OFDM-IM systems pose additional challenges. The maximum likelihood (ML) detector, while highly accurate, suffers from exponentially increased decoding complexity, making it impractical for larger values of subcarriers and modulation orders. On the other hand, simpler detection methods may not provide the required level of accuracy, especially in the presence of security mechanisms. Therefore, there is a need for a secure, efficient, and low-complexity encryption scheme that can protect both the subcarrier indices and the data symbols in OFDM-IM systems, while maintaining or improving performance and reliability of the transmission. In addition, an effective detection mechanism that balances accuracy and complexity is required. By introducing a new chaotic encryption approach, this work seeks to overcome such challenges. Furthermore, it employs a hybrid greedy detector (GD) alongside the ML detector to optimize the trade-off between computational complexity and detection accuracy, ensuring robust and efficient performance in practical applications.

5. System Model

The general model of a wiretap channel serves as a basis for security analysis [19]. As depicted in Fig. 4, Alice uses the wiretap channel for passive eavesdropping, whereas Eve

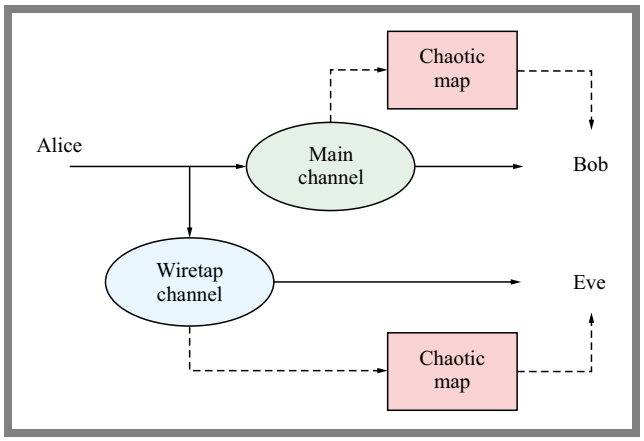

Fig. 4. General wiretap channel model.

tries to send Bob legitimate messages covertly via the secret channel. Given that Eve experiences independent channel fading, it is assumed that she is located half a wavelength away from the legitimate receiver. Because of the wireless channel's reciprocity, Bob and Alice could receive the same CSI if communication is conducted in the TDD mode and a perfect channel estimate is possible. It is important to remember that CSI must adhere to the wireless channel's reciprocity requirement and must not alter during the coherence time. Additionally, Bob and Alice establish the Lorenz chaotic system's initial values for the CSI's amplitude, frequency, and phase. Modulation of data symbols and subcarrier index selection are encrypted using the resulting chaotic sequences. Eve cannot acquire secret keys which represent initial Lorenz chaotic system.

The transmitter of the proposed chaotic encrypted OFDM-IM scheme, as depicted in Fig. 5, is mainly made up of a secure modulator of data symbols depending on chaotic maps and the index selector. Unlike traditional OFDM-IM systems, the method under consideration ensures two levels of security. Only the legitimate receiver is capable could accurately recovering original messages using the established mapping criteria.

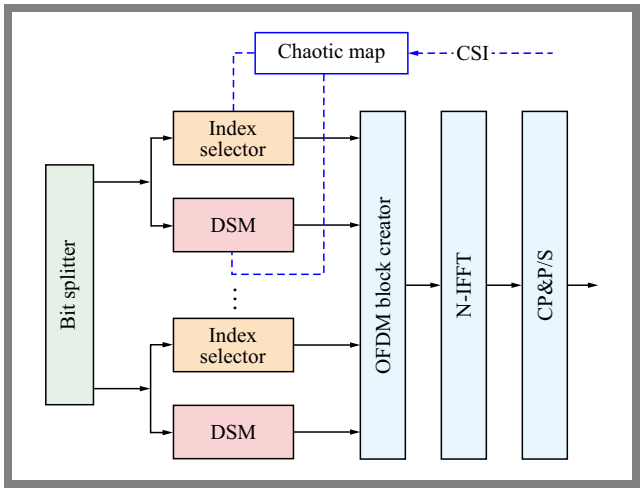

Fig. 5. Block diagram of the proposed transmitter of a chaotic encrypted OFDM-IM model.

5.1. Chaotic Maps on the Selections of the Subcarrier Indices

We provide an encrypted chaotic subcarrier index selector that is based on a combinatorial technique [1]. A natural number may be mapped to its corresponding combination by a combinatorial method, and the Lorenz chaotic system produces chaotic binary sequences which are used for randomizing the combination.

First, q_1 bits undergo conversion to the corresponding decimal value Z, and then this number is converted into the sequence $J = \{c_k, \cdots, c_1\}$ through the combinatory technique:

$$Z = C(c_k, k) + \ldots + C(c_2, 2) + C(c_1, 1)\,. \tag{16}$$

Indices regarding the active subcarrier are specified through $J + 1$. Let us assume that $D^k = \{d_k, \ldots, d_1\}$ denotes active sub-carrier indices and $k = 2^{r_1}, r_1 > 0$. We denote the corresponding chaotic binary sequence through $B_1^l = B_1(1), B_1(2), \ldots, B_1(l)$.

We convert each r_1 bit in B_1^l into a corresponding decimal number, therefore creating decimal sequence T^k that is k long. The decimal values that are contained in T^k traverse each one of the values between 0 and $k-1$ due to ergodicity of the chaotic sequences. After that, the indices of active subcarrier D^k are randomized.

Therefore, we acquire encrypted indices of active sub-carrier E^k:

$$\begin{cases} E(i_1) = D(i_2) \\ T(i_2) = i_1 \end{cases} \quad i_1, i_2 = 0, 1, \ldots, k-1\,. \tag{17}$$

5.2. Chaotic Maps on the Modulation of Data Symbols

We propose a data symbol modulator that is secure and relies on chaotic maps. Here, the chaotic encrypted data symbol modulation process is explained using 32-QAM as an example.

Let us assume $B_2^l = B_2(1), B_2(2), \ldots, B_2(l)$ represents a chaotic binary sequence that corresponds to modulation of encrypted data symbols. As shown in Fig. 6, $P_0(r_o, \theta_o)$ represents the original data symbol, while $P_c(r_c, \theta_c)$ represents a data symbol after mapping in a chaotic manner. Moreover, r_0, r_c represents symbol amplitude and θ_0, θ_c stands for symbols phase with following relation:

$$\begin{cases} r_c = r_0 \\ \theta_c = \theta_0 + 2\pi \frac{M_c}{M} \end{cases}, \tag{18}$$

where M represents a real constant and $M = 2^{r_2}, r_2 > 0$. Next, each r_2 bits in B_2^l are converted into decimal value M_c. Because of the ergodicity of chaotic maps, the range of M_c lies between 0 and $M-1$, thus $2 \times \pi \frac{M_c}{M}$ ranges between 0 and 2π. Let M be set as 256, r_2 be 8, and M_c could be specified through each eight bits in B_2^l.

5.3. Detection at the Receiver's Side

For the recovery of original messages, an ML detector has been used on the legitimate receiver side [1]. It is believed that the receiver side could accomplish a flawless CSI estimate.

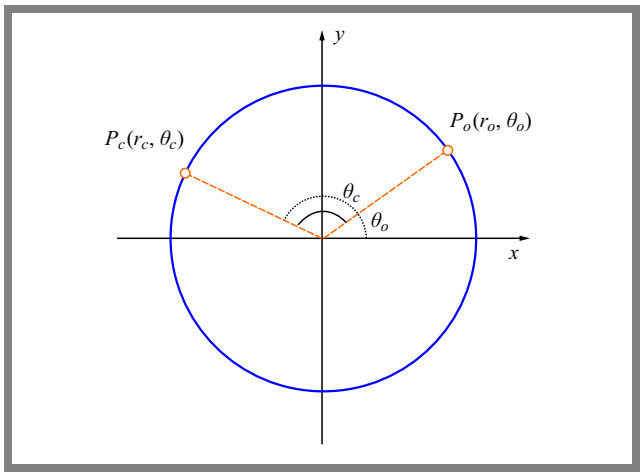

Fig. 6. Process of chaotic maps on the modulation of data symbols.

Inverse encryption operations could be carried out for Bob in order to recover the original key, because the secret users share secret keys as the well as the encryption method. The encrypted data symbol that has been carried by an active subcarrier could be recovered once indices of the sub-carrier are identified. Chaotic maps could be used to decrypt the secret key symbol using secret keys. Thus, at the legitimate receiver side, information carried through the encrypted modulated data symbols as well as the encrypted subcarrier indices could be recovered.

With regard to the multipath fading scenario, wireless channels are geographically separated. Therefore, Eve at a third party cannot acquire the same CSI as Bob and Alice. Eve is thus limited to trying to decode the message using a standard IM detector. Eve cannot decrypt the encrypted messages without secret keys, though, since the data symbol and indices of active the subcarrier are encrypted messages. Therefore, Eve could not get useful messages.

Furthermore, compared to schemes [12], [14], only modest modifications to the structure of the traditional OFDM-IM system are required throughout the detection phase. Because of the simplicity of the transceiver structure, this scheme is easy to implement in practical applications.

6. Performance Analysis

This section evaluates the performance of the proposed chaotic encryption method for the OFDM-IM systems, with a focus on both security and reliability.

6.1. Analysis of Secrecy

The proposed scheme leverages the initial values of the Lorenz chaotic system as secret keys. Given the sensitivity of the Lorenz system to initial conditions, even minute deviations (around 10^{-15}) can lead to entirely different output sequences. This characteristic results in a large key space, estimated at approximately 10^{45} ($10^{15}\times10^{15}\times10^{15}$), which offers robust protection against exhaustive search attacks and significantly improves system security compared to conventional encryption methods.

Traditional approaches, such as those referenced in [10] and [12], typically secure only one of these aspects, namely subcarrier indices or data symbols. Although method [13] addresses both, it is directly based on CSI without the additional chaotic encryption layer.

However, this approach integrates chaotic maps based on CSI, which are only known precisely to legitimate communication partners. Due to chaotic system's sensitivity, it is nearly impossible for an eavesdropper to replicate the same sequences without access to the initial conditions, making it exceedingly difficult for unauthorized parties to decode the intercepted signals. This dual-layer encryption significantly enhances OFDM-IM security.

6.2. Reliability Analysis

Monte Carlo simulations conducted over Rayleigh fading channels are used to validate the reliability of chaotic encrypted OFDM-IM system. The following parameters have been used in such simulations: each OFDM frame has 200 symbols, two of the four subcarriers are active, and the modulation method is 4QAM.

The results shown in Fig. 7 demonstrate that the bit error rate (BER) performance with respect to the encrypted OFDM-IM system is comparable to that of traditional OFDM-IM solutions. This indicates that the use of chaotic maps for encryption does not degrade the system's transmission performance.

Moreover, the simulation highlights that the transmission performance for unintended receivers – those without access to the secret keys – deteriorates significantly. This degradation validates the improved security offered by the proposed model, as eavesdroppers without private keys are not capable of effectively decoding transmitted signals.

6.3. Comparative Detection Performance

Besides the validation of transmission performance, the analysis compares the performance of two detection approaches: the hybrid greedy detector (GD) and the maximum likelihood

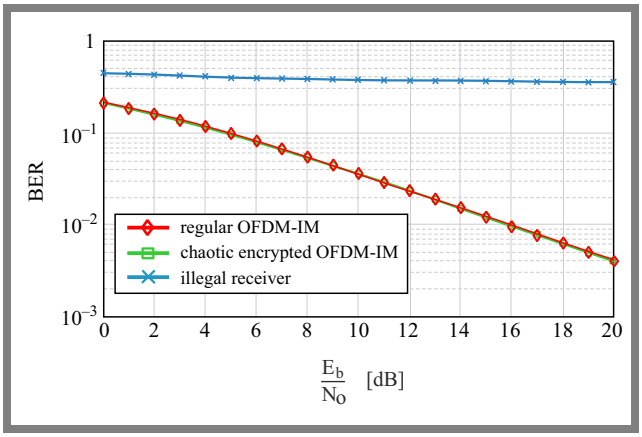

Fig. 7. BER performance of the proposed encrypted OFDM-IM system.

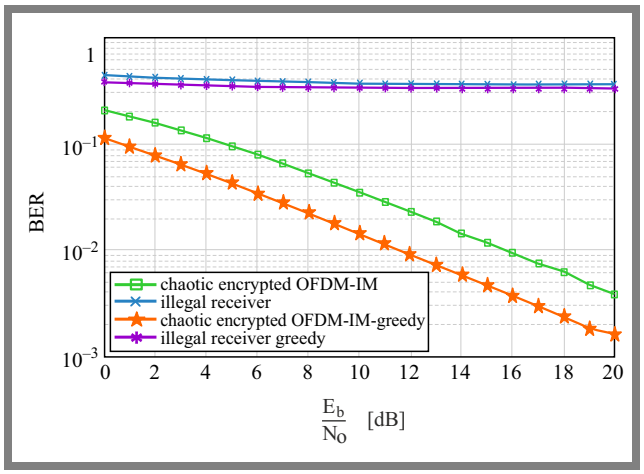

Fig. 8. Comparative detection performance: greedy vs. ML.

detector (ML). GD had demonstrated superior performance compared to ML in terms of computational efficiency and detection accuracy. GD operates in two stages: initially, it detects active subcarriers through the evaluation of received signal power, simplifying the detection process and reducing computational complexity. In the second stage, it applies the criterion of maximum potential for estimating M-ary symbols on identified active subcarriers. This method ensures good accuracy of the detection process and considerably reduces computational burden in comparison with the ML detector, in particular as the number of modulation orders and subcarriers is increased.

The ability of hybrid GD to efficiently and accurately detect active sub-carriers and their corresponding symbols makes it well-suited for the proposed encrypted OFDM-IM system. Its performance-related advantage over the ML detector is especially pronounced in scenarios with larger numbers of subcarriers and higher modulation schemes, where the computational complexity of the ML detector becomes impractical.

Although the proposed chaotic encryption approach based on Lorenz chaotic maps is responsible for boosting the security of OFDM-IM systems, the role of two-stage GD focuses primarily on reducing receiver side complexity and increasing detection robustness. Chaotic encryption randomizes both subcarrier indices and data symbol modulation, which enlarges the hypothesis space and increases index uncertainty at the receiver.

GD addresses this issue by performing an energy-based selection of active subcarriers in its first stage, thereby suppressing unlikely index hypotheses before symbol detection. This mechanism improves robustness against index uncertainty introduced by chaotic encryption and leads to improved BER performance that is shown in Fig. 8. It should be noted that GD does not enhance the security of the system itself, but rather enables efficient and reliable detection under the proposed encrypted OFDM-IM framework. This explains the BER behavior observed in Fig. 8 and confirms that the improved performance of GD under chaotic encryption does not contradict the theoretical optimality of the ML detector.

7. Conclusions

According to simulation results, the proposed scheme enhances the level of security without compromising transmission performance. Additionally, the illegal communication transmission performance has further validated the security. Furthermore, in comparison to the current encryption techniques, the added complexity and adjustment required for traditional OFDM-IM systems are negligible. It has been demonstrated that the suggested side-encrypted OFDM-IM system offers good prospects in terms of practical applications, taking into account both simulation results and theoretical analysis.

Future research could look into more techniques for removing secret keys from wireless channels. Moreover, various communication contexts could benefit from the application of chaotic encryption techniques to raise their security standards.

References

[1] B. Zhang and L. Liu, "Chaos-based Image Encryption: Review, Application, and Challenges", *Mathematics*, vol. 11, art. no. 2585, 2023 (https://doi.org/10.3390/math11112585).

[2] A. Dogukan and E. Basar, "Super-mode OFDM with Index Modulation", *IEEE Transactions on Wireless Communications*, vol. 19, pp. 7353–7362, 2020 (https://doi.org/10.1109/TWC.2020.3010839).

[3] X. Cheng, M. Zhang, M. Wen, and L. Yang, "Index Modulation for 5G: Striving to Do More with Less", *IEEE Wireless Communication*, vol. 25, pp. 126–132, 2018 (https://doi.org/10.1109/MWC.2018.1600355).

[4] X. Cheng, L. Wen, L. Yang, and Y. Li, "Index Modulated OFDM with Interleaved Grouping for V2X Communications", *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, 2014 (https://doi.org/10.1109/ITSC.2014.6957834).

[5] E. Basar, U. Aygolu, E. Panayirci, and H.V. Poor, "Orthogonal Frequency Division Multiplexing with Index Modulation", *IEEE Transactions on Signal Processing*, vol. 61, pp. 5536–5549, 2013 (https://doi.org/10.1109/TSP.2013.2279771).

[6] J. Wang *et al.*, "A High-security Physical Layer Encryption Scheme for Dual-mode Index Modulation–aided OFDM in Magnetic Induction Communication", *Optics Letters*, vol. 50, pp. 285–288, 2025 (https://doi.org/10.1364/OL.544682).

[7] H. Li, X. Wang, and J. Yves, "Eavesdropping-resilient OFDM System Using Sorted Subcarrier Interleaving", *IEEE Transactions on Wireless Communications*, vol. 14, pp. 1155–1165, 2015 (https://doi.org/10.1109/TWC.2014.2365031).

[8] A.A. Purwita, A. Yesilkaya, M. Safari, and H. Haas, "Generalized Time Slot Index Modulation for Optical Wireless Communications", *IEEE Transactions on Communications*, vol. 68, pp. 3706–3719, 2020 (https://doi.org/10.1109/TCOMM.2020.2979845).

[9] G. Cai *et al.*, "Design of a MISO-SWIPT-aided Code-index Modulated Multi-carrier M-DCSK System for e-health IoT", *IEEE Journal on Selected Areas in Communications*, vol. 39, pp. 311–324, 2021 (https://doi.org/10.1109/JSAC.2020.3020603).

[10] L. Fan, Q. Jiang, and F. Liu, "OFDM-based Waveform Design in Artificial Noise Aided Secure Communication", *2nd International Conference on Electronics Technology (ICET)*, Chengdu, China, 2019 (https://doi.org/10.1109/ELTECH.2019.8839524).

[11] L. Fan *et al.*, "Joint Resource Allocation for Temporal Artificial Noise Assisted Multiuser Wiretap OFDM Channels with Finite-alphabet Inputs", *2nd International Conference on Electronics Technology (ICET)*, Chengdu, China, 2019 (https://doi.org/10.1109/ELTECH.2019.8839391).

[12] J.M. Hamamreh, E. Basar, and H. Arslan, "OFDM-subcarrier Index Selection for Enhancing Security and Reliability of 5G URLLC Services", *IEEE ACCESS*, vol. 5, pp. 25863–25875, 2017 (https://doi.org/10.1109/ACCESS.2017.2768558).

[13] Y. Lee, H. Jo, Y. Ko, and J. Choi, "Secure Index and Data Symbol Modulation for OFDM-IM", *IEEE ACCESS*, vol. 5, pp. 24959–24974, 2017 (https://doi.org/10.1109/ACCESS.2017.2768540).

[14] X. Zhang, S. Zhang, and Z. Qiao, "A Chaos-based Encryption Scheme for OFDM-IM Systems", *2021 IEEE Symposium on Computers and Communications (ISCC)*, Athens, Greece, 2021 (https://doi.org/10.1109/ISCC53001.2021.9631386).

[15] H.N. Abdullah, T.R. Saeed, and A.H. Sahar, "Suboptimal Detection of Modified Logistic Map-based Chaos Shift Keying Modulation", *UPB Scientific Bulletin, Series C*, vol. 80, pp. 85–94, 2018.

[16] P. Robertson, E. Villebrun, and P. Hoeher, "A Comparison of Optimal and Sub-optimal MAP Decoding Algorithms Operating in the Log Domain", *IEEE International Conference on Communications ICC '95*, Seattle, USA, 1995 (https://doi.org/10.1109/ICC.1995.524253).

[17] H.N. Abdullah, T.R. Saeed, and A.H. Sahar, "Efficient Error Correcting Scheme for Chaos Shift Keying Signals", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, pp. 3550–3557, 2019 (http://doi.org/10.11591/ijece.v9i5.pp3550-3557).

[18] M. Wen, X. Cheng, and L. Yang, "Optimizing the Energy Efficiency of OFDM with Index Modulation", *2014 IEEE International Conference on Communication Systems*, Macau, China, 2014 (https://doi.org/10.1109/ICCS.2014.7024760).

[19] Y. Ko and J. Choi, "Sparse Multi-carrier Index Keying OFDM with Index Separation over Correlated Sub-carriers", *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, South Korea, 2015 (https://doi.org/10.1109/ICTC.2015.7354553).

Asaad H. Sahar, Ph.D.
https://orcid.org/0000-0003-3655-0162
E-mail: asaad.ha87@gmail.com
University of Baghdad, Baghdad, Iraq
https://en.uobaghdad.edu.iq

Aqiel N. Almamori, Ph.D.
https://orcid.org/0000-0001-5632-9989
University of Baghdad, Baghdad, Iraq
https://en.uobaghdad.edu.iq

Muhannad Y. Muhsin, Ph.D.
https://orcid.org/0000-0003-3937-4467
E-mail: muhannad.y.muhsin@uotechnology.edu.iq
University of Technology – Iraq, Bahdad, Iraq
https://uotechnology.edu.iq

Information for Authors

Journal of Telecommunications and Information Technology (JTIT) is published quarterly since 2000. 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.

We encourage submissions from a diverse range of authors from across all countries and backgrounds.

Manuscript

Latex files are preferred and Editorial Office provides a style to prepare the material along with the documentation. We also accept Microsoft Word and PDF files. A typical article is 10 pages long (approximately 6,000 words) and must include the following contents:

- Authors' names and affiliations in the following format:
  First name and surname (last name), academic title,
  Position held,
  ORCID number,
  E-mail address from the University's domain,
  Faculty and name of the University,
  Link to University website.
- Abstract (150-200 words). 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 related to the content of the article. About four keywords or phrases in alphabetical order should be used, separated by commas.
- The content of the article in a typical structure, i.e.: introduction, related work, conducted research, conclusions, references.

Figures, Tables and Photos

Together with the article, please send files with graphics with the highest resolution available, 150 dpi or more in bitmap resolution (jpg, png) and vector (cdr, svg, ps, pdf) formats are welcomed.

References

We use four main citation styles for a journal article, for an Internet article, for a conference paper, and for a book. Below are examples of citations. In each item, the DOI number or link to the PDF of the cited article should be provided.

[1] R.K. Meyers and A.H. Desoky, "An implementation of the blowfish cryptosystem", *2008 IEEE International Symposium on Signal Processing and Information Technology*, 2008 ((https://doi.org/10.1109/ISSPIT.2008.4775664).

[2] K. Nowicki and T. Uhl, *Ethernet End-to-End*, 1st ed. Germany, Shaker-Publisher, 2008 (ISBN: 978383832271404).

[3] C. Shorten and T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019 (https://doi.org/10.1186/s40537-019-0197-0).

[4] 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).

Submission

The paper with full PDF version and anonymous PDF version for the blind review process should be submitted on the JTIT website https://www.jtit.pl/jtit/about/submissions.

Reviewing Process

The article is initially approved by the Editor-In-Chief and if the decision is positive, is then sent to the reviewers. Depending on the subject of the article, it takes few weeks. In the next step, reviews are showed to authors who have 2 weeks to correct the article. Finally, the corrected text can be re-presented to the reviewer for reevaluation, which will take another 2 weeks.

As a result, after about 3 months, we are able to send the text for publication in the upcoming issue of JTIT.

When the reviews are inconsistent, additional corrections are necessary, or the reviewer expects additional verification because the corrections ordered by the author are insufficient or additional problems arise, the review of the article may be extended by another month or more.

Editorial Work

Positively reviewed and corrected article is next prepared by the editorial office for publication. At the end of this process the author receives an copyedited version for approval.

Licensing

Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript the author grants license to the National Institute of Telecommunications, for the use of the paper in the fields of exploitation: reproducing and fixing the paper, distributing the paper by means of introduction to trade, letting for use or rental of the original or copies, and distributing the paper by means of public exhibition, screening, presentation and broadcast as well as rebroadcast, and making the paper publicly available in such a manner that anyone could access it at a place and time selected thereby, or by making it available in a way not allowing selection of time or place, including by means of Internet or other networks.

Ghostwriting Declaration

We require formal declaration that the process of writing the paper was not influenced by any third party. In the article, all the contributions of other people are clearly indicated. The theories presented, methods used, analysis and research, as well as the copyrights to the drawings, photographs and other figures belong to the authors or are clearly credited in the text. The author must also indicate whether his work has received financial support and if the realization of the whole project was possible thanks to the permission and cooperation with scientific institutions, associations and others.

Other Information

- The JTIT being an Open Access Journal (OAJ) has no article processing charges (APCs). The published articles can be downloaded freely without payment.
- JTIT supports open access and using continuous publishing "publish-as-you-go" scheme. This means that we no longer wait to accumulate several articles into a quarterly issue before publication. Rather, articles are continuously added to current issues after acceptation. Publish-as-you-go reduces publication lag for our authors, and make the newest research available quickly. After completing the review process, an article is published online in the current issue with DOI registration. When the issue period ends, a new issue is activated. So accepted articles are published without waiting for the quarterly issue end.

National Institute of Telecommunications

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland
https://www.gov.pl/web/instytut-lacznosci

phone +48 22 512 81 83
fax +48 22 512 84 00

e-mail: journal@jtit.pl
www.jtit.pl

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 1/2026 vol.103