

# JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

3 / 2023

## Coverage Improvements for Sub-Terahertz Systems Under Shadowing Conditions

*W. Mohr*

*Paper*

1

## Comparative Analysis of NOMA and OMA Schemes: GSVD-based NOMA Systems and the Role of Mobile Edge Computing

*Y. Dursun*

*Paper*

11

## Performance Comparison of Several Algorithms for Localization of Wideband Sources

*H. Ougraz, S. Safi, A. Boumezzough, and M. Frikel*

*Paper*

21

## Implementation of a 4G/5G Base Station Using the srsRAN Software and the USRP Software Radio Module

*W. Flakowski, M. Krasicki, and R. Krenz*

*Paper*

30

## An Efficient Cooperative Spectrum Sensing Method Using Renyi Entropy Weighted Optimal Likelihood Ratio for CRN

*S.B. Shrote and S.D. Poshattiwar*

*Paper*

41

## Location-based Power Control Mechanism for D2D Communication Underlying a Cellular System

*M. Rodziejewicz*

*Paper*

49

## A Cross-layer Method for Scheduling and Routing Real-time Traffic Flow in Industrial IoT

*N. Azimian, R. Mohammadi, and M. Nassiri*

*Paper*

54

## Characterization of a Split Circle Element for Microstrip Reflectarrays

*K.H. Sayidmarie and L. S. Yahya*

*Paper*

62

(Contents continued on back cover)

## ***Editor-in-Chief***

**Adrian Kliks**, Poznan University of Technology, Poland

## ***Steering Editor***

**Paweł Pławiak**, National Institute of Telecommunications, Poland

## ***Editorial Advisory Board***

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

**Naveen Chilamkurti**, LaTrobe University, Australia

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

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

**Nikolaos Dimitriou**, NCSR “Demokritos” Athens, Greece

**Ciprian Dobre**, Politechnic University of Bucharest, Romania

**Filip Idzikowski**, Poznan University of Technology, Poland

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

**Albert Levi**, Sabanci University, Turkey

**Marian Marciniak**, National Institute of Telecommunications, Poland

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

**Constandinos Mavromoustakis**, University of Nicosia, Cyprus

**Klaus Mößner**, Technische Universität Chemnitz, Germany

**Imran Muhammad**, King Saud University, Saudi Arabia

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

**Daniel Negru**, University of Bordeaux, France

**Ewa Orłowska**, National Institute of Telecommunications, Poland

**Jordi Perez-Romero**, UPC, Spain

**Michał Pióro**, Warsaw University of Technology, Poland

**Konstantinos Psannis**, University of Macedonia, Greece

**Salvatore Signorello**, University of Lisboa, Portugal

**Adam Wolisz**, Technische Universität Berlin, Germany

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

## ***Editorial Team***

Content Editor: **Robert Magdziak**

Managing Editor: **Ewa Kapuściarek**

eISSN 1899-8852

© Copyright by National Institute of Telecommunications, Poland 2023

# Coverage Improvements for Sub-Terahertz Systems Under Shadowing Conditions

Werner Mohr

*Independent consultant, Munich, Germany*

<https://doi.org/10.26636/jtit.2023.3.1301>

**Abstract** — Radio propagation in the millimeter wave and sub-terahertz domain is heavily affected by shadowing conditions. The communication link is blocked without any additional technical means being used. Coverage improvements can be provided by using reflectors, RIS arrays, and repeaters to direct radio waves around corners or obstacles. These concepts show different performance and complexity levels affecting their network deployment. This paper investigates the achievable radio range or the received power to compare specific deployment concepts under realistic propagation conditions. Overall, the repeater solution provides either the largest radio range or the lowest necessary total transmit power compared to reflectors or RIS arrays and, thereby, is the most sustainable approach. A RIS array requires an additional centralized signal processing capacity for calculating optimized RIS settings and results in the highest level of network deployment complexity.

**Keywords** — *multipath propagation, radio range, received power, reconfigurable intelligent surface (RIS), shadowing*

## 1. Introduction

Sub-terahertz (sub-THz) mobile communication systems are currently discussed within the research community and industry as one of the technical components required for the development of sixth generation mobile networks (6G), with a view of enabling very wideband radio systems and high aggregated carrier throughput rates in the order of several hundred gigabits per second or up to a terabit per second (e.g., [1]). The wide carrier bandwidth can only be provided at very high frequencies in the sub-THz domain. However, with the increasing carrier frequency of the radio system, the propagation conditions are affected by additional atmospheric, as well as rain- and foliage-generated attenuation [2] compared with lower frequency bands. Distant-dependent pathloss and multipath propagation are encountered as well, resulting in fading (e.g., Rayleigh or Rice) and shadowing. Pathloss is particularly high in the millimeter wave (mmWave) and sub-THz domains, and the communication link is interrupted by obstacles or shadowing conditions created, for instance, by walls and buildings [2]–[5]. Therefore, technical means are needed to improve coverage and enable communication around corners in urban environments or indoor scenarios.

This paper investigates and compares the concepts of simple metallic reflectors (mirrors), reconfigurable intelligent surfaces (RIS), and repeaters to understand the performance, complexity, benefits and drawbacks for a practical network

deployment. Section 2 describes the shadowing scenario; Section 3 deals with the received power required by these technical means in order to overcome shadowing, while Section 4 compares the performance of the different concepts. Results are summarized in Section 5.

## 2. Shadowing Propagation Scenario

In [2], terahertz systems are investigated under idealized line-of-sight (LOS) conditions to obtain a basic understanding of the achievable range and channel capacity with the additional condition of legal radiation limits to avoid health effects.

Depending on the propagation conditions, and with additional atmospheric pathloss as well as potential rain- and/or foliage-generated attenuation taken into consideration, the range is significantly limited as the frequency increases, and channel capacity is decreasing at a rapid pace [2]. These additional attenuations need to be taken into account in the higher frequency bands. Furthermore, the intended much wider system carrier bandwidths or throughput rates achieved in the sub-THz domain contribute to reducing the achievable range even further.

The transmit power or EIRP is fixed in the following scenario to meet the radiation limit of 10 W/m<sup>2</sup> (at a reference distance), so as to ensure compliance with legal radiation limits [6], [7]. Radio range limitations resulting from radiation limits may be overcome by higher locations at which access point or base station antennas are installed [2]. However, this does not overcome the very high pathloss encountered due to shadowing.

### 2.1. Propagation Measurements and Models

Pathloss measurements and models existing in the sub-THz domain are investigated, for instance, in [8], [9], and [10]. In [8], the LOS model including additional attenuation for atmospheric pathloss is confirmed for the terahertz domain. Reflections from rough surfaces, such as buildings, are also considered, as they result in additional losses. Scattering increases along with the roughness of the reflecting surface. In the sub-THz domain, the roughness of surfaces existing in the environment, such as walls, is in the order of the wavelength ([3] p. H6 and [5] pp. 26). Paper [9] confirms basically the LOS model with a slightly higher pathloss decay factor of  $n > 2$ . Fading statistics are investigated in [10]. For different

indoor scenarios the fading statistics are modelled by  $\alpha - \mu$ , Rice, Nakagami- $m$  and log-normal distributions. The best fit was achieved with the use of the  $\alpha - \mu$  distribution. The Rice distribution, which is considered in this paper, is another reasonable approach.

## 2.2. Shadowing Propagation Scenario

Shadowing results in very high pathloss preventing data transmission in the mmWave and sub-THz domains. In addition, compared to low frequency ranges, additional attenuation components have to be considered, such as atmospheric pathloss, rain- and foliage-generated attenuation, as well as reflections and scattering depending on the propagation scenario [2]. Shadowing results in statistical fluctuations of the received power (according to a log-normal distribution) and, therefore, in a significantly higher pathloss  $L_{ln}$  compared to free space propagation [2], [4] p. 104. These fluctuations correspond to slow fading along the propagation path with a correlation length of several carrier wavelengths  $\lambda$  at low frequencies and many wavelengths in the mmWave and sub-THz domains.

$$L_{ln}(r) = \overline{L_{ln}(r_0)} + 10 n \log \frac{r}{r_0} + X_\sigma + \frac{L_r(f, r_0 = 1 \text{ km})}{1 \text{ km}} r + \frac{L_{at}(f, r_0 = 1 \text{ km})}{1 \text{ km}} r + L_{fol}(f, r) \text{ [dB]} \quad (1)$$

where:

$L_{ln}$  – propagation loss for log-normal fading [dB],

$r$  – distance [m],

$r_0$  – reference distance [m],

$\overline{L_{ln}(r_0)}$  – average pathloss at reference distance [dB],

$n$  – pathloss coefficient depending on the environment,

$X_\sigma$  – zero mean Gaussian process [dB] with a standard deviation of  $\sigma = 4$  to 10 dB, describing random shadowing depending on the environment,

$L_r$  – rain attenuation [11] pp. 91, [12]–[14] pp. 2–68, [15],

$L_{at}$  – atmospheric attenuation [3] p. H12, [11], [15],

$L_{fol}$  – foliage attenuation [16].

Equation (1) is valid for a constant carrier frequency. However, in the sub-THz domain, radio propagation under shadowing conditions results in much higher pathloss than for free space propagation depending on the radio environment and, thereby, in a blocked communication link. Therefore, technical system deployment means are needed which provide quasi-LOS conditions to increase the range under shadowing conditions while simultaneously complying with legal radiation limits.

## 3. Pathloss with a Single Reflection and Potential Secondary Reflections Around the Receiver

The technical means deployed between the transmitter and the receiver include artificial metallic reflectors, RIS or repeaters. Additional reflections and scattering caused by the environment, e.g., building or obstacles, are also taken into

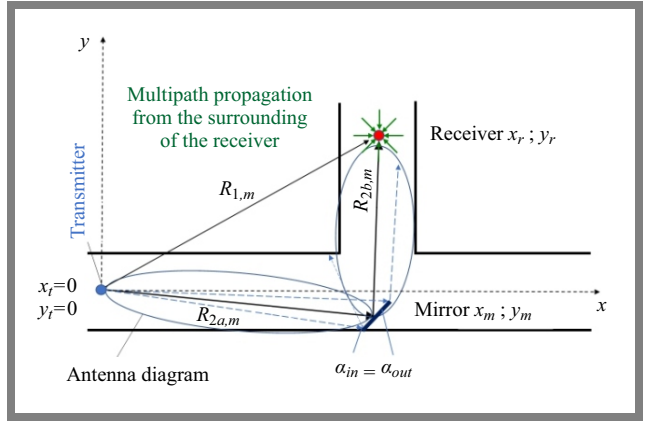


**Fig. 1.** Propagation scenario for indoor environments or street canyons, with the receiver not being visible from the transmitter and a reflector being used to ensure coverage.

consideration, as these direct electromagnetic waves around corners. In the following sections, these three technical means are investigated. For simplification purposes, only free space propagation is taken into consideration, without any atmospheric, rain- and foliage-generated attenuation.

### 3.1. Reflector Much Bigger Than the Wavelength

A metallic reflector with good conductivity and a physical size of  $\gg \lambda$  is assumed to mitigate shadowing effects. The high conductivity of the reflector ensures that the field resistance of the metal is much lower than the field resistance of the free space. Therefore, the reflection coefficient complies with [17] pp. 134 for horizontal and vertical polarization:

$$\varrho_m \rightarrow -1. \quad (2)$$

The angle of the outgoing wave  $\alpha_{out}$  is the same as that of the incoming wave  $\alpha_{in}$ . Figure 1 shows the propagation scenario with a corner, as experienced in indoor environments or street canyons. The receiver is obstructed and is not directly visible from the transmitter. Secondary reflections from walls around the receiver are experienced as well, where  $|\varrho| \ll 1$  may be assumed. Therefore, the path reflected by the reflector is much stronger than the secondary reflections (similarly as in a Rice channel). The direct path  $R_{1,m}$  is obstructed, for instance by buildings, and the effective length of the reflected propagation path at the reflector (as shown in Fig. 1) is given by:

$$R_{2m} = R_{2a,m} + R_{2b,m}, \quad (3)$$

where the reflected path is mirrored at the reflector. Communication at mmWave or sub-THz frequency ranges is only feasible in such scenarios via the reflected path.

If the incoming wave at the reflector is a plane wave, the reflected wave is a plane wave as well – see [5] pp. 24, [17] pp. 134. Radiation density at reflector  $S_{2a}$  and field strength  $E_{2a}$  are given by the Friis formula with the field resistance of  $Z_0 = 120\Omega$  [5] p. 18, where  $EIRP_T = P_T G_T$  is limited by applicable legal radiation limits (e.g., [2]):

$$S_{2a} = P_T G_T \frac{1}{4\pi R_{2a,m}^2} = EIRP_T \frac{1}{4\pi R_{2a,m}^2} = \frac{E_{2a}^2}{Z_0}. \quad (4)$$

The reflected field strength  $E'_{2a}$  and radiation density  $S'_{2a}$  result from the reflection coefficient used in Eq. (2):

$$E'_{2a} = \varrho_m e^{-j\Delta\phi} E_{2a}, \quad (5)$$

$$\begin{aligned} S'_{2a} &= |\varrho_m e^{-j\Delta\phi}|^2 P_T G_T \frac{1}{4\pi R_{2a,m}^2} \\ &= \frac{E_{2a}^2}{Z_0} = |\varrho_m e^{-j\Delta\phi}|^2 \frac{E_{2a}^2}{Z_0}. \end{aligned} \quad (6)$$

For  $\varrho_m < |-1|$ , the reflector may add attenuation to the propagation loss. This results in a specific relation between the received and transmitted powers, such as:

$$\begin{aligned} \frac{P_R}{P_T} \Big|_{\text{reflector}} &= |\varrho_m e^{-j\Delta\phi}|^2 G_T \frac{1}{4\pi R_{2,m}^2} A_R \\ &= |\varrho_m e^{-j\Delta\phi}|^2 G_T \frac{1}{4\pi (R_{2a,m} + R_{2b,m})^2} A_R \quad (7) \\ &= |\varrho_m e^{-j\Delta\phi}|^2 G_T G_R \left( \frac{\lambda}{4\pi R_{2,m}} \right)^2 \frac{1}{(R_{2a,m} + R_{2b,m})^2}. \end{aligned}$$

or with Eq. (2):

$$P_R \Big|_{\text{reflector}} \approx EIRP_T G_R \left( \frac{\lambda}{4\pi (R_{2a,m} + R_{2b,m})} \right)^2. \quad (8)$$

The received power depends on the square of the path length  $R_{2m} = R_{2a,m} + R_{2b,m}$ , the absolute value of the reflection coefficient  $\varrho_m$ , and the transmitter and receiver antenna gain. In the case of a metallic reflector,  $\varrho_m = |1|$ , is a realistic estimate. Figure 1 also indicates that secondary reflections around the receiver provide additional multipath propagation. A metallic reflector is characterized by the following implementation-critical features:

- it is passive device,
- no power supply or data connections are needed,
- it is only mounted at buildings, towers or lamp posts,
- it is a low complexity solution,
- the reflector may have the form of a plane, may be concave or convex, with a fixed, static beam shaped to suit the area to be covered,
- the reflector improves coverage, especially in shaded areas,
- no latency is added to the signal's propagation,
- for high conductivity of the reflector  $\varrho_m \approx |1|$ , nearly no insertion loss is added,
- the propagation corresponds to quasi-LOS propagation conditions around corners,
- it ensures a much longer range compared with direct propagation with high shadowing pathloss,
- LOS propagation provides a strong direct component and thereby establishes a Rice channel at the receiver,
- without the reflector, the received power would be much lower and the channel would correspond to a Rayleigh channel,
- at the receiver, delay spread is generated via secondary reflections around the receiver or from other reflecting obstacles present in the deployment area, with smaller reflection coefficients.

### 3.2. Reflector Smaller or in the Same Order as Wavelength and RIS Arrays

A RIS array comprises many small RIS elements with spacing between them equaling approx.  $\lambda/2$  of the medium carrier frequency of the frequency range in use. As in the antenna theory for array antennas with a group of antenna elements, it is assumed that the diagrams of the different elements are independent from each other ([18] pp. 254) to apply the superposition principle.

The radar cross section  $\sigma$  of an RIS element is modeled in this paper by a metallic sphere with a radius  $r$  as an idealized approach [19]–[22]:

$$\begin{aligned} \frac{\sigma}{\pi r^2} &= \frac{1}{kr} \sum_{n=1}^{\infty} (-1)^n (2n+1) \\ &\cdot \left[ \left( \frac{kr J_n(kr) - n J_n(kr)}{kr H_{n-1}(kr) - n H_n^1(kr)} \right) - \left( \frac{J_n(kr)}{H_n^1(kr)} \right) \right], \end{aligned} \quad (9)$$

where:  $r$  is the radius of the sphere,  $k = \frac{2\pi}{\lambda}$ ,  $\lambda$  is the wavelength,  $J_n$  is the spherical Bessel function of the first kind of order  $n$ ,  $H_n^1$  is the Hankel function of order  $n$ , described by  $H_n^1(kr) = J_n(kr) + j Y_n(kr)$ , and  $Y_n$  is spherical Bessel function of the second kind of order  $n$ .

A RIS inherent insertion loss [23] is not considered here.

Paper [24] indicates that a signal that is reflected or scattered in other directions than the incoming wave has a lower field strength than the reflection back to the source. Therefore, the radar cross section used below is an optimistic approach to reflections in other directions than that of the incoming wave. For a spacing of  $\lambda/2$  between RIS elements, radius  $r$  of the metallic sphere and its normalized size equal:

$$r = \frac{\lambda}{4}, \quad \frac{2\pi r}{\lambda} = \frac{2\pi \frac{\lambda}{4}}{\lambda} = \frac{\pi}{2} > 1 \quad (10)$$

Based on Eqs. (9) and (10), the normalized radar cross section is approximately:

$$\frac{\sigma}{\pi r^2} = 1 \quad (11)$$

Reflections from or scattering caused by small objects with a size  $< \lambda$  result in spherical waves. For a single sphere, the ratio between the received and transmitted power is given in [3], p. S1 as well as Eqs. (10) and (11) from the radar equation:

$$\begin{aligned} \frac{P_R}{P_T} \Big|_{\text{scatter,sphere}} &= \frac{G_T}{4\pi R_{2a,s}^2} \sigma \frac{A_e}{4\pi R_{2b,s}^2} \\ &= G_T G_R \pi \left( \frac{\lambda}{4} \right)^2 \frac{1}{(4\pi)^3} \frac{\lambda^2}{(R_{2a,s} R_{2b,s})^2} \end{aligned} \quad (12)$$

or by considering the limited  $EIRP_T = P_T G_T$  at the base station with respect to the legal radiation limits (e.g., [2]):

$$\begin{aligned} P_R \Big|_{\text{scatter,sphere}} &= EIRP_T G_R \pi \left( \frac{\lambda}{4} \right)^2 \\ \frac{1}{(4\pi)^3} \frac{\lambda^2}{(R_{2a,s} \cdot R_{2b,s})^2} &\propto \frac{1}{(R_{2a,s} \cdot R_{2b,s})^2}. \end{aligned} \quad (13)$$

The pathloss is proportional to the product of the square of the two partial distances  $R_{2a,s}$  and  $R_{2b,s}$ .

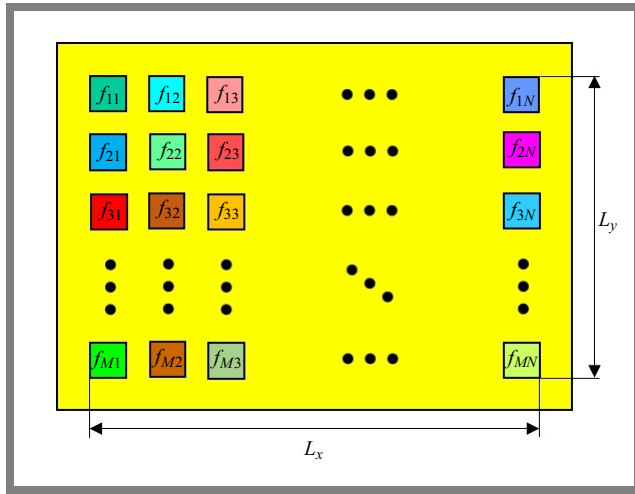


**Fig. 2.** RIS array made up of  $M \cdot N$  RIS elements.

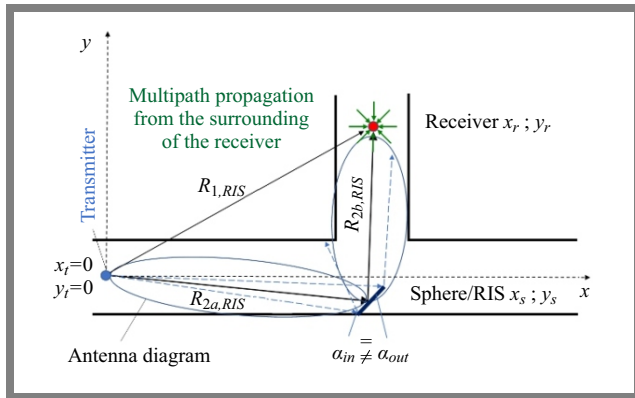


**Fig. 3.** Propagation scenario for indoor environments or street canyons, with the receiver not being visible from the transmitter and a sphere or an RIS array being used to ensure coverage.

A RIS array comprises many small reflectors ( $M \cdot N$ ) shown in Fig. 2. For simplicity, it is assumed that the different RIS elements do not influence each other. Then, the solution may be regarded as a phased array antenna which is illuminated by the base station or a user equipment antenna. The radar cross section of a single sphere is extended by the physical size of the RIS array  $(M \cdot N)(\frac{\lambda}{2})^2$ . A RIS array can be used for beamforming with the additional beamforming gain of  $(M \cdot N)$ , where compared to Subsection 2.1, the outgoing angle  $\alpha_{out}$  is not necessarily the same as the incoming angle  $\alpha_{in}$ . A propagation scenario similar to that shown in Fig. 1 is considered in Fig. 3. In the scenario in which phase shifts of the different RIS elements are random, the sum of the powers of the different reflected signals per RIS element is received. Then Eq. (12) is extended as in Eq. (14):

$$\frac{P_R}{P_T} \Big|_{RIS} = \frac{G_T}{4\pi R_{2a,RIS}^2} M \cdot N \sigma \frac{A_e}{4\pi R_{2b,RIS}^2} \\ = G_T G_R M \cdot N \pi \left( \frac{\lambda}{4} \right)^2 \frac{1}{(4\pi)^3} \frac{\lambda^2}{(R_{2a,RIS} R_{2b,RIS})^2} \cdot (14)$$

By relying on appropriate phase shifts, each element in the RIS array may be used for beamforming. The constructive superposition of each reflected signal provides a main lobe of the resulting beam from the RIS array towards the receiver

antenna. Under ideal conditions, the receiver antenna is summing up the field strength of each signal. This corresponds to a group of antennas with many individual elements ([18] pp. 254). With the multiplicative approach, where the directivity of an individual antenna element is multiplied with the characteristic of the array and all antenna elements have the same characteristic and same distance  $\lambda/2$  between the antenna elements like in Fig. 2, the maximum ratio between  $P_R$  and  $P_T$  follows with the array gain  $M \cdot N$ :

$$\frac{P_R}{P_T} \Big|_{RIS,max} = \frac{G_T}{4\pi R_{2a,RIS}^2} (M \cdot N)^2 \sigma \frac{A_e}{4\pi R_{2b,RIS}^2} \\ = G_T G_R (M \cdot N)^2 \pi \left( \frac{\lambda}{4} \right)^2 \frac{1}{(4\pi)^3} \frac{\lambda^2}{(R_{2a,RIS} R_{2b,RIS})^2} (15)$$

or

$$P_R \Big|_{RIS,max} = EIRP_T G_R (M \cdot N)^2 \pi \left( \frac{\lambda}{4} \right)^2 \frac{1}{(4\pi)^3} \\ \cdot \frac{\lambda^2}{(R_{2a,RIS} R_{2b,RIS})^2} \propto \frac{1}{(R_{2a,RIS} R_{2b,RIS})^2} \cdot (16)$$

The received power  $P_R$  is growing with the square of the RIS array size  $M \cdot N$ . With a growing number of RIS elements, the received power is increasing.

Pathloss is proportional to the product of the square of the two partial distances  $R_{2a,RIS}$  and  $R_{2b,RIS}$ .  $EIRP_T$  is limited by applicable radiation limits.  $P_R$  in Eqs. (13) or (16) corresponds to the direct reflected path. Figure 3 also indicates that secondary reflections around the receiver provide additional multipath propagation, where a reflection coefficient  $|\rho| \ll 1$  can be assumed.

A RIS array is characterized by the following technical implementation features:

- it is basically a passive device without amplification and without signal regeneration capabilities,
- by using appropriate phase shifters for different RIS elements, some beamforming is possible,
- this requires that signal processing performed by user equipment, the base station and the RIS array be centralized [25]. In this sense, it is an active system,
- a power supply is needed,
- a data connection from/to the central signal processing unit at the base station and the RIS feedback to the base station/signal processing unit is required [25],
- it is mounted on buildings, towers or lamp posts,
- with the power supply and an additional data connection, this is a very complex solution,
- no latency is added to the signal's propagation, because RIS elements only provide a phase shift,
- RIS improves coverage, especially in shaded areas,
- propagation corresponds to quasi-LOS propagation conditions around corners,
- it offers a longer range compared to direct propagation with the high shadowing pathloss,
- LOS propagation provides a strong direct component and thereby a Rice channel is established at the receiver,

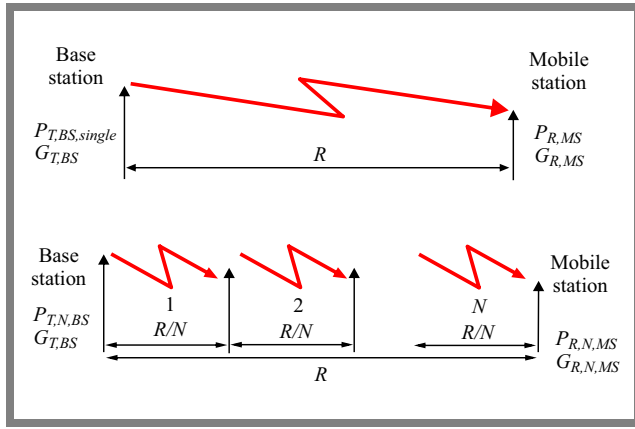


**Fig. 4.** Operation principle of a repeater-based or multi hop solution: a) single hop approach, b) multi hop or repeater-based approach.

- without a RIS array, the received power would be much lower and the channel would correspond to a Rayleigh channel,
- the delay spread is generated, at the receiver, via secondary reflections around the receiver or from other reflecting obstacles present in the deployment area and characterized by smaller reflection coefficients,
- for a specific user location, the radio channel can be influenced or adjusted. However, RIS settings are strongly dependent on the receiver's location,
- if RIS is used for beamforming, it operates in a manner that is similar to a directive antenna,
- with an increased number of RIS elements  $M \cdot N$  the RIS array gain and its physical size increase as well, in practice, the received power will be lower due to the inherent RIS insertion loss.

### 3.3. Repeater

In a solution with a repeater, the entire propagation path  $R$  is subdivided into shorter propagation paths (Fig. 4). In this simplified example, all transmit powers and antenna gains as well as the  $N$  radio path lengths are the same per hop. This means that the transmit power at the base station can be reduced for the same effective path length, as described in Subsection 3.1 and 3.2, and/or such an approach allows to transmit around corners under shadowing conditions, while achieving longer range. In the following, free space propagation per hop is assumed. In the single hop case, the received power  $P_R$  follows from the link budget in Eq. (17) and the necessary transmit power  $P_{T,BS,single}$  from Eq. (18):

$$P_R = P_{T,BS,single} G_{T,BS} G_{R,MS} \left( \frac{\lambda}{4\pi R} \right)^2, \quad (17)$$

$$P_{T,BS,single} = \frac{P_R}{G_{T,BS} G_{R,MS}} \left( \frac{4\pi R}{\lambda} \right)^2. \quad (18)$$

In the multi hop or repeater-based case, it is assumed that the received power  $P_R$  at the destination is the same as in the single hop case. Then, the necessary base station and repeater transmit powers  $P_{T,BS,multi\ hop}$  and  $P_{T,MS,N}$  are assumed

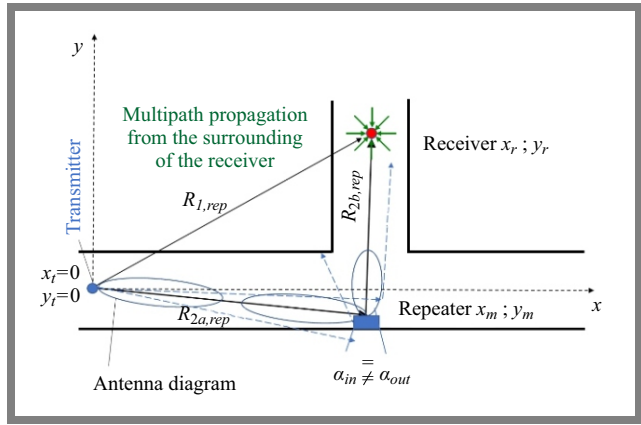


**Fig. 5.** Propagation scenario for indoor environments or street canyons with the receiver not visible from the transmitter's location and a repeater used to ensure coverage.

to be the same:

$$\begin{aligned} P_{T,total} &= N P_T = N \frac{P_R}{G_T G_R} \left( \frac{4\pi R}{\lambda N} \right)^2 \\ &= \frac{1}{N} \frac{P_R}{G_T G_R} \left( \frac{4\pi R}{\lambda} \right)^2 = \frac{P_{T,BS,single}}{N}. \end{aligned} \quad (19)$$

It follows from Eq. (19) that in the multi hop approach, the total needed transmit power is lower than that in the single hop scenario, and for the same distance  $R$  needs to be covered. It is more challenging, especially at very high frequencies, to generate high RF power levels. Therefore, the multi-hop solution makes millimeter wave and terahertz systems more suitable and allows to transmit around corners with lower energy consumption.

In the repeater-based solution, a propagation solution comparable with that shown in Fig. 1 and Fig. 3 is assumed (and shown Fig. 5). The repeater has an antenna looking towards the base station/access point and a second antenna pointing towards the intended coverage area which is shaded from the base station. In addition, the repeater provides an amplification from the received to the transmit signal and may perform signal regeneration and beamforming or MIMO transmission. The outgoing angle  $\alpha_{out}$  is not necessarily the same as the incoming angle  $\alpha_{in}$ . The link between the base station/access point and the repeater as well as the link between the repeater and the user equipment is given by the Friis formula [5], p. 18 and the repeater power amplification  $V$ , such as:

$$\begin{aligned} \frac{P_{R,repeater}}{P_{T,base\ station}} &= \\ &= G_{T,base\ station} G_{R,repeater} \left( \frac{\lambda}{4\pi R_{2a,rep}} \right)^2 \\ \frac{P_{R,user\ equipment}}{P_{T,repeater}} &= \frac{P_{R,user\ equipment}}{V \cdot P_{R,repeater}} \\ &= G_{T,repeater} G_{R,user\ equipment} \left( \frac{\lambda}{4\pi R_{2b,rep}} \right)^2. \end{aligned} \quad (20)$$

In the repeater-based solution, it is assumed that the base station and the repeater can generate the maximum  $EIRP_T$  values allowed under applicable radiation limits. It can also be

assumed that the minimum necessary received power  $P_R$  or receiver sensitivity are the same for the base station, repeater and user equipment. Due to the amplification performed in the repeater, the received power at the user equipment depends solely on the  $EIRP_T$  of the repeater and the distance  $R_{2b,rep}$ .

The amplification  $V$  of the repeater is limited, so that the maximum allowed  $EIRP_T$  based on radiation limits is complied with:

$$P_{T, \text{repeater}} G_{T, \text{repeater}} = V P_{R, \text{repeater}} G_{T, \text{repeater}} = EIRP_T, \quad (21)$$

$$V = \frac{P_{T, \text{repeater}}}{P_{R, \text{repeater}}} = \frac{EIRP_T}{P_{R, \text{repeater}} G_{T, \text{repeater}}}, \quad (22)$$

which provides the received power at the user equipment:

$$P_{R, \text{user equipment}} = EIRP_T G_{R, \text{user equipment}} \left( \frac{\lambda}{4\pi R_{2b,rep}} \right)^2. \quad (23)$$

Equation (23) provides the received power at the user equipment. The total maximum transmit power necessary to cover the entire path length  $R_{2,rep} = R_{2a,rep} + R_{2b,rep}$  is given as:

$$P_{T, \text{base station}} + P_{T, \text{repeater}} = \frac{EIRP_T}{G_{T, \text{base station}}} + \frac{EIRP_T}{G_{T, \text{repeater}}}. \quad (24)$$

Each of the two hops can cover, with the same  $EIRP_T$ , the same distance as achieved in the reflector-based solution from Subsection 2.1. Therefore, either twice the effective distance is feasible or the transmit power used at the base station and the repeater can be reduced.

The received power  $P_R$  from Eq. (23) corresponds to the direct reflected path. Figure 5 also indicates that secondary reflections around the receiver provide additional multipath propagation.

The repeater-based approach is characterized by the following technical implementation-related features:

- it is an active device with amplification and potentially with signal regeneration and beamforming or a MIMO transmission,
- there are basically two repeater options available.

In the first scenario an RF repeater is used only, and in the other the repeater regenerates the received signal by means of signal processing and beamforming or a MIMO transmission. In the first case RF amplification is applied only to achieve the maximum allowed  $EIRP_T$ , potentially with a shift to a different carrier frequency to avoid oscillation of the repeater due to positive feedback from the repeater's output, affecting its input. Repeater amplification and antenna gain can either increase the covered range compared to a reflector or a RIS, or may allow to reduce the overall RF power needed. This approach adds almost no latency, if analogue signal processing or amplification is applied only. The solution offers a medium complexity level.

In the second case, the repeater amplifies the signal to the

maximum allowed  $EIRP_T$ . Advanced antenna systems, such as those relying on beamforming or MIMO concepts, can be used and the repeater's amplification and the antenna gain may either increase the covered range compared to a reflector or a RIS or allow to reduce the overall RF power required.

No additional feedback from the user equipment or the base station to the repeater is needed, as is the case on the RIS approach. The channel estimation capabilities of the radio interface are sufficient to implement such antenna concepts. However, signal processing for signal regeneration and antenna algorithms is adding latency. It is a solution with a medium level of complexity, since the same microelectronic components as those used in base stations or user equipment may be applied:

- only a power supply is needed,
- it is mounted on buildings, towers or lamp posts,
- overall, its complexity level is between that of a metallic reflector and a RIS system,
- a repeater is more flexible than a metallic reflector in directing energy at specific users and a RIS by using advanced antenna concepts,
- a repeater is improving coverage significantly, especially in shaded areas,
- the propagation pattern corresponds to LOS propagation conditions between the base station and the repeater and a second LOS propagation between the repeater and user equipment to direct the signal around corners,
- this allows a bigger range compared to direct propagation with a high shadowing pathloss,
- LOS propagation provides a strong direct component and thereby a Rice channel is established at the receiver,
- without a repeater, the received power would be much lower and the channel would correspond to a Rayleigh channel,
- delay spread is generated via secondary reflections around the receiver or other reflecting obstacles in the deployment area characterized by smaller reflection coefficients.

## 4. Comparison of reflector, RIS array and repeater

The different solutions are compared under the same conditions. All concepts are using:

- the same maximum  $EIRP_T$  at the base station and also at the repeater,
- equal receiver antenna gains at user equipment,
- the same necessary received power at user equipment and at the repeater and
- the same carrier frequency or wavelength.

The size of the RIS array  $M \cdot N$  is variable.

Based on the formulas for the received power at user equipment for the three system concepts involving a reflector Eq. (8), a RIS array for optimal beamforming Eq. (16) and a repeater Eq. (23) the achievable received power for the same

radio path length is investigated. For comparison, the results are normalized to the reflector case. This results in the following normalized functions:

$$\frac{P_R|_{RIS,max}}{P_R|_{reflector}} = \frac{(M \cdot N)^2 \pi \left(\frac{\lambda}{4}\right)^2 \frac{1}{(4\pi)^3} \frac{\lambda^2}{(R_{2a,RIS} R_{2b,RIS})^2}}{\left(\frac{\lambda}{4\pi(R_{2a,m} + R_{2b,m})}\right)^2} = (M \cdot N)^2 \frac{1}{4^3} \frac{\lambda^4}{(R_{2a,RIS} R_{2b,RIS})^2} \frac{(R_{2a,m} + R_{2b,m})^2}{\lambda^2}. \quad (25)$$

$$\frac{P_R|_{repeater}}{P_R|_{reflector}} = \frac{\left(\frac{\lambda}{4\pi R_{2b,rep}}\right)^2}{\left(\frac{\lambda}{4\pi(R_{2a,m} + R_{2b,m})}\right)^2} = \frac{(R_{2a,m} + R_{2b,m})^2}{R_{2b,rep}^2}. \quad (26)$$

If the same geometry and distances are assumed with:

$$R_{2a,m} = R_{2a,RIS} = R_{2a,rep} = R_{2a}, \quad (27)$$

$$R_{2b,m} = R_{2b,RIS} = R_{2b,rep} = R_{2b}. \quad (28)$$

Eqs. (25) and (26) result in:

$$\frac{P_R|_{RIS,max}}{P_R|_{reflector}} = (M \cdot N)^2 \frac{1}{4^3} \frac{\lambda^4}{(R_{2a} R_{2b})^2} \frac{(R_{2a} + R_{2b})^2}{\lambda^2}, \quad (29)$$

$$\frac{P_R|_{repeater}}{P_R|_{reflector}} = \frac{(R_{2a} + R_{2b})^2}{R_{2b}^2} = \frac{R_2^2}{R_{2b}^2}. \quad (30)$$

#### 4.1. Received Power Ratio for the RIS Array vs. Reflector Approach

With the normalized distant-dependent terms  $R_{2a}/\lambda$ ,  $R_{2b}/\lambda$  and the total path length  $R_2/\lambda = R_{2a}/\lambda + R_{2b}/\lambda$  it follows for Eq. (29):

$$\frac{P_R|_{RIS,max}}{P_R|_{reflector}} = (M \cdot N)^2 \frac{1}{4^3} \frac{\lambda^2}{R_2^2} \frac{1}{\left(\frac{R_{2b}}{R_2}\right)^2 \left(1 - \frac{R_{2b}}{R_2}\right)^2}, \quad (31)$$

with  $0 \leq R_{2b}/R_2 \leq 1$ . As the RIS array will not be placed directly at the base station location ( $R_{2b}/R_2 = 1$ ) or at the location of user equipment ( $R_{2b}/R_2 = 0$ ) with poles in Eq. (31), a reasonable assumption for the RIS array location is somewhere in the middle of the overall propagation path within the following limits

$$0.2 \leq \frac{R_{2b}}{R_2} \leq 0.8. \quad (32)$$

The minimum size of the RIS array  $M \cdot N|_{min}$  for the same received power as for the reflector follows from Eq. (31):

$$\frac{P_R|_{RIS,max}}{P_R|_{reflector}} = 1, \quad (33)$$

$$M \cdot N|_{min} = 8 \frac{R_2}{\lambda} \frac{R_{2b}}{R_2} \left(1 - \frac{R_{2b}}{R_2}\right). \quad (34)$$

Due to the huge variation versus carrier frequency  $f$  (1, 3, 10, 30, 100, and 300 GHz) and entire path length  $R_2$  (1, 10, 100, and 200 m) the evaluation of Eq. (34) is presented logarithmically as  $\log(M \cdot N|_{min})$  in Fig. 6 versus the relative location  $R_{2b}/R_2$  of the reflector or the RIS array. The minimum RIS array size increases significantly with the carrier

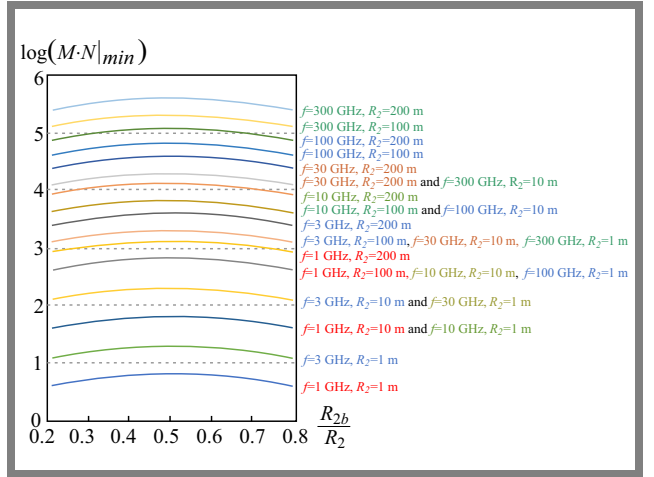


**Fig. 6.** Summary diagram of the minimum size of the RIS array  $M \cdot N|_{min}$  Eq. (34) for optimum phase shifts per RIS element versus carrier frequency  $f$ , entire path length  $R_2$  versus the relative location  $R_{2b}/R_2$  of the reflector or RIS array.

frequency and the entire path length. Especially, in the millimeter or sub-Terahertz frequency range, the number of RIS elements is in the order of several tens of thousands and more. Here, it is assumed that the maximum gain  $M \cdot N$  of the RIS array is available with appropriate phase shifts of the RIS elements.

If there is no RIS gain  $M \cdot N = 1$  in the case of random phase shifts per RIS element, the minimum number of RIS elements needs to grow by the square of  $M \cdot N|_{min}$  in Eq. (34):

$$\frac{P_R|_{RIS,no \text{ RIS gain}}}{P_R|_{reflector}} = M \cdot N \frac{1}{4^3} \frac{\lambda^2}{R_2^2} \frac{1}{\left(\frac{R_{2b}}{R_2}\right)^2 \left(1 - \frac{R_{2b}}{R_2}\right)^2}, \quad (35)$$

$$M \cdot N|_{min,no \text{ RIS gain}} = \left[8 \frac{R_2}{\lambda} \frac{R_{2b}}{R_2} \left(1 - \frac{R_{2b}}{R_2}\right)\right]^2. \quad (36)$$

In practice, the necessary RIS size will be within the limits of Eqs. (34) and (36):

$$\log(M \cdot N|_{min}) \leq \log(M \cdot N) \leq \log(M \cdot N|_{min,no \text{ RIS gain}}) = 2 \log(M \cdot N|_{min}). \quad (37)$$

The signal processing capacity required to adjust the different RIS elements will be very complex. For smaller RIS sizes, the received power at the user equipment will be lower than for a reflector.

#### 4.2. Received Power Ratio for the Repeater vs. Reflector Approach

In the scenario involving a repeater, the ratio of the received power between the repeater and reflector concepts depends solely on the ratio of the entire path length  $R_2$  and the relative location of the reflector or repeater  $R_{2b}$ , as long as the path length  $R_{2b}$  is short enough to ensure sufficient received power at the repeater (Fig. 7):

$$\frac{P_R|_{repeater}}{P_R|_{reflector}} = \frac{(R_{2a} + R_{2b})^2}{R_{2b}^2} = \frac{R_2^2}{R_{2b}^2}. \quad (38)$$

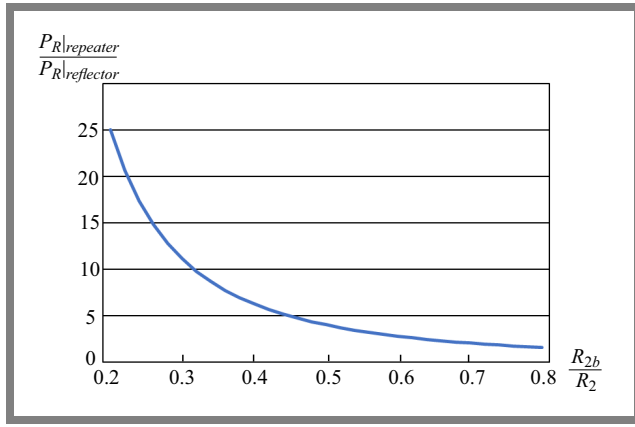


**Fig. 7.** Ratio of the power received by the repeater, normalized to a reflector solution, and the relative location of the reflector or repeater  $R_{2b}$ .

Equation (38) is valid under the condition that all transmitters have a radiated equivalent power  $EIRP_T$ . The repeater solution always provides a higher received power at the user equipment than the reflector. This gain is decreasing if the repeater is located closer to the base station. In that case, the length of the path of the second hop  $R_{2b}$  is similar to that of the entire path  $R_2$  in the reflector scenario. If the repeater is located in the middle of the entire radio path  $R_2$ , the transmit power necessary at the base station and at the repeater can be reduced for two hops  $N = 2$  by a factor of  $N^2 = 4$  – see Subsection 2.1.

## 5. Conclusions

In the millimeter wave and sub-terahertz frequency domains, pathloss is very high compared to lower frequency ranges (Section 1). Especially, under shadowing conditions, radio communication is basically blocked and without deploying special means, reflected multipath signals excessively lower the received power, since the roughness of surfaces characterized by a low reflection coefficient (e.g. walls walls) is in the order of the wavelength. Therefore, for practical applications, shadowing needs to be overcome by technical means which are directing radio signals around corners. The following options are considered:

- reflector or metallic mirror (Subsection 3.1),
- RIS array (Subsection 3.2),
- repeater (Subsection 3.3).

These technical means create a strong component in the channel impulse response at the receiver and thereby result in the establishment of a Rice channel.

From the comparison presented in Section 4 and Tables 1–2, it can be concluded that:

- A repeater is the best overall solution with a medium degree of complexity. The same electronic components as those used in base stations/access points and user equipment may be relied upon and only a power supply is needed. The overall transmit power is lower than for the other means. Therefore, this is the most sustainable solution;

- A reflector or a metallic mirror is a cheap alternative with low complexity. It does not require a power supply or an additional data connection;
- A RIS array requires a huge number of RIS elements and is the most complex solution in terms of signal processing. It also requires an additional power supply and a data connection to ensure centralized signal processing for RIS settings and to enable fast updates for moving users.

Especially for very wideband systems and in the millimeter wave and sub-terahertz domains, such as like 6G a Rice channel provides a significantly increased coherence bandwidth or reduced frequency dependency of the radio channel and increased received power compared to Rayleigh channels under shadowing conditions. This results in reduced effective intersymbol-interference (significantly more open eye-diagram), which reduces or avoids the need for time-domain equalization of the received signal.

In the scenarios involving a repeater or a reflector, coverage at high frequency ranges required by 6G systems may be improved significantly while maintaining reasonable complexity levels.

## References

- [1] ITU-R, “Working Party 5D: Workshop on IMT for 2030 and beyond”, Geneva, Switzerland, 2022 (<https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/Pages/wsp-int-vision-2030-and-beyond.aspx>).
- [2] W. Mohr, “Range and capacity considerations for Terahertz-Systems for 6G mobile and wireless communication”, *Journal of Mobile Multimedia*, vol. 19, no. 1, pp. 215, 2022 (<https://doi.org/10.13052/jmm1550-4646.19111>).
- [3] H. Meinke and F.W. Gundlach, “Taschenbuch der Hochfrequenztechnik”, Berlin: Springer Verlag, 870 p., 1986 (<https://doi.org/10.1007/978-3-642-96894-5>) (in German).
- [4] T.S. Rappaport, *Wireless Communications - Principles & Practice*, New Jersey: Prentice Hall, 640 p., 1996 (ISBN: 9780130422323).
- [5] J.B. Parsons, *The Mobile Radio Propagation Channel*, Pentech Press, London, 412 p., 1992 (<https://doi.org/10.1002/0470841524>).
- [6] ICNIRP (International Commission on Non-ionizing Radiation Protection), “ICNIRP Guidelines for limiting exposure to time-varying electric, magnetic and electromagnetic fields”, *Health Physics* vol. 74, no. 4, pp. 494–522; 1998 (<http://www.icnirp.org/cms/upload/publications/ICNIRPemfgdl.pdf>).
- [7] ICNIRP, “Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz)”, *Health Physics*, vol. 118, no. 5, pp. 483–524, 2020 (<https://doi.org/10.1097/HP.0000000000001210>).
- [8] S. Shah *et al.*, “Path Loss Model for Future Terahertz (THz) Wireless Communication Systems”, *Proceedings of the Pakistan Academy of Sciences*, vol. 56, no. 2, pp. 55–65, 2019 (ISSN: 2518-4245).
- [9] N.A. Abbasi, A. Hariharan, A.M. Nair, and A.F. Molisch, “Channel Measurements and Path loss Modeling for Indoor THz Communication”, in: *2020 14th European Conference on Antennas and Propagation (EuCAP)*, Copenhagen, Denmark, 2020 (<https://doi.org/10.23919/EuCAP48036.2020.9135643>).
- [10] E.N. Papasotiriou, A.A.A. Boullogeorgos, K. Haneda, M.F. de Guzman, and A. Alexiou, “An experimentally validated fading model for THz wireless systems”, *Scientific Reports*, art. no. 18717, 2021 (<https://doi.org/10.1038/s41598-021-98065-x>).
- [11] H. Brodhage and W. Hormuth, *Planung und Berechnung von Richtfunkverbindungen*, 10th updated ed., Siemens AG, 216 p., 1977 (ISBN: 9783800912421) (in German).
- [12] ITU-CCIR, “Propagation in Non-Ionized Media”, in: *Recommendations and Reports of the CCIR*, vol. 5, Geneva 1986 (ISBN: 9789261027415).

**Tab. 1.** Comparison of features

| Topic                           | Reflector                                                                                                                                       | RIS array                                                                                                                                                                                                                                        | Repeater                                                                                                                                                                                            |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mounting at                     | Buildings, towers or lamp posts                                                                                                                 |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                     |
| Power supply                    | No power supply                                                                                                                                 | Power supply needed                                                                                                                                                                                                                              |                                                                                                                                                                                                     |
| Data connection to the network  | No additional data connection                                                                                                                   | Additional data connection, deployment difficult and expensive                                                                                                                                                                                   | No additional data connection                                                                                                                                                                       |
| Active/passive device           | Passive                                                                                                                                         | Active due to signal processing for beamforming                                                                                                                                                                                                  | Active: amplification, potential signal processing for signal regeneration and beamforming                                                                                                          |
| Overall transmit power          | Transmit power needed to cover full range                                                                                                       | Transmit power needed to cover full range                                                                                                                                                                                                        | Overall transmit power lower than in other solutions for same range and therefore more sustainable                                                                                                  |
| Beamforming                     | Plane, concave or convex reflector, some fixed static beamforming feasible                                                                      | Beamforming by settings phase shifters of RIS elements; RIS gain (similar to antenna array gain)                                                                                                                                                 | Directive and/or MIMO antennas                                                                                                                                                                      |
| Update vs. receiver location    | Not needed, because reflector is covering an area around the receiver                                                                           | For a specific user location, the radio channel can be influenced or adjusted by RIS settings. However, RIS settings are sensitively linked to the receiver location, depending on beamwidth, and require fast updates due to receiver movements | Needed, because beamforming or MIMO antenna concepts applicable                                                                                                                                     |
| Amplification                   | No amplification                                                                                                                                |                                                                                                                                                                                                                                                  | Amplification                                                                                                                                                                                       |
| Signal regeneration             | No                                                                                                                                              |                                                                                                                                                                                                                                                  | RF repeater only: no, for analogue repeater (amplifier); Repeater including signal processing: yes, for repeater including signal processing by means of available information from radio interface |
| Signal processing               | No additional signal processing                                                                                                                 | Additional centralized signal processing between user equipment, base station and RIS                                                                                                                                                            | No additional signal processing in addition to the needs of the radio interface                                                                                                                     |
| Latency                         | No latency added to signal propagation                                                                                                          |                                                                                                                                                                                                                                                  | RF repeater only: nearly no latency added; Repeater including signal regeneration: signal processing for signal regeneration and antenna algorithms adds latency to signal propagation              |
| Shadowing mitigation            | Yes, allows to look around corners                                                                                                              |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                     |
| Coverage improvement            | Yes, shaded areas can be covered even in millimeter wave and sub-THz domains                                                                    |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                     |
| LOS like propagation conditions | Propagation corresponds to LOS propagation between base station and reflector and a second LOS propagation between reflector and user equipment | Propagation corresponds to LOS propagation between base station and RIS array and a second LOS propagation between RIS array and user equipment                                                                                                  | Propagation corresponds to LOS propagation between base station and repeater and a second LOS propagation between repeater and user equipment                                                       |

**Tab. 2.** Comparison of features – continued

| Topic                                                              | Reflector                                                                                                                                                                                                                                                | RIS array                                                                                                      | Repeater                                                                                                                                                    |
|--------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Range extension compared to propagation under shadowing conditions | Yes, based on quasi-LOS propagation around corners                                                                                                                                                                                                       | Yes, based on quasi-LOS propagation around corners, only for huge number of RIS elements better than reflector | Yes, based on amplification and directive antennas in repeater and quasi-LOS propagation around corners provides in maximum twice the range as for repeater |
| Strong component in channel impulse response                       | Yes, reflector, RIS array or repeater ensures a strong in channel impulse response compared to other secondary reflections around user equipment                                                                                                         |                                                                                                                |                                                                                                                                                             |
| Type of radio channel at receiver                                  | <b>Rice channel</b> due to strong component<br>Without reflector, RIS array or repeater received power much lower and radio channel corresponds to Rayleigh channel                                                                                      |                                                                                                                |                                                                                                                                                             |
| Delay spread at receiver                                           | Delay spread generated via secondary reflections around the receiver or other reflecting obstacles with smaller reflection coefficients in the deployment area.<br><b>Rice channel</b> – low delay spread, <b>Rayleigh channel</b> – higher delay spread |                                                                                                                |                                                                                                                                                             |
| Coherence bandwidth                                                | Enforced Rice channel at receiver provides very high coherence bandwidth                                                                                                                                                                                 |                                                                                                                |                                                                                                                                                             |
| Intersymbol interference                                           | With high coherence bandwidth little effective intersymbol interference, which reduces equalization effort                                                                                                                                               |                                                                                                                |                                                                                                                                                             |
| Coherence time                                                     | With increasing Rice factor of radio channel, coherence time is increasing.<br>Reduced requirements on equalizer update rate of channel estimation                                                                                                       |                                                                                                                |                                                                                                                                                             |
| Flexibility                                                        | Static, no flexibility                                                                                                                                                                                                                                   | Some beamforming due to RIS settings                                                                           | Very flexible by MIMO and signal regeneration                                                                                                               |
| Overall complexity                                                 | Low                                                                                                                                                                                                                                                      | High                                                                                                           | Medium                                                                                                                                                      |

- [13] ITU-CCIR, “Propagation in Non-Ionized Media”, in: *Recommendations and Reports of the CCIR*, Geneva 1982, vol. 5. (ISBN: 9789261014313).
- [14] V. Jung and H.-J. Warnecke, *Handbuch für die Telekommunikation*, Springer Verlag, Berlin, second ed., 2002 (<https://doi.org/10.1007/978-3-642-97702-2>) (in German).
- [15] T.S. Rappaport *et al.*, “Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond”, *IEEE Access*, vol. 7, pp. 78729–78757, 2019 (<https://doi.org/10.1109/ACCESS.2019.2921522>).
- [16] Federal Communications Commission (FCC), *Millimeter Wave Propagation: Spectrum Management Implications*, Office of Engineering and Technology, New Technology Development Division, Bulletin no. 70, 1997 ([https://transition.fcc.gov/Bureaus/Engineering\\_Technology/Documents/bulletins/oet70/oet70a.pdf](https://transition.fcc.gov/Bureaus/Engineering_Technology/Documents/bulletins/oet70/oet70a.pdf)).
- [17] G. Piefke, *Feldtheorie I*, vol. 771, 1977 (ISBN: 9783411007714) (in German).
- [18] O. Zinke and H. Brunswig, *Lehrbuch der Hochfrequenztechnik*, Springer Verlag, Berlin, Heidelberg, New York, 1973 (<https://doi.org/10.1007/978-3-662-00476-0>) (in German).
- [19] Wikimedia Commons [Online] [https://commons.wikimedia.org/wiki/File:Radar\\_cross\\_section\\_of\\_metal\\_sphere\\_from\\_Mie\\_theory.svg](https://commons.wikimedia.org/wiki/File:Radar_cross_section_of_metal_sphere_from_Mie_theory.svg).
- [20] M. Oziel, R. Korenstein, and B. Rubinsky, *The radar cross section (RCS),  $\sigma$ , as a function of the wavelength,  $\lambda$  (chart)*, in art. “Radar based technology for non-contact monitoring of accumulation of blood in the head: A numerical study”, 2017 (<https://doi.org/10.1371/journal.pone.0186381>).
- [21] H.-J. Li, Y.-W. Kiang, *The Electrical Engineering Handbook*, Elsevier, 2004 (ISBN: 0121709604).
- [22] “Radar cross section” [Online] <http://www.tscm.com/rcs.pdf>.
- [23] H. Mellein, “6G RIS – Shaping the radio channel for best connectivity”, *Rohde & Schwarz Webinar*, [Online] [https://www.rohde-schwarz.com/us/knowledge-center/webinars/webinar-ris-shaping-the-radio-channel-for-best-connectivity-registration\\_256353.html](https://www.rohde-schwarz.com/us/knowledge-center/webinars/webinar-ris-shaping-the-radio-channel-for-best-connectivity-registration_256353.html).
- [24] Radartutorial.eu, *Radar cross section*, [Online] <https://www.radartutorial.eu/01.basics/Radar%20Cross%20Section.en.html>.
- [25] Ariadne project: “On communication-theoretic models for programmable wireless environments enabled by reconfigurable intelligent surfaces”, 5G PPP white paper, ([https://5g-ppp.eu/wp-content/uploads/2021/06/ARIADNE\\_Beyond\\_Shannon-v1-0.pdf](https://5g-ppp.eu/wp-content/uploads/2021/06/ARIADNE_Beyond_Shannon-v1-0.pdf)).

Werner Mohr, Ph.D.

 <https://orcid.org/0009-0009-8957-2294>

E-mail: mohr\_werner@t-online.de

Independent consultant, Munich, Germany

# Comparative Analysis of NOMA and OMA Schemes: GSVD-based NOMA Systems and the Role of Mobile Edge Computing

Yunus Dursun

University of Manchester, Manchester, United Kingdom

<https://doi.org/10.26636/jtit.2023.3.1354>

**Abstract** — This paper presents a comprehensive study that examines the fundamental concept of the non-orthogonal multiple access (NOMA) scheme and provides its detailed comparison with the orthogonal multiple access (OMA) technique. Furthermore, the paper explores the application of the generalized singular value decomposition (GSVD) method in conjunction with NOMA, accompanied by a detailed review of GSVD-based NOMA systems. This study also introduces the concept of mobile edge computing (MEC) and extensively discusses its key parameters. Furthermore, a comprehensive analysis of NOMA MEC is presented, shedding light on its potential advantages and challenges. The aims of this study are to provide a comprehensive understanding of the aforementioned topics and contribute to the advancement of MIMO-NOMA systems.

**Keywords** — *generalized singular value decomposition, MIMO, mobile edge computing, non-orthogonal multiple-access*

## 1. Non-orthogonal Multiple Access

Non-orthogonal multiple access (NOMA) is a multiple access technique that allows multiple users to simultaneously share the same frequency, time, or code resources to communicate with a base station or an access point. Several NOMA schemes are distinguished in the literature [1]. Some of their well-known varieties are summarized below. The sparse code multiple access (SCMA) technique enables multiple users to communicate with a base station simultaneously by utilizing separate sparse codes assigned to each user based on a multi-dimensional codebook [2]. In the pattern division multiple access (PDMA) approach, the available bandwidth is divided into multiple non-overlapping frequency patterns or slots. Thus, the users rely on a unique pattern or slot to modulate their signals [3]. Resource spread multiple access (RSMA) employs a distinct spreading sequence for users to disperse their data over the frequency band. The receiver then reverses the spreading process by applying the identical spreading sequence to recover the user's information [4]. Multi-user shared access (MUSA) is based on code domain multiplexing, where symbols are multiplexed using the same spreading code. These symbols are transmitted over an orthogonal channel, such as a sub-carrier, as in OFDMA. At the receiver end, SIC decodes the received symbols [5]. Interleave-grid multiple access (IGMA) is another technique in which the user

data is segmented and interleaved based on a specific pattern, creating a grid-like structure that helps minimize user interference and improves the overall spectral efficiency of the system [6]. Rate-splitting multiple access is also evoking significant interest within the research community. In rate-splitting multiple access, users partition their data into shared and exclusive components [7]. The shared components of each user are aggregated and modulated jointly, while the unique components of each user are modulated separately. This results in a transmitted signal containing shared and unique components for all users. Both users initially decode the shared component at the receiver, treating any interference from the unique signals as noise. Both users use SIC to decode their exclusive signals in the subsequent stage.

In 3GPP Release 13, the standardization of power domain NOMA (PD-NOMA), known as multi-user superposition transmission (MUST), has been introduced for a broadcast channel. In PD-NOMA, multiple users use different power levels to share the same time and frequency resources. At the transmitter, superposition coding is employed to multiplex the users, while the receiver uses successive interference cancellation to decode the superposed signals [8]. PD-NOMA is considered a promising technique for 5G and beyond wireless communication systems, as it is capable of significantly increasing spectral efficiency and supporting multiple users with diverse communication requirements. PD-NOMA can also improve user fairness and energy efficiency, enabling users with weaker channel conditions to share the same resources with stronger users, without sacrificing their quality of service [9].

### 1.1. Overview of Power-domain NOMA

This section introduces the basic concepts of PD-NOMA for downlink and uplink networks. Additionally, we analyze and compare the sum rate and signal-to-interference-plus-noise ratio (SINR) of NOMA and OMA.

### 1.2. Downlink NOMA Network

Figure 1 illustrates a downlink NOMA scheme consisting of a base station or an access point and  $K$  receivers, where the BS broadcasts a superposed signal to all the receivers. The BS combines complex-valued symbols with superpo-

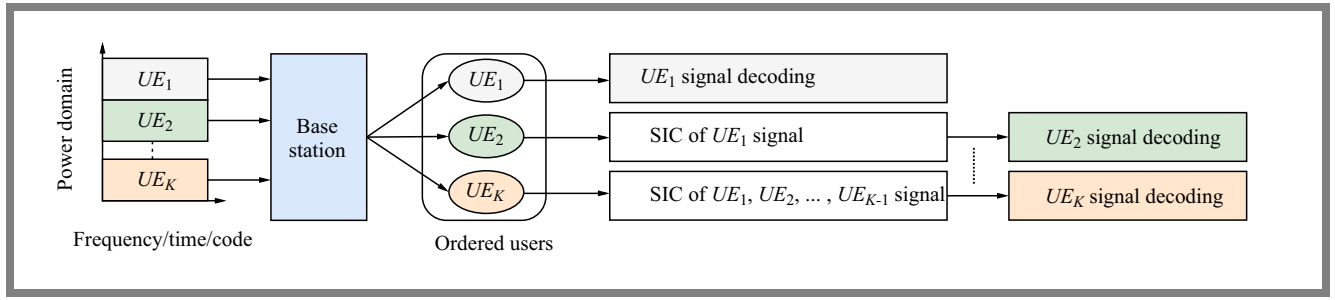


Fig. 1. Basic concept of a downlink NOMA [9].

sition coding (SC), and the receivers employ the successive interference cancellation (SIC) technique to decode their respective signals. Each receiver, except for the weakest one or receivers without SIC capabilities, performs the SIC process at the receiver. Even though there is a tendency for the SIC process as follows: the users first extract the strongest signal from the combined signal and then subtract it to eliminate interference from the remaining signals, and the SIC process is repeated until the receiver's signal is decoded. This strategy may not be optimal. Ding *et al.* showed that dynamic decoding orders relying on users' QoS and CSI-based SIC orders might improve the system's performance [10]. To simplify the analysis, we consider a downlink NOMA system with a base station and two users to derive the SINRs and sum rates. Additionally, we assume that the base station and users are equipped with a single antenna and the system bandwidth  $B$  is one. The information-bearing signals,  $x_N$  for the near user  $UE_1$  and  $x_F$  for the far user  $UE_2$ , are superimposed at the transmitter as follows:

$$x = \sqrt{P_N}x_N + \sqrt{P_F}x_F, \quad (1)$$

where  $P_N$  and  $P_F$  denote the transmission power allocation coefficients for the near and far users, respectively.  $P_{tot}$  represents the total transmit power which equals the sum of  $P_N$  and  $P_F$ . The received signal at the receivers are:

$$y_i = h_i x + n_i, \quad i \in \{N, F\}, \quad (2)$$

where  $h_i$  denotes the channel coefficient between the BS and user  $UE_i$  and  $n_i$  represents the additive white Gaussian noise (AWGN) with zero mean and  $\sigma_i^2$  variance for  $UE_i$ . Let's assume the users are ordered using the CSI-based method at the receiver and the near user has a strong signal than of the far user, i.e.,  $\frac{|h_N|^2}{\sigma_N^2} \geq \frac{|h_F|^2}{\sigma_F^2}$ . Therefore, the SINR expression of the near user and far user are given by:

$$SINR_N = \frac{P_N |h_N|^2}{\sigma_N^2}, \quad (3)$$

$$SINR_F = \frac{P_F |h_F|^2}{P_N |h_N|^2 + \sigma_F^2}. \quad (4)$$

Accordingly, the data rate for the near user and far user can be written as follows:

$$R_N = \log_2 \left( 1 + \frac{P_N |h_N|^2}{\sigma_N^2} \right), \quad (5)$$

$$R_F = \log_2 \left( 1 + \frac{P_F |h_F|^2}{P_N |h_N|^2 + \sigma_F^2} \right). \quad (6)$$

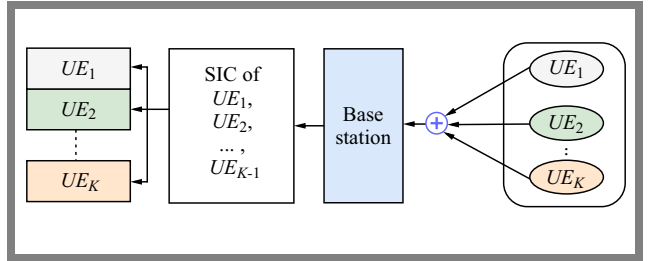


Fig. 2. Basic concept of uplink NOMA [9].

### 1.3. Uplink NOMA Network

As illustrated in Fig. 2, an uplink NOMA system allows  $K$  users to simultaneously transmit their data to the BS using the same spectrum. The base station employs SIC to decode the signals from different users. We again assume that the near user has better channel gain than the far user, i.e.  $\frac{|h_N|^2}{\sigma_N^2} \geq \frac{|h_F|^2}{\sigma_F^2}$ . The received signal at the receiver is:

$$y = h_N x_N + h_F x_F + n_B, \quad (7)$$

where  $n_B$  is an AWGN with zero mean and  $\sigma_N^2$  variance at the receiver. If the BS decodes the received signal in descending order, the data rate for the near and far users are:

$$R_N = \log_2 \left( 1 + \frac{P_N |h_N|^2}{P_F |h_F|^2 + \sigma_B^2} \right), \quad (8)$$

$$R_F = \log_2 \left( 1 + \frac{P_F |h_F|^2}{\sigma_B^2} \right). \quad (9)$$

On the other hand, if the BS decodes the received signal in ascending order, the data rate for the near and far users becomes:

$$R_N = \log_2 \left( 1 + \frac{P_N |h_N|^2}{\sigma_B^2} \right), \quad (10)$$

$$R_F = \log_2 \left( 1 + \frac{P_F |h_F|^2}{P_N |h_N|^2 + \sigma_B^2} \right). \quad (11)$$

It is worth mentioning that in each case, the sum rate for the users is the same as given in:

$$R_N + R_F = \log_2 \left( \frac{P_N |h_N|^2 + P_F |h_F|^2 + \sigma_B^2}{\sigma_B^2} \right). \quad (12)$$

In other words, the sum rate in the uplink NOMA does not depend on the order of SIC, assuming no error propagation occurs in the SIC process. However, according to Benjebbour [8], it is more practical to perform SIC in the descending order of channel quality levels.

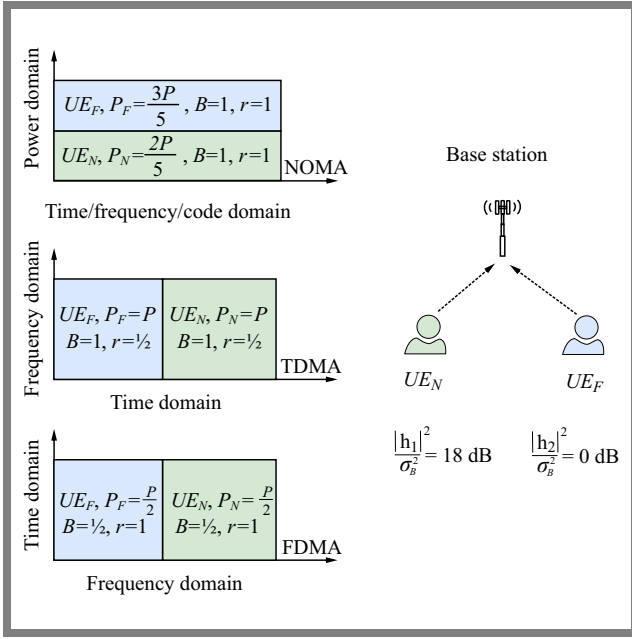


**Fig. 3.** Sum rate comparison between NOMA, TDMA, and FDMA networks.

#### 1.4. Sum Rate Comparison Between NOMA, TDMA, and FDMA Networks

This section compares NOMA with TDMA and FDMA schemes for an uplink scenario, as illustrated in Fig. 3. The data rates of the near and far users in the NOMA, FDMA, and TDMA systems are given as follows [11]:

$$\begin{aligned}
 \text{NOMA} &= \begin{cases} R_N^{\text{NOMA}} = B \log_2(1 + \frac{P_N |h_N|^2}{B \sigma_B^2}) \\ R_F^{\text{NOMA}} = B \log_2(1 + \frac{P_F |h_F|^2}{P_N |h_N|^2 + B \sigma_B^2}) \\ 0 \leq P_N, P_F \leq P \end{cases} \\
 \text{TDMA} &= \begin{cases} R_N^{\text{TDMA}} = B(1 - \tau) \log_2(1 + \frac{P_N |h_N|^2}{B \sigma_B^2}) \\ R_F^{\text{TDMA}} = B\tau \log_2(1 + \frac{P_F |h_F|^2}{B \sigma_B^2}) \\ 0 \leq \tau \leq 1 \end{cases} \\
 \text{FDMA} &= \begin{cases} R_N^{\text{FDMA}} = B(1 - \omega) \log_2(1 + \frac{P_N |h_N|^2}{B(1 - \omega) \sigma_B^2}) \\ R_F^{\text{FDMA}} = B\omega \log_2(1 + \frac{P_F |h_F|^2}{B\omega \sigma_B^2}) \\ 0 \leq \omega \leq 1. \end{cases}
 \end{aligned} \quad (13)$$

Assuming two uplink users  $UE_N$  and  $UE_F$  with channel gains of  $\frac{h_N}{\sigma_B} = 18$  dB and  $\frac{h_F}{\sigma_B} = 0$  dB, respectively, the total power is the same in all schemes such that  $P_N + P_F = P$ , where  $P$  is the maximum transmit power [8]. In the TDMA scheme, the users are allocated equal time slots, i.e.,  $\tau = 0.5$ . The data rates for the near and far users are  $R_N^{\text{TDMA}} = 3.0011$  bps and  $R_F^{\text{TDMA}} = 0.5$  bps, respectively. In the FDMA scheme, the bandwidth is split equally between the users, i.e.,  $\omega = 0.5$ , and the resulting data rates are  $R_N^{\text{FDMA}} = 3.0011$  bps and  $R_F^{\text{FDMA}} = 0.5$  bps for the near and far users, respectively. In the NOMA case, the power is split between the users by  $\delta$ , with two out of five for the near user and three out of five for the far user. Thus,

the data rates for the near and far users are  $R_N^{\text{NOMA}} = 4.0682$  bps and  $R_F^{\text{NOMA}} = 0.6781$  bps. The total sum rates achieved by the TDMA, FDMA, and NOMA schemes are 3.5011 bps, 3.5011 bps, and 4.7463 bps, respectively. Based on this example, it can be concluded that NOMA offers a significant advantage over OMA schemes in terms of spectral efficiency, resulting in 35.57% higher sum data rates. Figure 4 emphasizes the influence of the power allocation coefficient on the sum rate performance. In Fig. 5, a comparison is presented between the sum rates of NOMA, TDMA, and FDMA, with the power allocation coefficient  $\delta$ , the bandwidth allocation coefficient  $\omega$ , and the time allocation coefficient  $\tau$  set to 0.3. As the channel gain difference increases, it is observed that NOMA outperforms OMA schemes, exhibiting greater improvement in performance.

## 2. Enhancing MIMO-NOMA Systems Through GSVD

Generalized singular value decomposition (GSVD) is a powerful matrix factorization technique that extends the standard singular value decomposition (SVD) to accommodate rectangular matrices of potentially different dimensions. This technique encompasses two primary types: real-valued GSVD and complex-valued GSVD. Real GSVD is utilized for real-valued matrices, while complex GSVD is tailored for complex-valued matrices. Various algorithms, such as Van Loan's, which was first introduced in 1976, as well as Paige and Saunders' algorithms [12], facilitate the computation of GSVD. By decomposing matrices into their singular components, GSVD finds a widespread application across diverse domains. It is employed in signal processing tasks, such as adaptive filtering, blind source separation, and channel estimation, as well as in analyzing biological data in bioinformatics [13], [14]. In wireless communication, GSVD decomposes MIMO channels into orthogonal SISO channels. In other words, the main use case of the GSVD is beamforming design. Table 1 lists

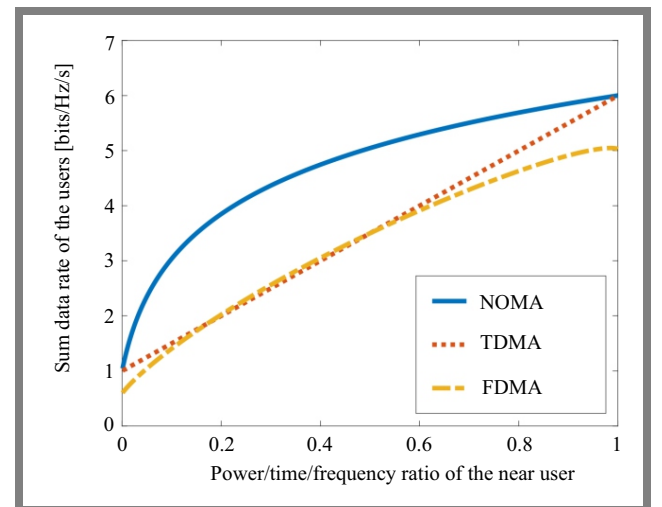


**Fig. 4.** Sum data rate comparison for different resource ratios for the near user.

some studies that employ GSVD to solve some problems in wireless communication.

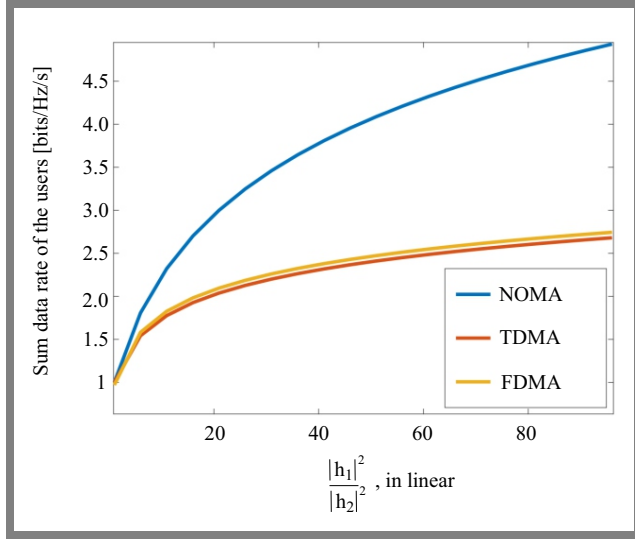


Fig. 5. Sum data rate comparison under different SNR rates.

### 2.1. Definition of GSVD

Let us consider two matrices,  $\mathbf{H}_1 \in \mathbb{C}^{m \times n}$  and  $\mathbf{H}_2 \in \mathbb{C}^{m \times n}$ . By applying the GSVD method, we can decompose these matrices into three components: a unitary matrix, a non-singular matrix, and a non-negative singular matrix. The decomposition can be expressed as follows [15]:

$$\Sigma_1 = \mathbf{U}\mathbf{H}_1\mathbf{Q} \quad \text{and} \quad \Sigma_2 = \mathbf{V}\mathbf{H}_2\mathbf{Q}, \quad (14)$$

where the matrices  $\mathbf{U} \in \mathbb{C}^{m \times m}$  and  $\mathbf{V} \in \mathbb{C}^{m \times m}$  are unitary, while  $\mathbf{Q} \in \mathbb{C}^{n \times n}$  is non-singular, and  $\Sigma_1 \in \mathbb{C}^{m \times n}$  and  $\Sigma_2 \in \mathbb{C}^{m \times n}$  are diagonal matrices with non-negative elements. The dimension of the matrices form  $\Sigma_1$  and  $\Sigma_2$  as follows:

- If  $m \geq n$ , then  $\Sigma_1 = \begin{pmatrix} \mathbf{0}_{(m-n) \times n} \\ \mathbf{S}_1 \end{pmatrix}$   
and  $\Sigma_2 = \begin{pmatrix} \mathbf{S}_2 \\ \mathbf{0}_{(m-n) \times n} \end{pmatrix}$ .
- If  $m \leq n \leq 2m$ ,  $r = n - m$  and  $q = 2m - n$ , then  
 $\Sigma_1 = \begin{pmatrix} \mathbf{I}_r & \mathbf{O}_{r \times q} & \mathbf{O}_{r \times r} \\ \mathbf{O}_{q \times r} & \mathbf{S}_1 & \mathbf{O}_{q \times r} \end{pmatrix}$   
and  $\Sigma_2 = \begin{pmatrix} \mathbf{O}_{q \times r} & \mathbf{S}_2 & \mathbf{O}_{q \times r} \\ \mathbf{O}_{r \times r} & \mathbf{O}_{r \times q} & \mathbf{I}_r \end{pmatrix}$ .
- If  $2m \geq n$ , then  $\Sigma_1 = \begin{pmatrix} \mathbf{I}_m & \mathbf{O}_{m \times (n-m)} \end{pmatrix}$   
and  $\Sigma_2 = \begin{pmatrix} \mathbf{O}_{m \times (n-m)} & \mathbf{I}_m \end{pmatrix}$ ,

where  $\mathbf{O}$  and  $\mathbf{I}$  represent the zero and identity matrices, respectively. Moreover,  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are non-negative diagonal matrices, with elements between zero and one. Notably, the elements of  $\mathbf{S}_1$  are sorted in descending order, while those of  $\mathbf{S}_2$  are sorted in ascending order.

### 2.2. Application of GSVD in MIMO-NOMA

We consider a base station (BS) with  $n$  antennas communicating with two downlink users, each equipped with  $m$  antennas. The channels between the BS and the users can be represented by  $\mathbf{G}_i = \frac{\mathbf{H}_i}{\sqrt{d_i^\tau}}$ , where  $\mathbf{H}_i$  denotes the small-scale fading coefficients. The near user, denoted as  $UE_N$ , and the far user, denoted as  $UE_F$ , are sorted based on their large-scale fading element  $\sqrt{d_i^\tau}$ . Here,  $d$  represents the distance of the  $i$ -th user, and  $\tau$  represents their path loss component. The received signals at the receivers are:

$$\mathbf{y}_N = \frac{\mathbf{H}_N \mathbf{x}}{\sqrt{d_N^\tau}} + \mathbf{n}_N, \quad (15)$$

$$\mathbf{y}_F = \frac{\mathbf{H}_F \mathbf{x}}{\sqrt{d_F^\tau}} + \mathbf{n}_F. \quad (16)$$

The noise at the  $i$ -th receiver,  $\mathbf{n}_i$ ,  $i \in N, F$ , modeled by the additive white Gaussian noise, is given by mutually independent and identically distributed elements with zero mean and variance  $\sigma_i$ . The transmitted signal, denoted as  $\mathbf{x} \in \mathbb{C}^{n \times 1}$ , is subject to interference mitigation techniques using precoding and decoding matrices  $\mathbf{P}_b \in \mathbb{C}^{n \times n}$  and  $\mathbf{D}_i \in \mathbb{C}^{m \times m}$ , respectively. The decomposition of channels, as described in Eq. (14), leads to the selection of detection matrices  $\mathbf{D}_i$  as  $\mathbf{U}$  and  $\mathbf{V}$  for the near and far users. Additionally, the precoding matrix  $\mathbf{P}_b$  is modified to  $\mathbf{Q}\sqrt{P}/t$ , where  $P$  represents the maximum transmission power and  $t$  is a power normalization coefficient [15]. Consequently, the MIMO receivers obtain the transmitted signal as follows:

$$\begin{aligned} \mathbf{U}\mathbf{y}_N &= \mathbf{U}\mathbf{H}_N\mathbf{P}_b\mathbf{x} + \mathbf{U}\mathbf{n}_N = \frac{P}{t\sqrt{d_N^\tau}}\Sigma_N + \mathbf{U}\mathbf{n}_N \\ \mathbf{V}\mathbf{y}_F &= \mathbf{V}\mathbf{H}_F\mathbf{P}_b\mathbf{x} + \mathbf{V}\mathbf{n}_F = \frac{P}{t\sqrt{d_F^\tau}}\Sigma_F + \mathbf{V}\mathbf{n}_F. \end{aligned} \quad (17)$$

Please note that  $\mathbf{U}$  and  $\mathbf{V}$  are the unitary matrices. Therefore, the unitary matrices  $\mathbf{U}$  and  $\mathbf{V}$  preserve the variance of noise after multiplication with therewith.

**Example:** In this example, we analyze a basic setup comprising a BS with four transmitter antennas. The near user is equipped with three receiver antennas, while the distant users have two receiver antennas. The near, which small-scale channel coefficient denoted as  $\mathbf{H}_N \in \mathbb{C}^{3 \times 4}$ , is located  $d_N = 40$  meters away from the base station. On the other hand, the far user, represented by  $\mathbf{H}_F \in \mathbb{C}^{2 \times 4}$ , is located  $d_F = 75$  meters away from the base station. The value of the path loss component, denoted as  $\alpha$ , is 3.2.

$$\mathbf{H}_N = \begin{bmatrix} 0.629 + 0.735i & 0.066 + 0.931i \\ 0.210 + 0.772i & 0.260 + 0.013i \\ 0.752 + 0.907i & 0.804 + 0.234i \\ 0.193 + 0.616i & 0.924 + 0.556i \\ 0.639 + 0.949i & 0.263 + 0.915i \\ 0.524 + 0.950i & 0.065 + 0.641i \end{bmatrix}.$$

$$\mathbf{H}_F = \begin{bmatrix} 0.390 + 0.173i & 0.604 + 0.135i \\ 0.485 + 0.126i & 0.549 + 0.505i \\ 0.926 + 0.021i & 0.394 + 0.827i \\ 0.918 + 0.947i & 0.963 + 0.015i \end{bmatrix}.$$

Decoding matrices  $\mathbf{U} \in \mathbb{C}^{3 \times 3}$  and  $\mathbf{V} \in \mathbb{C}^{2 \times 2}$  can be found as follows:

$$\mathbf{U} = \begin{bmatrix} 0.5886 + 0.0000i & 0.0000 + 0.0000i \\ 0.2571 - 0.6199i & 0.4855 + 0.2742i \\ -0.2400 + 0.3816i & 0.6442 + 0.5235i \\ 0.8084 + 0.0000i \\ -0.1872 + 0.4513i \\ 0.1747 - 0.2778i \end{bmatrix}.$$

$$\mathbf{V} = \begin{bmatrix} -0.5294 + 0.7844i & 0.3201 + 0.0453i \\ -0.2029 - 0.2517i & 0.4024 - 0.8565i \end{bmatrix},$$

and the precoding matrix  $\mathbf{Q}$  becomes:

$$\mathbf{Q} = \begin{bmatrix} -0.2004 + 0.3013i & 0.2468 - 0.5664i \\ -0.4527 + 0.5098i & -0.0139 - 0.7738i \\ -0.8984 + 0.6990i & -0.1623 - 1.2709i \\ 0.2405 + 0.5082i & 0.6040 - 1.2088i \\ 1.2740 - 0.5080i & 0.6976 - 0.7226i \\ 0.7713 + 0.3349i & 0.0869 - 0.8978i \\ 1.4061 - 0.6229i & 0.2928 - 0.3442i \\ 0.7572 - 0.7511i & 0.9449 - 0.2902i \end{bmatrix}.$$

Let's rearrange the super-positioned transmitted signal  $\mathbf{x} \in \mathbb{C}^{4 \times 1}$  with power allocation such that  $\mathbf{x} = \mathbf{P}\mathbf{s}$ , where  $\mathbf{P} \in \mathbb{C}^{4 \times 4}$  is the diagonal non-negative power allocation matrix and  $\mathbf{s} \in \mathbb{C}^{4 \times 1}$  contains the coded signals for both users as follows:

$$\mathbf{x} = \underbrace{\begin{bmatrix} p_{1,1} & 0 & 0 & 0 \\ 0 & p_{2,2} & 0 & 0 \\ 0 & 0 & p_{3,3} & 0 \\ 0 & 0 & 0 & p_{4,4} \end{bmatrix}}_{\mathbf{P}} \times \underbrace{\begin{bmatrix} l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2} \\ l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2} \\ l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2} \\ l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2} \end{bmatrix}}_{\mathbf{s}}$$

$$= \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix},$$

where  $s_{i,1}$  and  $s_{i,2}$  represent the corresponding message, and  $l_{i,1}$  and  $l_{i,2}$  are the power allocation coefficients for the near and far users, respectively. Also, we assume that  $\mathbf{s}$  encoded with unit power i.e.,  $\|\mathbf{s}_i\|^2 = 1$ ,  $i \in \{1, 2, 3, 4\}$ . After the GSVD is applied to the downlink channels, the channels become:

$$\mathbf{\Sigma}_N = \begin{bmatrix} 0 & 0.4526 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{\Sigma}_F = \begin{bmatrix} 1.0000 & 0 & 0 & 0 \\ 0 & 0.8917 & 0 & 0 \end{bmatrix}.$$

Observations at the near user are equal to  $\mathbf{y}_N = \frac{\mathbf{\Sigma}_N \mathbf{x}}{\sqrt{d_N^\alpha}} + \mathbf{n}_N$  that can be written as follows:

$$\mathbf{y}_N = \begin{bmatrix} 0 & 0.4526 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix} \\ \times \frac{1}{\sqrt{d_N^\alpha}} + \begin{bmatrix} n_{N,1} \\ n_{N,2} \\ n_{N,3} \end{bmatrix} \\ = \begin{bmatrix} 0.4526 \times (l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \sqrt{d_N^{-\alpha}} + n_{N,1} \\ 1 \times (l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \sqrt{d_N^{-\alpha}} + n_{N,2} \\ 1 \times (l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \sqrt{d_N^{-\alpha}} + n_{N,3} \end{bmatrix}.$$

Similarly, observations at the far user  $\mathbf{y}_F$  are equal to  $\mathbf{y}_F = \frac{\mathbf{\Sigma}_F \mathbf{x}}{\sqrt{d_F^\alpha}} + \mathbf{n}_F$  and can be given as follows:

$$\mathbf{y}_F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8917 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix} \\ \times \frac{1}{\sqrt{d_F^\alpha}} + \begin{bmatrix} n_{F,1} \\ n_{F,2} \end{bmatrix} \\ = \begin{bmatrix} 1 \times (l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \sqrt{d_F^{-\alpha}} + n_{F,1} \\ 0.8917 \times (l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \sqrt{d_F^{-\alpha}} + n_{F,2} \end{bmatrix}.$$

We can consider each sub-channel an individual SISO channel. Therefore, they may require different SIC ordering regard-

ing their effective channel gains. Furthermore, by examining the expressions for  $\mathbf{y}_N$  and  $\mathbf{y}_F$ , we can deduce that  $\mathbf{s}_1$  corresponds to a private stream intended for the far user. On the other hand,  $\mathbf{s}_3$  and  $\mathbf{s}_4$  represent private streams dedicated to the near user, while  $\mathbf{s}_2$  serves as the common stream shared by both users. The OMA transmission strategy can be employed for private streams. For example, the first stream may equal  $\mathbf{s}_1 = \mathbf{s}_{1,2}$ .

Now, let us have a look at the decoding of the common stream. It can be calculated that the near user has a stronger channel gain than the far user, given by  $\frac{0.4525}{\sqrt{40^{3.2}}} \geq \frac{0.8917}{\sqrt{75^{3.2}}}$ . Therefore, the near and far users decode their signals as follows:

$$R_{N,2} = \log_2 \left( 1 + \frac{p_{2,2} \times l_2^2 \frac{0.4525}{40^{3.2}}}{\text{Var}(n_{N,2})} \right),$$

$$R_{F,2} = \log_2 \left( 1 + \frac{p_{2,2} \times (1 - l_2^2) \times \frac{0.8917}{75^{3.2}}}{p_{2,2} \times l_2^2 \frac{0.4525}{40^{3.2}} + \text{Var}(n_{F,2})} \right).$$

Likewise, data rate expressions can be derived for private channels by eliminating inter-user interference. Additionally, the transmitted power from each antenna, denoted as  $\mathbf{P}$ , holds a specific physical interpretation, such as maximum transmit power from each antenna element. Hence, the opportunity for further exploration emerges from optimizing the elements within  $\mathbf{P}$ , encompassing enhancing secrecy, improving data rate, ensuring fairness, and selecting optimal antennas.

### 3. Mobile Edge Computing

Mobile edge computing, also known as multi-access edge computing, brings the processing of traffic and services from centralized cloud servers to the edge of the network, closer to the end-users, as illustrated in Fig. 6. Instead of transmitting all the data to the cloud for analysis, MEC devices are responsible for the processing, storage, and analysis of the data [24]. This approach minimizes latency, thus improving performance of high-bandwidth applications in real-time [25]. The combination of NOMA and MEC holds immense potential, as it not only enhances the spectral efficiency of MEC users, but also empowers IoT devices at the edge to handle computationally intensive tasks. Combining NOMA and MEC requires an optimal approach to resource and power allocation and time management. In order to minimize the time spent offloading tasks,  $UE$  needs to determine the optimal task partition coefficient  $\beta$  and power allocation  $p_{off}$ . The offloading time ( $T_{off}$ ) can be defined as follows:

$$T_{off} = \frac{\beta N}{R} \text{ [s]}, \quad (18)$$

Here,  $N$  represents the data size of the task, and  $R$  is the data rate of the  $UE$ .

The energy consumed during the offloading time  $T_{off}$  can be calculated as:

$$E_{off} = T_{off} \times p_{off} \text{ [J]}, \quad (19)$$

In Eq. (19),  $p_{off}$  denotes the transmit power of the  $UE$ . Once the data is offloaded to the MEC server, the duration for the

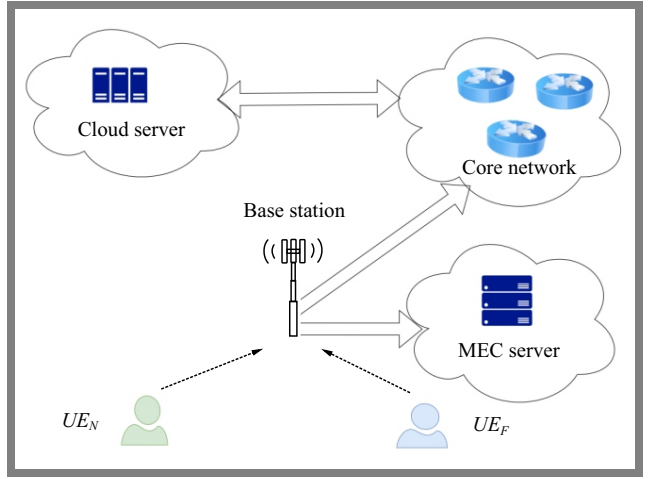


Fig. 6. NOMA-assisted MEC model.

mobile execution time  $T_{mec}$  can be determined using:

$$T_{mec} = \frac{\beta N C_m}{f_m} \text{ [s]}, \quad (20)$$

In this equation,  $C_m$  represents the required CPU cycles to execute a bit, and  $f_m$  is the CPU frequency of the MEC server. The energy consumption during  $T_{mec}$  can be calculated as:

$$E_{mec} = \xi \beta N C_m f_m^2 \text{ [J]}, \quad (21)$$

where  $\xi$  denotes the energy consumption coefficient for the MEC. Tables 2–3 summarize existing works on combining NOMA and MEC. It provides valuable insights into different research papers that have explored this subject, with a particular emphasis placed on potential optimization benefits.

### 4. Discussions and Future Works

In this paper, we have presented a comprehensive analysis of NOMA and compared it with OMA schemes. The study also explored the integration of NOMA with MIMO technologies using the GSVD method. Additionally, we discussed the key parameters of the MEC technology and provided a literature review on NOMA with MEC.

Future research directions in this field involve addressing the limitations that come with assuming availability of perfect channel state information (CSI) at the transmitters and receivers. It is recommended to investigate the implications and benefits of incorporating imperfect channel estimation with random error matrices, as demonstrated in [36]. By considering such realistic scenarios with imperfect CSI, the findings of this study can be extended to real-world applications. Furthermore, the constraints of the GSVD technique, which currently allow the combination of only two users simultaneously, pose another limitation. However, recent advancements have shown that the GSVD technique can be extended to accommodate more than two users, as highlighted in [37] and [38]. The results of these studies indicate the potential of assigning more than two users to a resource block, using the proposed methods.

An intriguing path for future research involves integrating the

**Tab. 1.** GSVD-based MIMO applications.

| Ref. | Objective                                                                                      | Technology  | System analysis                       | Optimization variable                             | Constraints                                                    | UL/DL | Result                                                                                                                                        |
|------|------------------------------------------------------------------------------------------------|-------------|---------------------------------------|---------------------------------------------------|----------------------------------------------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| [16] | Minimize outage probability while improving physical layer security                            | NOMA-MIMO   | Performance analysis and optimization | Power allocation coefficients                     | N/A                                                            | DL    | Compared with GSVD-OMA based transmission, NOMA has superior outage performance                                                               |
| [17] | Design low complexity and highly efficient GSVD-based beamforming to maximize secrecy capacity | OMA-MIMO    | Optimization                          | Power allocation coefficients                     | Average power consumption                                      | DL    | GSVD-MIMO achieves nearly identical performance with secure dirty paper coding (S-DPC)                                                        |
| [18] | Maximize secrecy rate                                                                          | OMA-MIMO    | Optimization                          | Sub-channel and power allocation                  | Quality of service                                             | DL    | GSVD-based precoding outperforms a TDMA-based system                                                                                          |
| [19] | Obtain the expressions of the average data rate and outage in a MIMO-NOMA relaying             | NOMA-MIMO   | Performance analysis                  | N/A                                               | Finite number of users                                         | DL    | GSVD-NOMA achieves a higher sum rate than GSVD-OMA                                                                                            |
| [20] | Maximize minimum data rate                                                                     | NOMA-MIMO   | Optimization                          | Power allocation coefficients                     | Imperfect channel estimation                                   | DL    | The SINR balancing problem was solved using error bounds. The proposed solution has better performance than non-robust or OMA-based solutions |
| [21] | Minimize offloading delay                                                                      | H-NOMA-MIMO | Optimization                          | Power allocation coefficients                     | Total power                                                    | UL    | Hybrid NOMA-MIMO based solution has better delay performance compared with OMA-based solution                                                 |
| [22] | Minimize energy consumption                                                                    | NOMA-MIMO   | Optimization                          | Task assignment and power allocation coefficients | Total power, offloading time, and RF chains energy consumption | UL    | NOMA-MIMO performs better than OMA, especially when the data is high and time is stringent                                                    |
| [23] | Maximize secrecy sum rate                                                                      | NOMA-MIMO   | Optimization                          | Power allocation coefficients                     | Total power and QoS                                            | UL    | NOMA has better SSR performance than OMA                                                                                                      |

**Tab. 2.** Summary of some existing works on NOMA-MEC.

| Ref. | Objective                                                            | Method                                  | Technology            | Constraints                                          | Optimization parameters                              | Offloading policy |
|------|----------------------------------------------------------------------|-----------------------------------------|-----------------------|------------------------------------------------------|------------------------------------------------------|-------------------|
| [26] | Minimize task offloading and computing delay                         | Linear problem using auxiliary variable | Massive MIMO          | Transmit power, MEC computing capacity               | Power allocation, computation frequency allocation   | Partial           |
| [24] | Minimize delay                                                       | Bisection search                        | SISO-MEC              | Energy and offloading power, computation time at MEC | Task partition coefficient, power allocation         | Partial           |
| [22] | Minimize total energy during local task offloading and MEC computing | AO-SCA                                  | MIMO-MEC              | Total power, time, energy consumption on RF chains   | Task partition coefficient, power allocation         | Partial           |
| [23] | Minimize delay                                                       | Dinkelbach transform-SCA                | MIMO-MEC              | Total power                                          | Power allocations                                    | Full              |
| [27] | Minimize system energy consumption                                   | SCA                                     | SISO-MEC              | Time, transmission power, decoding power             | Power allocation                                     | Full              |
| [28] | Minimize delay                                                       | Alternating optimization (AO)           | UAV assisted SISO-MEC | Energy and QoS                                       | Trajectory of UAV, power allocation, user scheduling | Full              |

proposed system models with emerging technologies, such as intelligent reflecting surfaces and unmanned aerial vehicles. The integration of these technologies holds significant promise and provides opportunities to enhance the signal-to-interference-plus-noise ratio (SINR) of MIMO-NOMA systems. Overall, this research contributes to the understanding of NOMA, its comparison with OMA schemes, the application of GSVD in NOMA systems, and the exploration of key parameters associated with the MEC technology. The identified research directions open up avenues for further advancements in wireless communication systems, enabling more efficient and reliable transmissions in diverse scenarios.

## References

- [1] B. Makki *et al.*, “A Survey of NOMA: Current Status and Open Research Challenges”, *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020 (<https://doi.org/10.1109/OJCOMS.2020.2969899>).
- [2] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, “SC-MA Codebook Design”, in: *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Vancouver, Canada, pp. 1–5, 2014 (<https://doi.org/10.1109/VTCFall.2014.6966170>).
- [3] J. Zeng *et al.*, “Pattern Division Multiple Access (PDMA) for Cellular Future Radio Access”, in: *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, Nanjing, China, pp. 1–5, 2015 (<https://doi.org/10.1109/WCSP.2015.7341229>).
- [4] Y. Cao, H. Sun, J. Soriaga, and T. Ji, “Resource Spread Multiple Access – A Novel Transmission Scheme for 5G Uplink”, in: *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Toronto, Canada, 2017 (<https://doi.org/10.1109/VTCFall.2017.8288412>).
- [5] Z. Yuan *et al.*, “Multi-User Shared Access for Internet of Things”, in: *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016 (<https://doi.org/10.1109/VTCSpring.2016.7504361>).
- [6] S. Hu *et al.*, “Nonorthogonal Interleave-Grid Multiple Access Scheme for Industrial Internet of Things in 5G Network”, *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5436–5446, 2018 (<https://doi.org/10.1109/TII.2018.2858142>).
- [7] B. Clerckx *et al.*, “Rate Splitting for MIMO Wireless Networks: a Promising Physical Layer Strategy for LTE Evolution”, *IEEE Communications Magazine*, vol. 54, no. 5, pp. 98–105, 2016 (<https://doi.org/10.1109/MCOM.2016.7470942>).
- [8] A. Benjebbour, “An Overview of Non-Orthogonal Multiple Access”, *ZTE Communications*, vol. 15, no. 1, pp. 28–34, 2017 (<https://doi.org/10.3969/j.issn.1673-5188.2017.01.005>).
- [9] M. Aldababsa *et al.*, “A Tutorial on Non-orthogonal Multiple Access for 5G and Beyond”, *Wireless Communications and Mobile Computing*, pp. 1–24, 2018 (<https://doi.org/10.1155/2018/9713450>).
- [10] Z. Ding, R. Schober, and H.V. Poor, “Unveiling the Importance of SIC in NOMA Systems – Part 1: State of the Art and Recent Findings”, *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373–2377, 2020 (<https://doi.org/10.1109/LCOMM.2020.3012604>).
- [11] Z. Yang *et al.*, “Sum-Rate Maximization of Uplink Rate Splitting Multiple Access (RSMA) Communication”, in: *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, USA, 2019 (<https://doi.org/10.1109/GLOBECOM38437.2019.9013344>).
- [12] C.C. Paige, “Computing the Generalized Singular Value Decomposition”, *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1126–1146, 1986 (<https://doi.org/10.1137/0907077>).

**Tab. 3.** Summary of some existing works on NOMA-MEC – continued.

| Ref. | Objective                         | Method                                   | Technology                     | Constraints                            | Optimization parameters                                                                                        | Offloading policy     |
|------|-----------------------------------|------------------------------------------|--------------------------------|----------------------------------------|----------------------------------------------------------------------------------------------------------------|-----------------------|
| [29] | Maximize computation capacity     | AO, DC programming                       | Backscatter-assisted SISO_NOMA | Energy and QoS                         | Energy harvesting time coefficient, BackCom time coefficient, transmission time, computing resource allocation | Partial               |
| [30] | Maximize computation capacity     | AO, concave-convex procedure and SDR     | IRS and UAV assisted SISO-NOMA | Total transmit power                   | Phase shift of IRS, transmit power, computational resource allocation, the trajectory of UAV                   | Binary                |
| [31] | Minimize system energy            | AO, matching algorithm, SCA              | SISO-MEC                       | Latency                                | Power allocation, time, sub-channel allocation                                                                 | Binary                |
| [32] | Maximize computation capacity     | Deep reinforcement learning              | SISO-multi-MEC                 | Delay, limited sub-channel             | Task scheduling, power allocation                                                                              | Binary                |
| [25] | Minimize latency                  | AO                                       | WPT, IRS                       | Transmit power                         | Power allocation, phase shift of IRS                                                                           | Partial               |
| [33] | Minimize total energy consumption | TD3                                      | SISO-MEC                       | Transmit power, latency                | Task partition, power allocation                                                                               | Binary                |
| [34] | Maximize computation probability  | Meta-heuristic-based algorithms, PSO, GA | SISO-MEC                       | Transmit power, computational resource | Task partition, power allocation                                                                               | Binary, partial, full |
| [35] | Minimize power consumption        | AO, Riemann gradient descent             | IRS-MEC                        | Delay, computational resource          | Bandwidth allocation, computational resource allocation, power allocation, the phase shift of IRS              | Binary                |

- [13] S. Doclo and M. Moonen, “GSVD-based Optimal Filtering for Single and Multi-microphone Speech Enhancement”, *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002 (<https://doi.org/10.1109/TSP.2002.801937>).
- [14] O. Alter, P.O. Brown, and D. Botstein, “Generalized Singular Value Decomposition for Comparative Analysis of Genome-scale Expression Data Sets of Two Different Organisms”, *Proceedings of the National Academy of Sciences*, vol. 100, no. 6, pp. 3351–3356, 2003 (<https://doi.org/10.1073/pnas.0530258100>).
- [15] Z. Chen, Z. Ding, X. Dai, and R. Schober, “Asymptotic Performance Analysis of GSVD-NOMA Systems with a Large-Scale Antenna Array”, *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 575–590, 2019 (<https://doi.org/10.1109/TWC.2018.2883102>).
- [16] C. Rao, Z. Ding, and X. Dai, “GSVD-Based MIMO-NOMA Security Transmission”, *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1484–1487, 2021 (<https://doi.org/10.1109/LWC.2021.3071365>).
- [17] S. Ali., A. Fakoorian, and A.L. Swindlehurst, “Dirty Paper Coding versus Linear GSVD-Based Precoding in MIMO Broadcast Channel

- with Confidential Messages”, in: *2011 IEEE Global Telecommunications Conference – GLOBECOM 2011*, Houston, USA, 2011 (<https://doi.org/10.1109/GLOCOM.2011.6134129>).
- [18] W. Mei, Z. Chen, and J. Fang, “GSVD-based Precoding in MIMO Systems with Integrated Services”, *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1528–1532, 2016 (<https://doi.org/10.1109/LSP.2016.2606349>).
- [19] C. Rao, Z. Ding, and X. Dai, “Application of GSVD-based Precoding in MIMO-NOMA Relaying Systems”, *IET Communications*, vol. 14, no. 21, pp. 3802–3812, 2020 (<https://doi.org/10.1049/iet-com.2020.0555>).
- [20] M.F. Hanif and Z. Ding, “Robust Power Allocation in MIMO-NOMA Systems”, *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1541–1545, 2019 (<https://doi.org/10.1109/LWC.2019.2926277>).
- [21] Y. Dursun, F. Fang, and Z. Ding, “Hybrid NOMA based MIMO Offloading for Mobile Edge Computing in 6G Networks”, *China Communications*, vol. 19, no. 10, pp. 12–20, 2022 (<https://doi.org/10.23919/JCC.2022.00.024>).
- [22] Y. Dursun, M.B. Goktas, and Z. Ding, “Green NOMA based MU-

- MIMO Transmission for MEC in 6G Networks”, *Computer Networks*, vol. 228, art. no. 109749, 2023 (<https://doi.org/10.1016/j.comnet.2023.109749>).
- [23] Y. Dursun, K. Wang, and Z. Ding, “Secrecy sum rate maximization for a MIMO-NOMA uplink transmission in 6G networks”, *Physical Communication*, vol. 53, art. no. 101675, 2022 (<https://doi.org/10.1016/j.phycom.2022.101675>).
- [24] F. Fang *et al.*, “Optimal Resource Allocation for Delay Minimization in NOMA-MEC Networks”, *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7867–7881, 2020 (<https://doi.org/10.1109/TCOMM.2020.3020068>).
- [25] G. Li *et al.*, “Latency Minimization for IRS-Aided NOMA MEC Systems with WPT-enabled IoT Devices”, *IEEE Internet of Things Journal*, vol. 10, no. 14, pp. 12156–12168, 2023 (<https://doi.org/10.1109/JIOT.2023.3240395>).
- [26] S.S. Yilmaz and B. Ozbek, “Massive MIMO-NOMA Based MEC in Task Offloading for Delay Minimization”, *IEEE Access*, vol. 11, pp. 162–170, 2022 (<https://doi.org/10.1109/ACCESS.2022.3232731>).
- [27] Y. Zhang, Z. Na, Y. Wang, and C. Ji, “Joint power allocation and deployment optimization for HAP-assisted NOMA-MEC System”, *Wireless Networks*, pp. 1–13, 2022 (<https://doi.org/10.1007/s11276-022-03201-8>).
- [28] F. Guo *et al.*, “Joint Trajectory and Computation Offloading Optimization for UAV-assisted MEC with NOMA”, in: *IEEE INFOCOM 2019 – IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Paris, France, 2019, (<https://doi.org/10.1109/INFOCOMWKSHPS47286.2019.9093764>).
- [29] C. Zheng and W. Zhou, “Computation Bits Maximization in Backscatter-assisted Wireless-powered NOMA-MEC Networks”, *EURASIP J. on Wireless Communications and Networking*, art. no. 23, 2022 (<https://doi.org/10.1186/s13638-022-02097-4>).
- [30] Y. Xu, T. Zhang, Y. Zou, and Y. Liu, “Reconfigurable Intelligence Surface Aided UAV-MEC Systems with NOMA”, *IEEE Communications Letters*, vol. 26, no. 9, pp. 2121–2125, 2022 (<https://doi.org/10.1109/LCOMM.2022.3183285>).
- [31] I. Altin and M. Akar, “A Joint Resource Allocation Method for Hybrid NOMA MEC Offloading”, *Physical Communication*, vol. 54, art. no. 101809, 2022 (<https://doi.org/10.1016/j.phycom.2022.101809>).
- [32] L. Lin, W. Zhou, Z. Yang, and J. Liu, “Deep Reinforcement Learning-based Task Scheduling and Resource Allocation for NOMA-MEC in Industrial Internet of Things”, *Peer-to-Peer Networking and Applications*, vol. 16, no. 1, pp. 170–188, 2023 (<https://doi.org/10.1007/s12083-022-01348-x>).
- [33] C. Li, H. Wang, and R. Song, “Mobility-Aware Offloading and Resource Allocation in NOMA-MEC Systems via DC”, *IEEE Communications Letters*, vol. 26, no. 5, pp. 1091–1095, 2022 (<https://doi.org/10.1109/LCOMM.2022.3154434>).
- [34] T.V. Truong and A. Nayyar, “System Performance and Optimization in NOMA Mobile Edge Computing Surveillance Network using GA and PSO”, *Computer Networks*, vol. 223, art. no. 109575, 2023 (<https://doi.org/10.1016/j.comnet.2023.109575>).
- [35] M. Chen, Y. Wan, M. Wen, and T. Zhou, “Fairness Optimization in IRS-assisted MEC Computational Offloading”, *Physical Communication*, vol. 54, art. no. 101855, 2022 (<https://doi.org/10.1016/j.phycom.2022.101855>).
- [36] C.-L. Wang, Y.-C. Ding, Y.-C. Wang, “A Low-Complexity Power Allocation Scheme for MIMO-NOMA Systems with Imperfect Channel Estimation”, in: *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Kyoto, Japan, pp. 234–239, 2022 (<https://doi.org/10.1109/PIMRC54779.2022.9977487>).
- [37] L. Khamidullina, A.L.F. de Almeida, and M. Haardt, “Multilinear Generalized Singular Value Decomposition (ML-GSVD) and its Application to Multiuser MIMO Systems”, *IEEE Transactions on Signal Processing*, vol. 70, pp. 2783–2797, 2022 (<https://doi.org/10.1109/TSP.2022.3178902>).
- [38] L. Khamidullina, A.L.F. de Almeida, and M. Haardt, “Rate Splitting and Precoding Strategies for Multi-User MIMO Broadcast Channels with Common and Private Streams”, in: *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023 (<https://doi.org/10.1109/ICASSP49357.2023.10095138>).

#### Yunus Dursun, Ph.D.

School of Electrical and Electronic Engineering

 <https://orcid.org/0000-0003-3405-8369>

E-mail: yunus.dursun@manchester.ac.uk

University of Manchester, Manchester, United Kingdom

<https://www.manchester.ac.uk>

# Performance Comparison of Several Algorithms for Localization of Wideband Sources

Hassan Ougraz<sup>1</sup>, Said Safi<sup>1</sup>, Ahmed Boumezzough<sup>1</sup>, and Miloud Frikel<sup>2</sup>

<sup>1</sup>University of Sultan Moulay Slimane Beni Mellal, Morocco,

<sup>2</sup>National Graduate School of Engineering and Research Center Caen, France

<https://doi.org/10.26636/jtit.2023.3.1359>

**Abstract** — In recent years, researchers have tried to estimate the direction-of-arrival (DOA) of wideband sources and several novel techniques have been proposed. In this paper, we compare six algorithms for calculating the DOA of broadband signals, namely coherent subspace signal method (CSSM), two-sided correlation transformation (TCT), incoherent multiple signal classification (IMUSIC), test of orthogonality of frequency subspaces (TOFS), test of orthogonality of projected subspaces (TOPS), and squared TOPS (S-TOPS). The comparison is made through computer simulations for different parameters, such as signal-to-noise ratio (SNR), in order to establish the efficiency and performance of the discussed methods in noisy environments. CSSM and TCT require initial values, but the remaining approaches do not need any preprocessing.

**Keywords** — coherent, DOA estimation, incoherent, localization, wideband sources

## 1. Introduction

Many applications rely on sensor arrays. These include radar signal processing [1]–[3], sonar, medical imaging, wireless communication systems [4]–[7], and Internet of Things [8]–[10]. In the simplest scenario, signals received by the sensors are scaled and contain delayed replicas of the waveform emitted by a single source. In a complex case, numerous sources and multiple propagation pathways from the sources to the sensors may exist. In the Fourier domain, the most frequent strategy is to decompose the broadband source into narrowband elements [11]–[12]. Moreover, localization of wideband sources is primarily based on figuring out how to employ numerous covariance matrices at various frequencies to construct precise DOA estimation [11], [13]. A variety of approaches have been suggested to fix the problem of the direction-of-arrival (DOA) estimation of various broadband signals. These are categorized as incoherent signal subspace algorithms [14], [15] and coherent signal subspace algorithms [16], [17].

In this paper, we discuss the capabilities of some algorithms used for estimating the location of wideband signal sources. Being one of the fundamental wideband DOA estimation methods, incoherent methods rely on numerous narrowband sources that have been incoherently extracted from a broad-

band source [18]. In particular, these techniques apply independently narrowband approaches, such as multiple signal classification (MUSIC) [19], to signal sources. The first approach is incoherent MUSIC (IMUSIC). Its prediction resolution improves in high signal-to-noise ratio (SNR) zones and degrades when the SNR of several frequencies is minimal. Moreover, weak estimations from several bands of frequencies will have an impact on the performance of the overall estimation. The second approach is the test of orthogonality of frequency subspaces (TOFS) [20] which generates steering vectors for each DOA and frequency available. When SNR is high, the estimation accuracy is quite good, since TOFS is an incoherent technique [20]. However, when SNR is low, TOFS is unable to resolve the desired DOAs. Another algorithm, known as the test of orthogonality of projected subspaces (TOPS) [21], applies the signal and noise subspaces of various frequencies to offer better DOA estimation performance without the requirement to provide any starting values. Thus, TOPS fills the gap between coherent and incoherent methods. This technique offers the following advantages [21]:

- it does not require a beamforming matrix or focusing angles,
- when SNR is high, it is not biased,
- at low SNR, it merges frequency bands more efficiently than incoherent approaches.

The drawback of such an approach is that it is challenging to determine the real DOA of sources, since the spatial spectra computed using the TOPS method contain several incorrect lobes. The squared TOPS (S-TOPS) approach [28] employs the signal subspace and noise subspace twice and achieves the squared test of orthogonality of both subspaces. S-TOPS outperforms TOPS at small SNR values, since its resolution has been enhanced.

The coherent signal subspace method (CSSM) has been suggested to resolve these drawbacks and to estimate DOA [16]. The covariance matrix of each frequency band is focused by transformation arrays and the focusing arrays are averaged to produce a novel covariance matrix in the coherent signal subspace approaches. The main point of this approach is to precisely focus these covariance matrices. The two-sided correlation transformation (TCT) approach, an alternative of CSSM, is proposed to improve performance.

Various approaches for extracting a proper focusing array have been suggested, e.g. in [23], [24]. Nevertheless, each focusing algorithm entails starting values which are the predetermined angles of arriving sources, and the efficiency of coherent signal subspace methods is related to an initial value [25].

In this paper, we concentrate on the comparison of the techniques mentioned above, with various SNR values, different snapshots and for various angles of sources, in order to provide a more accurate comparison in terms of the advantages and drawbacks of these approaches. The remaining part of this paper has the following structure. In Section 3, the wideband model and the problem of localization will be investigated. The wideband methods are investigated in Section 4, while the experimental results are discussed in Section 5. Finally, Section 6 presents the conclusions and future outlooks.

## 2. Related Works

Many papers focusing on localization and wideband DOA estimation for incoherent sources are closely related to our work in this domain. The first is [13] as it proposes a novel approach to estimating the wideband sources, namely weighted squared TOPS, and confirms that it outperforms other methods. Paper [13] does something similar to this specific work. However, we present a comparative study using IMUSIC, TOFS, TOPS, S-TOPS, CSSM, and TCT to estimate the wideband sources, and to determine which one of them performs the best. Additionally, the comparative study of three wideband DOA estimation algorithms (TCT, TOPS, W-CMSR), as presented in [11], demonstrates that W-CMSR surpasses both TCT and TOPS in terms of resolution capability.

## 3. Problem Formulation

The  $M$  elements of a uniform linear array (ULA) are evenly spaced, with the distance  $d$  between succeeding elements being no more than half of the wavelength. We assume that  $L$  wideband sources ( $L < M$ ) are either known or can be determined [26] and are impinging on the array elements from different angles  $(\theta_1, \theta_2, \dots, \theta_L)$ . Figure 1 shows the ULA geometry for the localization of broadband signals.

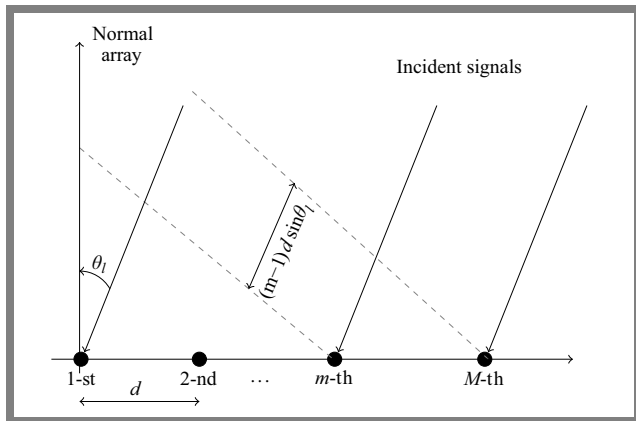


Fig. 1. ULA geometry of  $M$  antennas [11].

Our goal is to estimate the direction of arrival of  $L$  wideband sources using an ULA of  $M$  antennas supposing that all sources are uncorrelated and occur in the bandwidth between the smallest  $\omega_s$  and the largest  $\omega_L$  source frequency. Thus, the received data at the  $m$ -th antenna are given as follows:

$$x_m(t) = \sum_{l=1}^L s_l(t - v_m \sin \theta_l) + n_m(t), \quad (1)$$

where  $s_l(t)$  is the  $l$ -th source,  $n_m(t)$  is an additive white Gaussian noise (AWGN) at the  $m$ -th component,  $v_m = (m-1)d/c$ , where  $d$  is the distance between the components of ULA, and  $c$  is the velocity of light.  $\theta_l$  is the angle to be computed. The received data is split into  $K$  narrowband signals. Thus, the discrete Fourier transform (DFT) of the received signals at  $m$ -th element is:

$$x_m(\omega) = \sum_{l=1}^L s_l(\omega) e^{-j\omega v_m \sin \theta_l} + n_m(\omega). \quad (2)$$

The DFT output signals can then be represented in vector form [11]:

$$x(\omega_i) = A(\omega_i, \theta) s(\omega_i) + n(\omega_i), \quad i = 1, 2, \dots, K, \quad (3)$$

where  $\omega_s < \omega_i < \omega_L$  for  $i = 1, 2, \dots, K$ ,

$$A(\omega_i, \theta) = [a(\omega_i, \theta_1) \ a(\omega_i, \theta_2) \ \dots \ a(\omega_i, \theta_L)], \quad (4)$$

$$a(\omega_i, \theta_l) = \begin{bmatrix} 1 \\ e^{-j\omega_i v_1 \sin \theta_l} \\ \vdots \\ e^{-j\omega_i v_{M-1} \sin \theta_l} \end{bmatrix}, \quad (5)$$

where  $T$  is the transpose of a given matrix.

The correlation matrix is calculated in the following manner [11]:

$$\begin{aligned} R_{xx}(\omega_i) &= E[x(\omega_i)x^H(\omega_i)] \\ &= A(\omega_i, \theta) R_{ss}(\omega_i) A^H(\omega_i, \theta) + \sigma_n^2 I, \end{aligned} \quad (6)$$

where  $E[\cdot]$  is the expectation value operator,  $H$  is the Hermitian operator,  $R_{ss}(\omega_i) = E[s(\omega_i)s^H(\omega_i)]$ ,  $\sigma_n^2$  is the noise power, and  $I$  is the  $M \times M$  unit matrix. Supposing that  $R_{ss}(\omega_i)$  has a complete rank, the signal subspace  $F_s(\omega_i)$  and the noise subspace  $F_n(\omega_i)$  matrices at the frequency  $\omega_i$  can be generated using the eigen-values decomposition (EVD) or the singular value decomposition (SVD) of the correlation array as:

$$F_s(\omega_i) = [e_1(\omega_i), e_2(\omega_i), \dots, e_L(\omega_i)], \quad (7)$$

$$F_n(\omega_i) = [e_{L+1}(\omega_i), e_{L+2}(\omega_i), \dots, e_M(\omega_i)], \quad (8)$$

where  $e_1(\omega_i), \dots, e_M(\omega_i)$  are the perpendicular eigenvectors of  $R_{xx}(\omega_i)$ , ranked in decreasing order by their corresponding eigen-values as follows:

$$\lambda_1(\omega_i) \geq \dots \geq \lambda_L(\omega_i) \geq \lambda_{L+1}(\omega_i), \dots, \lambda_M(\omega_i) = \sigma_n^2. \quad (9)$$

## 4. Wideband Algorithms

In this part, we investigate several techniques for localization of broadband signals. As described in Section 3, DFT is used to break down every wideband source into  $K$  narrowband signals.

### 4.1. Coherent Signal Subspace Method (CSSM)

Many strategies have been proposed to address the drawbacks of incoherent methods by finding a good way to integrate the signal subspaces coherently into one general correlation matrix in which narrowband high-resolution methods may be used. CSSM, proposed by Wang and Kaveh [27], was the first technique to coherently sum the correlation matrices of different frequency bins [27]. It converts the correlation matrices at many frequency bins into a single general correlation matrix at a single focusing frequency using a frequency-dependent transformation matrix (focusing matrix). The focusing matrix  $F$  is intended to provide the solution to the following equation:

$$F_i A_i = A_0, \quad i = 1, \dots, K. \quad (10)$$

To estimate the angles-of-arrival that may be employed to generate  $A_0$ , a pre-process is required. The observation vectors at different frequency bins are converted into the focusing subspace using the focusing matrices. The standard proposed in [27] is based on the transformation matrices produced from the constrained minimization problem:

$$\begin{cases} \min_{F_i} \| A_0 - F_i A_i \| \\ F_i^H F_i = I \end{cases}. \quad (11)$$

Hung *et al.* established in [14] that the focusing matrix that solves the problem given by Eq. (11) is:

$$F_i^{CSSM} = V_i W_i^H. \quad (12)$$

where  $V_i \Sigma_i W_i^H$  is the singular values decomposition of  $A_0 A_i^H$ . Following the determination of the focusing matrix, a high resolution approach such as MUSIC [19] is employed to determine the DOAs. The DOAs are calculated by determining the angular position of peaks in the MUSIC algorithm's spatial spectrum.

#### Algorithm 1. CSSM algorithm.

- 1: Use an ordinary beamformer to scan the space and find an initial estimate of the DOAs.
- 2: Apply a DFT to the array output to sample the spectrum of data.
- 3: **For**  $i = 1, 2, \dots, K$  **do**
- 4: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 5: **End for**
- 6: **For**  $i = 1, 2, \dots, K$  **do**
- 7: Determine the focusing matrices by using the Eq. (12).
- 8: **End for**

- 9: Apply MUSIC or any other high-resolution spectral estimation method to find the DOAs.

### 4.2. TCT Method

Two-sided correlation transformation (TCT) [29] is similar to CSSM introduced by Hung *et al.* in that signal subspaces are transformed using focusing matrices. The TCT transforms the correlation matrix using a two-sided unitary transformation. Let us consider the covariance matrix at the  $i$ -th frequency band in a noise-free floor:

$$P_i = A_i R_{ss}(\omega_i) A_i^H, \quad (13)$$

and  $P_0$  is the focusing noise free covariance matrix. The TCT focusing matrices can be found by solving the minimization problem:

$$\begin{cases} \min_{F_i} \| P_0 - F_i P_i F_i^H \| \\ F_i^H F_i = I \end{cases}. \quad (14)$$

It is shown in [29] that the optimal solution of the problem is given by the eigenvectors of the covariance matrix at the frequencies  $\omega_0$  and  $\omega_k$  as:

$$F_k^{TCT} = V_0 V_i^H, \quad (15)$$

where  $V_0$  and  $V_i$  are the eigenvector matrices of  $P_0$  and  $P_i$ , respectively.

#### Algorithm 2. TCT algorithm.

- 1: Use an ordinary beamformer to scan the space and find an initial estimate of the DOAs.
- 2: Split the data into  $K$  blocks of the same size.
- 3: Apply a DFT to the array output to sample the spectrum of data.
- 4: **For**  $i = 1, 2, \dots, K$  **do**
- 5: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 6: Form  $A_i$ , and  $S_i$ , using the results of the preprocessing step and Eq. (14).
- 7: **End for**
- 8: Average the source correlation matrices to obtain  $S_0$  as in [30].
- 9: Find  $P_0 = A_0 S_0 A_0^H$ .
- 10: **For**  $i = 1, 2, \dots, K$  **do**
- 11: Calculate  $P_i$  by using Eq. (13).
- 12: **End for**
- 13: Determine the unitary transformation matrices.
- 14: Multiply these matrices by the sample correlation matrices, and average the results.
- 15: Apply MUSIC or any other high-resolution spectral estimation method to find the DOAs.
- 16: To improve the performance, iterate steps 3 to 5.

### 4.3. IMUSIC Approach

Because the narrowband MUSIC technique is used for all frequencies simultaneously, this approach is known as incoherent

MUSIC (IMUSIC) [13] and is considered to be one of the most fundamental DOA estimation techniques for wideband signals. Thus, IMUSIC calculates the DOA of broadband signals using the following formula:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^K a^H(\omega_i, \theta) F_n(\omega_i) F_n^H(\omega_i) a(\omega_i, \theta), \quad (16)$$

where  $F_n(\omega_i)$  is the noise subspace at every frequency  $\omega_i$ . The noise subspace array  $F_n(\omega_i)$  is obtained from the covariance matrix  $R_{xx}(\omega_i)$  [21]. As long as the DOAs computed by Eq. (16), IMUSIC results are the averages of the outcomes of each range of frequencies. Weak estimations from one frequency range degrade the overall estimation accuracy.

#### Algorithm 3. IMUSIC algorithm.

- 1: Split the data into  $K$  blocks of the same size.
- 2: Calculate the DFT of the  $K$  blocks.
- 3: **For**  $i = 1, 2, \dots, K$  **do**
- 4: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 5: **End for**
- 6: **For**  $i = 1, 2, \dots, K$  **do**
- 7: Calculate the noise subspace  $F_n(\omega_i)$  of each frequency  $\omega_i$  by EVD/SVD of  $R_{xx}$ .
- 8: **End for**
- 9: Estimate the final DOA  $\hat{\theta}$  using the Eq. (16).

IMUSIC is generally effective in noisy environments, and when the signals are well separated from one to another. However, when the SNR is low, it suffers from failures and generates side lobes at erroneous directions, which occurs in many situations. Additionally, the noise level is expected to be uniform across the band of frequencies, which is not usual [21]. The limitations of incoherent approaches encouraged the development of coherent approaches that are capable of addressing these drawbacks.

#### 4.4. TOFS Approach

For each frequency, the TOFS approach utilizes the noise subspace created by the EVD/SVD of the covariance matrix [20]. To compute the DOA of every arriving broadband signal, TOFS relies on orthogonality between the steering array and the noise subspace. Since represents the single DOA of arriving broadband signals, then  $\theta$  satisfies the formula [30]:

$$a^H(\omega_i, \theta) F_n(\omega_i) F_n^H(\omega_i) a(\omega_i, \theta) = 0. \quad (17)$$

The vector  $D_i(\theta)$  is defined as [13]:

$$D_i(\theta) = \begin{bmatrix} a^H(\omega_1, \theta) F_n(\omega_1) F_n^H(\omega_1) a(\omega_1, \theta) \\ a^H(\omega_2, \theta) F_n(\omega_2) F_n^H(\omega_2) a(\omega_2, \theta) \\ \vdots \\ a^H(\omega_K, \theta) F_n(\omega_K) F_n^H(\omega_K) a(\omega_K, \theta) \end{bmatrix}^T. \quad (18)$$

Since the DOA of arriving broadband sources is  $\theta$ , all elements of the vector  $D_i(\theta)$  are zero.

#### Algorithm 4. TOFS algorithm [20].

- 1: Split the data into  $K$  blocks of the same size.
- 2: Calculate the DFT of the  $K$  blocks.
- 3: **For**  $i = 1, 2, \dots, K$  **do**
- 4: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 5: **End for**
- 6: **For**  $i = 1, 2, \dots, K$  **do**
- 7: Calculate the noise subspace  $F_n(\omega_i)$  of each frequency  $\omega_i$  by EVD/SVD of  $R_{xx}$ .
- 8: Generate  $D_i(\theta)$  by Eq. (18) for each hypothesized DOA  $\theta$ .
- 9: **End for**
- 10: Estimate the final DOA  $\hat{\theta}$  by:

$$\hat{\theta} = \arg \max_{\theta} \frac{1}{\sigma_{\min}(\theta)}. \quad (19)$$

where  $\sigma_{\min}(\theta)$  is the shortest singular value of  $D_i(\theta)$ .

By employing the noise subspaces derived from the covariance matrix, TOFS offers better DOA estimation in noisy environments. However, when the SNR is low TOFS is unable to resolve closely spaced targets.

#### 4.5. TOPS Approach

TOPS calculates the DOA of arriving broadband sources by employing the signal subspace and noise subspace of each frequency range [21]. The first step is to extract the signal subspace  $F_s(\omega_i)$  and the noise subspace  $F_n(\omega_i)$  from the EVD/SVD of the covariance array of each range of frequencies. Thus, one range of frequencies  $\omega_i$  should be opted, and the signal subspace  $F_s(\omega_i)$  of the opted range should be transformed into other bands. A diagonal unitary transformation array is used by TOPS. The  $m$ -th component on the diagonal of the frequency transform array  $\Psi(\omega_i, \theta)$  is defined as:

$$[\Psi(\omega_i, \theta)]_{(m,m)} = e^{-j\omega_i \frac{md}{c} \sin \theta}. \quad (20)$$

The signal subspace  $F_s(\omega_i)$  of the band of frequencies  $\omega_i$  is transformed into the other frequency band  $\omega_j$  by using  $\Psi(\omega_i, \theta)$ , where the transformed signal subspace  $U_{ij}(\theta)$  is [21]:

$$U_{ij}(\theta) = \Psi(\Delta\omega, \theta) F_s(\omega_i), \quad i \neq j, \quad (21)$$

where  $\Delta\omega = \omega_j - \omega_i$ .

The Eq. (21) can be written as follows:

$$\begin{aligned} U_{ij}(\theta) &= \Psi(\Delta\omega, \theta) A(\omega_i, \theta) G_i \\ &= A(\omega_j, \hat{\theta}) G_i, \end{aligned} \quad (22)$$

where  $\hat{\theta}$  is the converted  $\theta$  by using the frequency transform matrix  $\Psi(\omega_i, \theta)$ , and  $G_i$  is a complete rank square array that proves  $F_s(\omega_i) = A(\omega_i, \theta) G_i$ . An array manifold at any frequency and direction can be converted into another array

manifold at a different frequency by using this transformation process. Consequently, the transformed matrix has a complete rank and can be employed in the following orthogonality test between transformed array and noise subspace, investigated in detail in [21]. Given that the frequency range of relevance is  $\omega_1$ , the estimator  $D'(\theta)$  is given as [30]:

$$D'_i(\theta) = \begin{bmatrix} U_{12}^H(\theta) F_n(\omega_2) \\ U_{13}^H(\theta) F_n(\omega_3) \\ \vdots \\ U_{1K}^H(\theta) F_n(\omega_K) \end{bmatrix}^T. \quad (23)$$

When  $\theta$  is one of the arriving wideband signals, the rank of the matrix  $D'_i(\theta)$  also degrades same as TOFS. The performance of DOA estimation is dictated by the correctness of generating the covariance array, that is largely determined by the number of snapshots and the SNR of the received data. TOPS employs the subspace projection strategy to decrease leakage of signal subspace components in the predicted noise subspace. The projection array  $P_i(\theta)$  is then computed as [30]:

$$P_i(\theta) = I - (a^H(\omega_i, \theta) a(\omega_i, \theta))^{-1} a(\omega_i, \theta) a^H(\omega_i, \theta), \quad (24)$$

where  $I$  is an  $M \times M$  identity matrix. The noise robust matrix  $D''(\theta)$  is obtained by substituting the component  $U_{ij}(\theta)$  of Eq. (23) with a novel transformed signal subspace array  $U'_{ij}(\theta)$ .

$$U'_{ij}(\theta) = P_j(\theta) U_{ij}(\theta), \quad (25)$$

$$D''_i(\theta) = \begin{bmatrix} U_{12}'^H(\theta) F_n(\omega_2) \\ U_{13}'^H(\theta) F_n(\omega_3) \\ \vdots \\ U_{1K}'^H(\theta) F_n(\omega_K) \end{bmatrix}^T. \quad (26)$$

Since the projection matrix  $P_i(\theta)$  eliminates subspace prediction errors, TOPS shows better performance when the Eq. (26) with  $D''_i(\theta)$  is used.

**Algorithm 5.** TOPS algorithm [21].

- 1: Split the data into  $K$  blocks of the same size.
- 2: Calculate the DFT of the  $K$  blocks.
- 3: **For**  $i = 1, 2, \dots, K$  **do**
- 4: Select  $x(\omega_i)$  at preselected  $\omega_i$ .
- 5: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 6: **End for**
- 7: **For**  $i = 1, 2, \dots, K$  **do**
- 8: Calculate the signal subspace  $F_s(\omega_1)$ .
- 9: Compute the noise subspace  $F_n(\omega_i)$  by EVD/SVD of the calculated correlation matrix  $R_{xx}$ .
- 10: Generate  $D''_i(\theta)$  using Eqs. (25) and (26) for each hypothesized DOA  $\theta$ .

**11: End for**

**12:** Estimate the final DOA  $\hat{\theta}$  by

$$\hat{\theta} = \arg \max_{\theta} \frac{1}{\sigma_{\min}(\theta)}, \quad (27)$$

where  $\sigma_{\min}(\theta)$  is the shortest singular value of  $D''_i(\theta)$ .

The output data of a DFT or bandpass filter is not necessarily an accurate narrowband signal. Moreover, the filtered signals may degrade DOA prediction performance. The TOPS approach may minimize these drawbacks by applying a particular signal subspace generated by the measured covariance matrix, rather than the steering array of the frequency range. This signifies that the approach used to select the range of frequencies from which the signal subspace is converted by Eq. (21) impacts the localization of wideband sources. On the other hand, the residual error of the subspaces results in certain unwanted rank reductions of the matrix  $D''_i(\theta)$ . Consequently, TOPS suffers from a considerable limitation when it comes to generating certain spurious peaks in the spectrum simulated by Eq. (27).

#### 4.6. S-TOPS Approach

To enhance DOA estimation performance, squared TOPS (S-TOPS) incorporates two strategies into the TOPS method. The first is to determine the frequency band from which the signal subspace will be employed. The other is a strategy to increase the sensitivity of reduction rank of the estimator  $D''_i(\theta)$  when  $\theta$  is the desired DOA of the arriving broadband signals. The band of frequencies with the maximum SNR should operate as the reference frequency, which is considered to be the frequency range in which the signal subspace will be employed. S-TOPS employs the range of frequencies with the difference between the lowest signal eigen-value  $\lambda_L(\omega_i)$  and the highest noise eigen-value  $\lambda_{L+1}(\omega_i)$  is higher than in the reference frequency [22]. Then, Eq. (21) transforms the signal subspace of the reference range of frequencies into another range of frequency. Suppose that the range of frequencies  $\omega_i$  is chosen, and the signal subspace  $F_s(\omega_i)$  is converted into another frequency band  $\omega_j$ . We generate the matrix  $Q_i(\theta)$  for testing the squared orthogonality of projected subspaces using the converted signal subspace  $U'_{ij}(\theta)$  and the noise subspace  $F_n(\omega_j)$ , as follows:

$$Q_i(\theta) = \begin{bmatrix} U_{12}'^H F_n(\omega_2) F_n(\omega_2)^H U_{12}' \\ U_{13}'^H F_n(\omega_3) F_n(\omega_3)^H U_{13}' \\ \vdots \\ U_{1K}'^H F_n(\omega_K) F_n(\omega_K)^H U_{1K}' \end{bmatrix}^T. \quad (28)$$

**Algorithm 6.** S-TOPS algorithm.

- 1: Split the data into  $K$  blocks of the same size.
- 2: Calculate the DFT of the  $K$  blocks.
- 3: **For**  $i = 1, 2, \dots, K$  **do**
- 4: Select  $x(\omega_i)$  at preselected  $\omega_i$ .

- 5: Calculate the correlation matrix  $R_{xx}$  of every frequency  $\omega_i$ .
- 6: End for
- 7: For  $i = 1, 2, \dots, K$  do
- 8: Calculate the signal subspace  $F_s(\omega_1)$ .
- 9: Compute the noise subspace  $F_n(\omega_i)$  by EVD/SVD of the calculated correlation matrix  $R_{xx}$ .
- 10: Generate  $D_i''(\theta)$  using Eqs. (25) and (26) for each hypothesized DOA  $\theta$ .
- 11: Calculate the squared orthogonality of projected subspaces  $Q_i(\theta)$  using Eq. (28) for each hypothesized DOA  $\theta$ .
- 12: End for
- 13: Estimate the final DOA  $\hat{\theta}$  by:

$$\hat{\theta} = \arg \max_{\theta} \frac{1}{\sigma_{\min}(\theta)}, \quad (29)$$

where  $\sigma_{\min}(\theta)$  is the shortest singular value of  $Q_i(\theta)$ .

When  $\theta$  is the DOA of the arriving broadband signal, the row and column components of the matrix  $Q_i(\theta)$  produced by the squared process should be close to zero. This demonstrates that the procedure enhances the sensitivity of estimating rank reduction of the orthogonality evaluation matrix. Hence, it improves the estimated resolution of S-TOPS. Nevertheless, the unwanted lobes in the spectra still exist.

## 5. Simulation Results

In this part, several simulations were run to evaluate the efficiency of these approaches. Matlab 2020a has been employed to demonstrate the performance of these approaches. The received data are split into  $K = 256$  blocks, with the number of snapshots in each block equaling 100, using an ULA array of 10 sensors to show various scenarios.  $\lambda$  is the wavelength that refers to the center frequency  $\omega_0$ . The signals share identical frequency bands. Table 1 shows the parameters used in the simulations.

**Tab. 1.** Simulation parameters.

| Description                | Symbols       | Value                     |
|----------------------------|---------------|---------------------------|
| Antennas spacing           | $d$           | $\lambda/2$               |
| Number of sources          | $L$           | 3                         |
| Number of antennas         | $M$           | 10                        |
| Smallest frequency         | $\omega_S$    | 2 MHz                     |
| Largest frequency          | $\omega_L$    | 4 MHz                     |
| Center frequency           | $\omega_0$    | $(\omega_S + \omega_L)/2$ |
| Nyquist sampling frequency | $\omega_{NS}$ | 10 MHz                    |

### 5.1. Spatial Spectrum

For each hypothesis, the spectra depict the results achieved with the use of the DOA estimation methods. The estimated

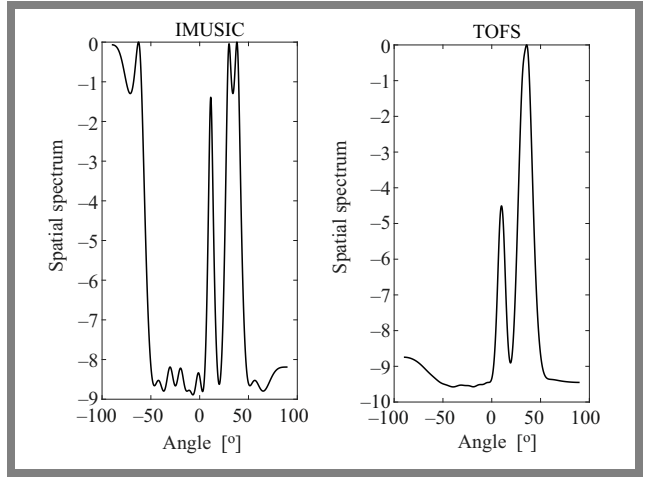


**Fig. 2.** Spatial spectra of the IMUSIC and TOFS algorithms at  $M = 10$ ,  $L = 3$  ( $10^\circ$ ,  $30^\circ$ ,  $38^\circ$ ), and  $\text{SNR} = 0$  dB.

DOAs are indicated by the peaks. For the simulation scenarios, Figs. 2, 3, and 4 depict the spatial spectra of all methods mentioned in this paper for a ten-sensor ULA. The spatial spectra are estimated at the angles of  $\theta_1 = 10^\circ$ ,  $\theta_2 = 30^\circ$ , and  $\theta_3 = 38^\circ$  to illustrate the capabilities of each technique in a noisy environment with  $\text{SNR} = 0$  dB. The sources are detected by all methods, which results in high peaks of real DOAs. TOFS, CSSM, and TCT eliminate all false peaks. The spatial spectra indicate false lobes at erroneous angles that can be detected as DOAs using the TOPS and S-TOPS methods, as opposed to smaller peaks at true DOAs. The IMUSIC and TOFS techniques that can be used for reconstructing the arriving signals and TCT reach higher power values than the remaining approaches.

From Fig. 2, when SNR is 0 dB, it is clear that the TOFS algorithm is incapable of resolving the nearest sources ( $30^\circ$ ,  $38^\circ$ ), while the IMUSIC method can detect the two closest sources. However, it suffers from side fluctuations at approx.  $-60^\circ$ . Figure 3 indicates that all the estimation approaches have solved the true DOAs successfully, with some fake peaks generated between the calculated DOAs. Figure 4 illustrates the spectra of CSSM and TCT approaches, when SNR is 0 dB. TCT performs the best when SNR is low, without showing any false peaks, whereas CSSM cannot detect closely spaced signal sources (difference of  $8^\circ$ ).

According to simulation results, TCT, TOPS, S-TOPS, and IMUSIC are capable of detecting two closely separated targets with a greater degree of precision in mid-to-high SNR conditions. TOPS and S-TOPS both help us improve average performance in the entire SNR range, with an advantage over other methods in low SNR, while S-TOPS shows small fake peaks compared with the TOPS algorithm.

In terms of simplicity, TOFS is similar to IMUSIC, however TOPS and S-TOPS still have significant computational costs due to the complicated calculations required. Note that CSSM and TCT methods require initial values.

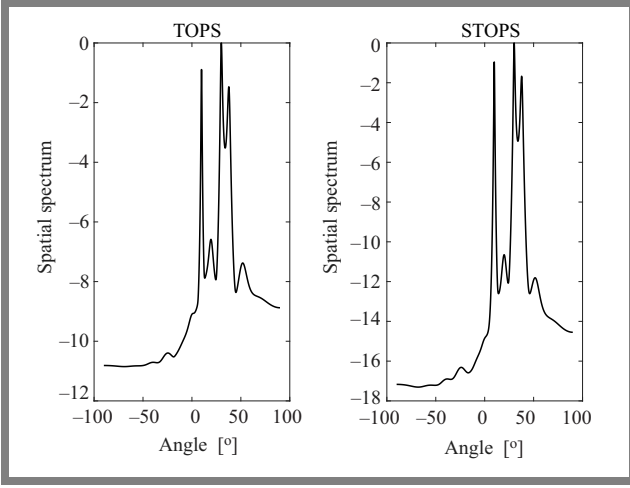


**Fig. 3.** Spatial spectra of the TOPS and S-TOPS algorithms at  $M = 10$ ,  $L = 3$  ( $10^\circ$ ,  $30^\circ$ ,  $38^\circ$ ), and SNR = 0 dB.

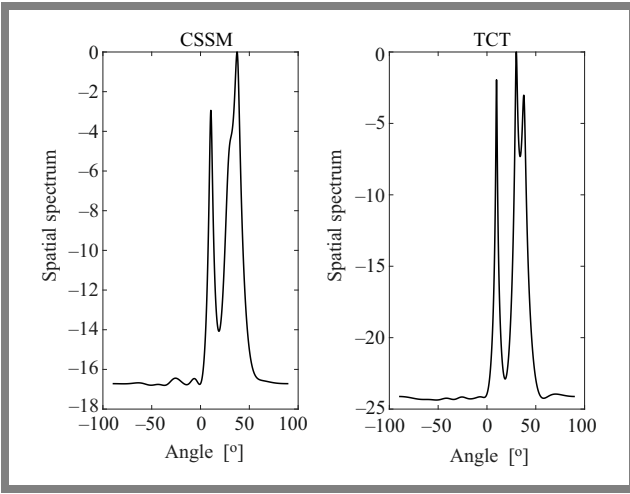


**Fig. 4.** Spatial spectra of the CSSM and TCT algorithms at  $M = 10$ ,  $L = 3$  ( $10^\circ$ ,  $30^\circ$ ,  $38^\circ$ ), and SNR = 0 dB.

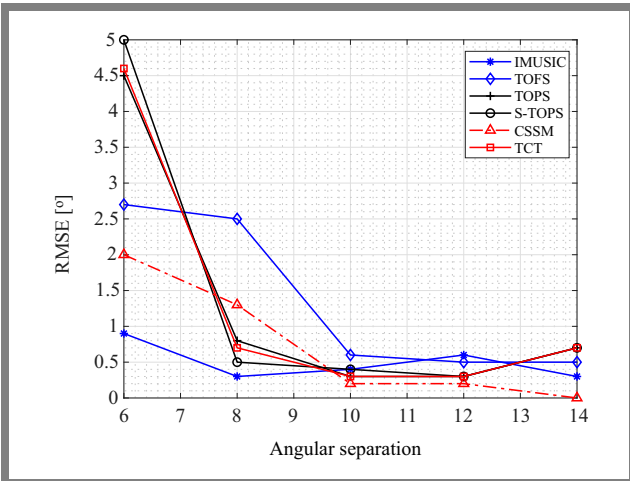


**Fig. 5.** RMSE against angular separation for SNR = 0 dB.

## 5.2. RMSE

For each simulation case, the root mean square error (RMSE) for the estimated DOAs of all sources is calculated as:

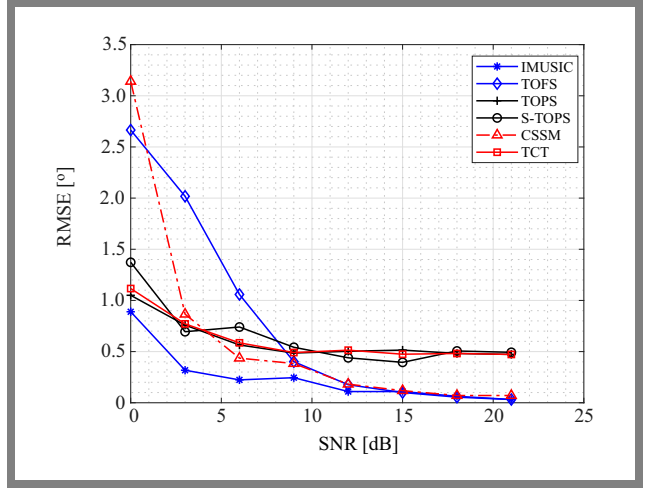


**Fig. 6.** RMSE versus SNR for the detected source  $38^\circ$ .

$$\text{RMSE} = \sqrt{\frac{1}{PL} \sum_{l=1}^L \sum_{n=1}^P (\hat{\theta}_l(n) - \theta_l)^2}, \quad (30)$$

where  $\theta_l$  is the  $l$ -th real DOA and  $\hat{\theta}_l(n)$  is the  $l$ -th calculated DOA at every simulation trial, where the number of trials is  $P = 10$ . An average of the number of arriving signals is obtained in order to calculate the overall errors for all predicted DOAs.

Figure 5 shows RMSE plots for IMUSIC, TOFS, TOPS, S-TOPS, CSSM, and TCT algorithms for different angular separations, with a changing value of the signal source  $\theta_3$ , with SNR being fixed – in this case at 0 dB – and the snapshot number equaling 100.

In Fig. 6, we examine the performance of IMUSIC, TOFS, TOPS, S-TOPS, CSSM, and TCT algorithms in terms of RMSE. Depicted in the Fig. 6 are the RMSE outcomes of the compared approaches versus SNR, where the average SNR ranges from 0 dB to 21 dB, and the snapshot number is 100. The associated DOA is  $\theta = 38^\circ$ .

In terms of RMSE, TOPS and S-TOPS performed well across the entire range, while IMUSIC and TOFS performed similarly and outperformed the other algorithms in mid-to-high SNR ranges. IMUSIC achieves nearly the same performance as TOFS under good conditions ( $\text{SNR} \geq 10$  dB). TOPS and S-TOPS provided average performance as the number of snapshots increased. With a high number of snapshots, IMUSIC and TOFS can effectively identify the right DOAs. As the angular separation varies, we consistently observe similar outcomes, with improved performance as the separation increases. Nevertheless, TOPS, S-TOPS, and TCT exhibit limitations when the angular separation between two signal sources is small.

## 6. Conclusion

This paper presents a detailed comparison of six high-resolution techniques used for calculating the DOA of wideband sources. For various SNR values, varied number of

snapshots and varied number of antennas, a comparison of six alternative methods for calculating DOA was provided. A ULA of ten antennas constitutes the receiver system. When the number of sources is known, computer simulations illustrate how well each algorithm performs. IMUSIC, TOPS, S-TOPS and TCT are capable of handling two closely spaced targets with a greater resolution in noisy environments (low SNR). With mid-to-high SNR values, TOFS, IMUSIC and CSSM perform best, whereas TOPS, S-TOPS and TCT show improved average performance in the entire SNR range. In the future, the authors will focus on deep/machine learning techniques to enhance the resolution of these approaches.

## References

- [1] S. Ebihara, Y. Kimura, T. Shimomura, R. Uchimura, and H. Choshi, "Coaxial-fed circular dipole array antenna with ferrite loading for thin directional borehole radar sonde", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1842–1854, 2015 (<https://doi.org/10.1109/TGRS.2014.2349921>).
- [2] U. Nielsen and J. Dall, "Direction-of-arrival estimation for radar ice sounding surface clutter suppression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5170–5179, 2015 (<https://doi.org/10.1109/TGRS.2015.2418221>).
- [3] R. Takahashi, T. Inaba, and T. Takahashi, "Digital Monopulse Beamforming for Achieving the CRLB for Angle Accuracy", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 315–323, 2018 (<https://doi.org/10.1109/TAES.2017.2756519>).
- [4] B. Cantrell, J. Rao, G. Tavik, and V. Krichevsky, "Wideband array antenna concept", in *IEEE International Radar Conference*, Arlington, USA, pp. 680–684, 2005 (<https://doi.org/10.1109/MAES.2006.1581120>).
- [5] Z.M. Liu, Z.T. Huang, and Y.Y. Zhou, "Direction-of-Arrival Estimation of Wideband Signals via Covariance Matrix Sparse Representation", *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4256–4270, 2011 (<https://doi.org/10.1109/TSP.2011.2159214>).
- [6] A. Toktas and A. Akdagli, "Compact multiple-input multiple-output antenna with low correlation for ultra-wideband applications", *IET Microwaves, Antennas and Propagation*, vol. 9, no. 8, pp. 822–829, 2015 (<https://doi.org/10.1049/iet-map.2014.0086>).
- [7] M. Frikel, S. Safi, and Y. Khmou, "Focusing Operators and Tracking Moving Wideband Sources", *Journal of Telecommunications and Information Technology*, no. 4, pp. 53–59, 2016 (<https://www.itl.waw.pl/czasopisma/JTIT/2016/4/53.pdf>).
- [8] I. Atanasov, A. Nikolov, and E. Pencheva, "An approach to transform Internet of Things data into knowledge", *International Journal of Embedded Systems*, vol. 9, no. 5, pp. 401–412, 2017 (<https://doi.org/10.1504/IJES.2017.086722>).
- [9] P. Thanapal and M.A.S. Durai, "A framework for computational offloading to extend the energy of mobile devices in mobile cloud computing", *International Journal of Embedded Systems*, vol. 9, no. 5, pp. 444–455, 2017 (<https://doi.org/10.1504/IJES.2017.086725>).
- [10] S. Narendrakumar *et al.*, "Token security for Internet of Things", *International Journal of Embedded Systems*, vol. 10, no. 4, pp. 334–343, 2018 (<https://doi.org/10.1504/IJES.2018.093689>).
- [11] H. Ougraz, S. Safi, and M. Frikel, "Analysis of Several Algorithms for DOA Estimation in Two Different Communication Models by a Comparative Study", *Lecture Notes in Business Information Processing*, vol. 449, pp. 219–230, 2022 ([https://doi.org/10.1007/978-3-031-06458-6\\_18](https://doi.org/10.1007/978-3-031-06458-6_18)).
- [12] J. Zhang, J. Dai, and Z. Ye, "An extended TOPS algorithm based on incoherent signal subspace method", *Signal Processing*, vol. 90, no. 12, pp. 3317–3324, 2010 (<https://doi.org/10.1016/j.sigpro.2010.05.031>).
- [13] H. Hayashi and T. Ohtsuki, "DOA estimation for wideband signals based on weighted squared TOPS", *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1–12, 2016 (<https://doi.org/10.1186/s13638-016-0743-9>).
- [14] H. Hung and M. Kaveh, "Focussing matrices for coherent signal subspace processing", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 8, pp. 1272–1281, 1998 (<https://doi.org/10.1109/29.1655>).
- [15] Z. Ahmad, Y. Song, and Q. Du, "Wideband DOA estimation based on incoherent signal subspace method", *COMPEL International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 37, no. 8, pp. 1271–1289, 2018 (<https://doi.org/10.1108/COMPEL-10-2017-0443>).
- [16] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985 (<https://doi.org/10.1109/TASSP.1985.1164667>).
- [17] K.S. Huang, Y.Y. Zhou, and G.Z. Zhang, "A fast DOA estimating method of wideband signals based on Krylov subspace", *Journal of Astronautics*, vol. 26, no. 4, pp. 461–465, 2005.
- [18] M. Wax, T.J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 4, pp. 817–827, 1984 (<https://doi.org/10.1109/TASSP.1984.1164400>).
- [19] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas Propagation*, vol. 34, no. 3, pp. 276–280, 1986 (<https://doi.org/10.1109/TAP.1986.1143830>).
- [20] H. Yu, J. Liu, Z. Huang, Y. Zhou, and X. Xu, "TOFS: A new method for wideband DOA estimation", in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, vol. 28, Shanghai, China, pp. 598–601, 2007 (<https://doi.org/10.1109/WICOM.2007.155>).
- [21] Y.S. Yoon, L.M. Kaplan, and J.H. McClellan, "TOPS: new DOA estimator for wideband signals", *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1977–1989, 2006 (<https://doi.org/10.1109/TSP.2006.872581>).
- [22] K. Okane and T. Ohtsuki, "Resolution Improvement of Wideband Direction-Of-Arrival Estimation Squared-TOPS", in *2010 IEEE International Conference on Communications*, Cape Town, South Africa, pp. 1–5, 2010 (<https://doi.org/10.1109/ICC.2010.5502480>).
- [23] M.A. Doron and A.J. Weiss, "On focusing matrices for wide-band array processing", *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1295–1302, 1992 (<https://doi.org/10.1109/78.139236>).
- [24] F. Sellone, "Robust auto-focusing wideband DOA estimation", *Signal Processing*, vol. 86, no. 1, pp. 17–37, 2006 (<https://doi.org/10.1016/j.sigpro.2005.04.009>).
- [25] D.N. Swingler and J. Krolik, "Source location bias in the coherently focused high-resolution broadband beamformer", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 143–145, 1989 (<https://doi.org/10.1109/29.17516>).
- [26] P.J. Chung, J.F. Bohme, C.F. Mecklenbrauker, and A.O. Hero, "Detection of the number of signals using the Benjamini-Hochberg procedure", *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2497–2508, 2007 (<https://doi.org/10.1109/TSP.2007.893749>).
- [27] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985 (<https://doi.org/10.1109/TASSP.1985.1164667>).
- [28] S. Valaee and P. Kabal, "Wideband array processing using a two-sided correlation transformation", *IEEE Transactions on Signal processing*, vol. 43, no. 1, pp. 160–172, 1995 (<https://doi.org/10.1109/78.365295>).
- [29] Y. Zeng and G. Lu, "Efficient wideband signals direction of arrival estimation method with unknown number of signals", *International Journal of Distributed Sensor Networks*, vol. 12, no. 11, 2016 (<https://doi.org/10.1177/1550147716676557>).

- [30] A. Asadzadeh, S.M. Alavi, M. Karimi, and H. Amiri, "Coherent wide-band signals DOA estimation by the new CTOPS algorithm", *Multidimensional Systems and Signal Processing*, vol. 31, pp. 1075–1089, 2020 (<https://doi.org/10.1007/s11045-020-00699-z>).

---

**Hassan Ougraz, M.Sc.**

Department of Mathematics and Computer Sciences, LIMATI Laboratory

 <https://orcid.org/0000-0002-0450-4893>

E-mail: hassan.ougraz.24@gmail.com

University of Sultan Moulay Slimane Beni Mellal, Morocco

<https://www.usms.ac.ma>

**Said Safi, Ph.D., Professor**

Department of Mathematics and Computer Sciences

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

University of Sultan Moulay Slimane Beni Mellal, Morocco

<https://www.usms.ac.ma>

**Ahmed Boumezzough, Ph.D., Professor**

Applied Physics and New Technologies Team

E-mail: ahmed.boumezzough@gmail.com

University of Sultan Moulay Slimane Beni Mellal, Morocco

<https://www.usms.ac.ma>

**Miloud Frikel, Ph.D., Professor**

Department of Physics

 <https://orcid.org/0000-0003-2178-1814>

National Graduate School of Engineering and Research Center Caen, France

<https://www.ensicaen.fr>

# Implementation of a 4G/5G Base Station Using the srsRAN Software and the USRP Software Radio Module

Wojciech Flakowski, Maciej Krasicki, and Rafał Krenz

*Poznan University of Technology, Poznan, Poland*

<https://doi.org/10.26636/jtit.2023.3.1298>

**Abstract** — This article presents the potential applications and scenarios for the implementation of a software-defined radio (SDR) module operating as a base station in 4G/5G networks. The paper presents test configurations of the srsRAN software environment in conjunction with various types of programmable radio modules. Then, the key performance metrics of the mobile telephony system and potential problems that may be encountered while implementing hardware and software layers are presented.

**Keywords** — base station, radio access network, radio channel power, radio link throughput, software-defined radio, user equipment

## 1. Introduction

This article presents the implementation of a cellular base station using SDR programmable radio. Such a concept allows for a significant development and improvement of radio applications. This requires smooth and quick changes in the operating parameters of the radio module to adapt to the current application scenario. This is possible thanks to the use of an FPGA-based system which enables the operating parameters to be reconfigured quickly.

An important role in ensuring the versatility of the SDR module is played by the transfer of some of the data processing and signal processing to the software domain. Such an approach results in a significant scalability of signal processing algorithms. SDR radio modules may be successfully used to implement the radio access layer for 2G, 3G, 4G and 5G digital cellular networks. In such a case, the signal is processed and the parameters of the carrier are controlled in the digital domain, with the aid of a computer. Having been processed in the software layer, the signal samples are sent to the SDR module. Using built-in FPGA and analog components of the RF path (i.e., amplifiers, filters and oscillators), the carrier is generated for the downlink and tracked in the uplink. The software-defined structure of the programmable radio and the extensive configuration possibilities enable the base station to quickly adapt to the specific type and standard of the wireless transmission supported by the user's mobile terminal. The increasing computational power of standard PCs and the wide availability of programmable radio modules have resulted in the design and development of software working as a complete base station. The srsRAN package is one of

the most popular open-source projects of this type. It offers the ability to emulate a base station using a PC and a programmable radio. The modular structure of the software package enables separate emulation of the core part of the network and the radio access layer (Radio Access Network, RAN), as well as the client part (User Equipment, UE). Thanks to such an approach, it is possible to observe the parameters of the data transferred between the individual modules.

In recent years, implementation work has been in progress to verify the operation of emulated base stations and UE for cellular systems. However, the propagation conditions of the complete radio path using commercial terminals have not been evaluated.

In [1], the authors tested an srsRAN system using USRP and LimeSDR modules. The measurements of the radio modules' transmit power were carried out using wires connecting the transmitter and the receiver. This allowed the absolute power values of the generated RF signal to be measured, but it did not take into account free-space losses and radio wave propagation mechanisms. On the other hand, there were also some implementations of software radio systems working in earlier RAN generations. As an example, a GSM system developed with the use of the OpenBTS software package and the USRP N210 radio module was described in [2]. The said paper provided an overview of the modular software configuration of the system, but failed to describe either 4G or 5G RAN.

The system described in this paper is used to test the open-source software of the srsRAN radio package and to measure its performance. Link capacity tests and power measurements are carried out in real world conditions. Such an approach allows to simulate and validate the system model in various radio environments. Therefore, it is possible to improve the transmission and receiving parameters. According to the authors' best knowledge, this is the first hardware implementation using a commercial UE and USRP radio module.

The paper is structured as follows. Section 2 presents the software components of the srsRAN package, as well as describes their functionality and operating characteristics. Next, in Section 3, the capabilities of the SDR programmable radio module and its cooperation with the software layer are described. Section 4 is a detailed description of the system's implementation, its features and functionalities. It shows the detailed parameters of the computers responsible for emulating base

station functions. In Section 5, the process of configuring, starting up and connecting the system is described. That part contains the settings of individual software modules, the SDR module and network functions. Such presets are necessary to establish a connection between mobile terminals. Then, in Section 6, radio link throughput measurements are presented, taking into account various configurations of the radio module, terminal location and transmission status. Similarly, high-frequency power measurements are carried out in an operational system, in the downlink and uplink channels, in Section 7. Additional experiments are conducted in Section 8, with some remarks regarding the stability of the system's operation under specific radio conditions and with different base station settings added.

## 2. Characteristics of the Software Layer

### 2.1. Overview of Software Layer Modules

In the application scenarios under consideration, the base station system is implemented using open source srsRAN software [3]. The srsRAN software package is characterized by a modular structure and is a complete solution that is required to run the base station and emulate user equipment (UE) on a PC computer. It also includes a communication layer with an SDR controller, making it a universal tool to run 4G and 5G networks.

The basic modules of the srsRAN package include: srsEPC, srsENB, and srsUE modules.

### 2.2. srsEPC Module

The srsEPC module emulates the core network for the 4G solution. It consists of four basic components performing strictly defined tasks. The first one is the mobility management entity (MME) subsystem. It is the main control element of the 4G core network, managing mobility and the connection of a mobile terminal to the network. It is also responsible for exchanging mobile terminal authentication data with the home subscriber server (HSS) database. The srsEPC module also connects (through the S1-M interface) with the srsENB module that is responsible for the radio access layer. The MME sends information via the radio resource control (RRC) protocol to activate the terminal's idle or active mode.

The HSS database is the second component closely cooperating with the MME subsystem. In the srsRAN environment, it is implemented using the *user\_db.csv* file which contains the authentication data of SIM cards installed in mobile terminals. It should be noted that the identification of terminals is done by authorizing the international mobile subscriber identity (IMSI) number which enables network services to be launched on any terminal with a registered SIM card. The MME main control unit is connected to the HSS register subsystem via a virtualized S6-M interface.

The third important component of the srsEPC module is the service gateway (S-GW) subsystem which is responsible for routing packets between the UE and the external P-GW network gateway. It sets up GPRS tunneling protocol (GTP)

sessions between the srsENB module and the P-GW gateway subsystem. It is also responsible for assigning IP addresses to mobile terminals communicating with the base station and is connected directly to the srsENB module through the virtualized S1-M interface.

The fourth subsystem included in the srsEPC module is the packet data network gateway (P-GW) which redirects network traffic between the S-GW subsystem and the external Internet network. The Internet connection is physically provided by connecting a PC network card to a LAN network via the Ethernet port. The P-GW subsystem does not have a direct connection with the srsENB module, but it connects to the S-GW unit through virtualized S5 and S8 interfaces. The S-GW unit is connected to the MME unit through the S11 interface.

### 2.3. srsENB Module

The srsENB module performs the function of the base station: eNodeB in the 4G network and gNodeB in the 5G network, respectively. In the test configuration, the srsENB module shares a computing unit with the srsEPC module. It is also possible to run the srsENB and srsEPC modules on separate PCs. When both modules work on a single PC unit, communication takes place via the IP address of the local loopback system interface. The srsENB program supports the LTE/4G mode in the 1.4, 3, 5, 10, 15 and 20 MHz bands, providing a wide range of configuration options for the available channels.

It is possible to connect the srsENB module to an external 5G core network module in order to implement a 5G SA (standalone) network in a scenario in which the srsEPC module is not used. The third available operating mode of the srsENB enables a 5G NSA non-standalone connection. This mode allows to set up two radio channels using 4G and 5G technology. In that case, the UE establishes a connection with the 4G channel to exchange authorization and control data. Simultaneously, it establishes an additional connection using the 5G carrier in order to implement a broadband channel dedicated to the user's data traffic. The srsEPC module is still responsible for managing data transmissions with the external network. In the case of a 5G NSA connection, the srsENB module creates two virtual base stations (eNodeB and gNodeB), but the data within the core network is exchanged between the subsystems of the eNodeB station only (via S1-M interfaces). In addition, communication between the eNodeB and gNodeB subsystems is established via the X2 interface inside the srsENB module, so the gNodeB subsystem does not have a direct connection to the srsEPC module.

In the case of the 5G SA operating mode, where the srsENB module activates the gNodeB subsystem only, the connection established via N2 and N3 interfaces requires an independent software module that performs the functions of the 5G core network. The N2 interface is then used to transmit information from the srsENB to the access and mobility management function (AMF) subsystem of the 5G core network. The AMF is responsible for management, access control and autho-

rization of a terminal connecting to the 5G network. The N3 interface is used to connect srsENB to the user plane function (UPF) subsystem which is responsible for routing data packets between the UE and the external Internet network. In addition to the basic single-input single-output (SISO) transmission mode (mode 1), three  $2 \times 2$  multiple-input multiple-output (MIMO) modes are supported by the srsENB and selected SDR modules. Mode 2 is the broadcast diversity mode, where replicas of the signal are transmitted through all antennas using individual frequency resources and coding schemes. This mode of operation boosts the signal-to-noise ratio (SNR) and also makes the transmission more resilient to channel interference. Mode 3 is based on open-loop spatial multiplexing and cyclic prefix delay diversity. This means that each of the ports transmits the subcarriers using a different cyclic prefix delay. This solution increases resistance to interference in channels exhibiting a high degree of variability. The last of the implemented transmission modes is the spatial multiplexing mode using a closed feedback loop (mode 4). In this mode, each antenna transmits a separate data stream used to increase the radio connection's total bit rate. In order to perform the correct channel estimation on the mobile terminal side, the base station transmits reference signals (RS) in specific resource blocks and time slots. Those reference signals are used to reduce the error rate in the transmission between eNodeB and the UE. It is worth noting that the terminal sends channel quality reports, including the selection of the appropriate precoding matrix. The precoding matrices are selected using their associated codebook indexes – precoding matrix indicators (PMIs) for each antenna. The srsENB software works in the frequency domain division (FDD) mode, where the downlink channel and the uplink channel transmit on separate frequencies with a defined channel spacing (this is made possible by specifying the EARFCN parameter).

#### 2.4. srsUE Module

The srsUE module is an independent software package that may be run on a separate PC computer together with the software radio module (SDR). SDR is responsible for communicating, via the physical layer, with the base station. The srsUE software also supports authentication by emulating a virtual SIM card whose parameters are defined in a file known as *ue.conf* (soft USIM). However, it is possible to use a hardware SIM card reader connected to the PC. In the test scenario, a virtual USIM card is used.

The srsUE module performs all UE functions. It supports both 4G and 5G in NSA and SA modes. This module cooperates directly with the SDR hardware module, as such an approach offers great configurability along with easy software setup of the available bands and frequency resources. It is possible to rely on frequency division duplex (FDD) or time division duplex (TDD) techniques. In addition, all MIMO transmission modes listed above and all available radio channel widths of the LTE standard (1.4, 3, 5, 10, 15 and 20 MHz) are supported.

#### 2.5. Radio Controller Command Interpreter

It is important that both srsUE and srsENB modules have a built-in interpreter of radio controller commands, as this allows to define the operating parameters of the radio module based on respective configuration files: *ue.conf* (for the srsUE module) and *enb.conf* (for the srsENB module).

### 3. Hardware Layer Characteristics

#### 3.1. Internal Structure of the Radio Module

The scalability of the solution based on virtualized cellular networks is additionally enhanced by the support of various types of radio modules available on the market. srsRAN software may communicate with them using hardware drivers which manage the operation of specific radio modules. The programmable radio controller works as an interface between the antenna and the L1 (physical or PHY) layer. It contains RF front-end as well as analog-to-digital (ADC) and digital-to-analog (DAC) converters. The srsENB software emulates the entire protocol stack of L1, L2 (MAC, RLC, PDCP) and L3 (S1-AP, RRC, GTP-U) layers. It initiates communication with the programmable radio controller and processes signal samples in the digital domain. The tested hardware platform uses various types of programmable radios from the USRP Ettus Research family, but it is possible to use stations with other types of radio modules, such as BladeRF and LimeSDR. In the case of the USRP radio modules, the UHD hardware driver is used. USRP 1, USRP B210, USRP N200/N210 and USRP E320 modules are employed for the tests, assuming a wide range of carrier frequencies (from 70 MHz to 6 GHz in the case of USRP B210 and E320).

The SDR modules can be divided into two main categories, namely those responsible for the individual stages of processing received and transmitted signals. The first component is the motherboard. It is tasked with processing the signal in the baseband and the intermediate frequency (IF) band. Meanwhile, in the daughterboard, the target radio frequency (RF) carrier is generated by local oscillators and the mixers multiply the IF signal with the carrier. Such a structure is present in the USRP N200/N210 and USRP 1 radio modules, where it is possible to physically replace the daughterboard with a model operating in a different band. All components of USRP B210 and USRP E320 modules are integrated on a single PCB and cannot be reconfigured.

#### 3.2. Radio Front-end Tuning Process

The programmable radio tuning process is divided into two stages. Coarse tuning is performed by specifying the frequency of the local oscillator. Such an approach enables the carrier frequency to be tuned in the full range offered by a given module. Fine tuning is realized in the digital domain and consists in FPGA programming. By default, the signal in the digital domain is sampled with a frequency of the main system clock. This limits the maximum tuning range in the digital domain to the master clock rate. The maximum signal band-

width is also restricted by the main system clock. Fine tuning of the carrier frequency can be done by shifting the frequency in the digital domain within the tuning range (*lo\_offset*), so that the center frequency (which is the frequency of the local oscillator) determines a central point in the tuning range. It must be pointed out that the maximum frequency of the TX radio channel should not exceed the limit defined by the maximum tuning range, since it may cause signal leakage and aliasing. In the case of radio modules equipped with an anti-aliasing filter, if the permissible tuning range is exceeded, the center frequency will be shifted and wrapped to the lowest allowable frequency. For example, if the maximum allowable tuning frequency is contained in the tuning range of 180 – 200 MHz and a 15 MHz digital frequency shift is added to the oscillator (190 MHz), the tuning range is increased to the value of 205 MHz. When the anti-aliasing filter is activated, the final carrier frequency will be moved to the value of 185 MHz.

### 3.3. Principles of Channel Mapping in Radio Module

In the case of radio modules using a distributed hardware structure (separate motherboard and daughterboard), the channel mapping rule is defined by the input parameters of the UHD driver. Channel mapping parameters are specified in the *enb.conf* file in the *device\_argsfield* using the *rx\_subdev\_spec* formula for RX channels and the *tx\_subdev\_spec* formula for TX channels. A general channel mapping scheme is A:0 B:0, where variables A and B specify the daughterboards (corresponding to the slots on the motherboard), while 0 identifies the antenna port of the daughterboard. Note that the daughterboard may have two antenna ports (indexed as 0 or 1). The direction of subsequent ports is specified by individual variables: *tx\_subdeb\_spec* (group of TX ports) and *rx\_subdev\_spec* (group of RX ports). In the case of a radio module that can accommodate two daughterboards, the channel mapping configuration is the following: *tx\_subdev\_spec*=A:0 B:0, *rx\_subdev\_spec*=A:0 B:0. It enables TX port 0 and RX port 0 on both (A and B) daughterboards. Such a configuration is possible only for the USRP 1 module which enables installation of two daughterboards. The N210 module has only one slot for an external daughterboard, so it is possible to configure the channels of one daughterboard only – A:x A:y where x, y are the daughterboard's port numbers. In the case of modules where the RF front-end and the ADC and DAC converters are integrated (B210 and E320 models), there is a different, specific type of channel mapping. In the case of the B210 model, which has two TX ports and two RX ports, the channel mapping scheme is the following – A:A A:B. In this case, they are virtually connected to one motherboard. In the E320 model, the channel mapping scheme is A:0 A:1, i.e., there are also two channels (2 TX ports and 2 RX ports).

### 3.4. Configuration of Internal Oscillators

The internal configuration of local oscillators is an additional aspect that affects the individual programmable radio models.

In the case of the USRP B210 and E320 radio modules, two local oscillators are available sharing their carrier frequencies between the TX and RX ports. This means that basic tuning of individual TX and RX channels cannot be performed separately. Instead, one can set a common carrier frequency for all TX channels and a common frequency for all RX channels. As a consequence, the USRP B210 and E320 modules support only the  $2 \times 2$  MIMO transmission mode. This prohibits the use of the 5G NSA mode using separate frequency bands in the FDD mode, e.g., 1800 MHz band for the 5G broadband channel and 2600 MHz band for the 4G control data channel. The manner in which channels are mapped by the srsENB program is an additional implementation problem. In the case of the USRP B210 radio, it is not possible to use the RF B port to generate a separate, independent carrier due to the unusual specificity of the channel mapping scheme – A:A A:B. As a result, the B210 radio cannot be used to set up a 5G network in the NSA mode. However, it is still possible to use the  $2 \times 2$  MIMO mode for a single 4G network cell.

### 3.5. Multiple Radio Modules Configuration Problem

It is not possible to use two B210 or E320 modules to generate two independent carriers, because these models are not supported by the *multi\_usrp* function which enables physical binding of more USRP modules into one virtual device. However, this is possible with the use of USRP N210 modules. The UHD driver is responsible for the process of combining individual modules into one logical device.

It should be mentioned that the USRP N200/N210 module is not supported by the srsENB program due to the conflict in providing the information necessary to initialize the UHD driver. For proper operation, the USRP N210 module requires specifying the reference time source, defined by the *time\_source* argument. In the latest revision of the srsRAN suite, this function is not supported by the srsENB software. The *time\_source* parameter cannot be defined manually in the *enb.conf* configuration file and the UHD driver does not initialize correctly in that case.

### 3.6. System Bandwidth Requirements

The width of the radio channel is an important implementation aspect as well. When the srsENB software module operates in the 5G NSA network mode, it is possible to set up two 10 MHz channels: a 10 MHz channel for the 4G standard carrier and a 10 MHz channel for the 5G NSA standard carrier. Such a configuration requires a total processing bandwidth of 20 MHz. Although the USRP 1 radio module meets the requirements of channel mapping through the srsENB application, it does not have the ability to process the total signal bandwidth. In fact, the available total bandwidth with 8-bit sampling is 16 MHz – the value is too low for 5G NSA mode transmissions.

### 3.7. Digital Frontend Tuning Procedure

Despite the inability to separate the carrier frequencies for individual frequency bands in the case of the USRP E320

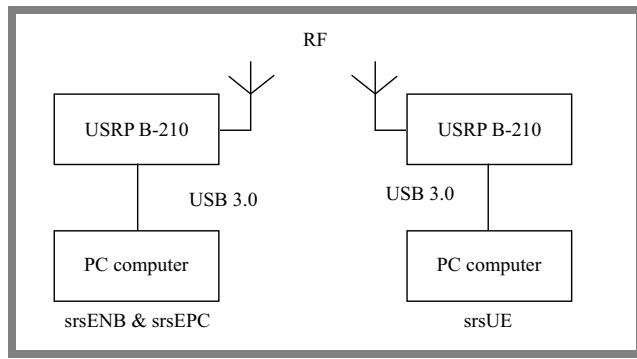


**Fig. 1.** eNodeB and srsUE system configuration.

module, it is possible to run the 5G NSA mode thanks to the digital domain tuning. In this scenario, identical values of EARFCN and ARFCN parameters need to be entered for both carriers, e.g., 2600 MHz range. EARFCN and ARFCN values allow to set downlink and uplink frequencies for 4G and 5G carriers by specifying a set of frequencies defined by EARFCN (for 4G) and ARFCN (for 5G). These settings can be set in the *rr.conf* file describing the radio resources of the eNodeB station. Thereafter, using the *lo\_offset* command, the frequencies need to be separated in the digital tuning range, with the tuning range defined by the main system clock – *master\_clock\_rate*. It needs to be borne in mind that clock frequencies have to be used that are integer multiples of the signal sampling frequency.

### 3.8. Interface Bandwidth Requirements

Another problem encountered when implementing a 5G NSA network using the USRP E320 platform is that the bandwidth of the 1 Gbit/s Ethernet communication port is too low. This port is capable of transferring up to 25 Mbaud/s, which translates into the maximum available bandwidth of 20 MHz. In the case of the 5G NSA mode, four 10 MHz channels are indispensable, meaning that a total system bandwidth of 40 MHz is required. To fulfill such a requirement, it would be necessary to equip a PC-class unit, responsible for emulating the functions of the 5G NSA base station, with a 10 Gbit/s Ethernet card. Replacing the card can provide a throughput of 200 Mbaud/s. Using the available system clock frequency of 46.08 MHz, it is possible to set up 4 channels with a total bandwidth of 40 MHz – 2 RX channels and 2 TX channels.

## 4. 4G Base Station System Implementation

### 4.1. Description of Hardware Components

This section describes the process of configuring and optimizing the 4G cellular network developed with the use of the USRP B210 programmable radio hardware platform. As a test platform, a PC equipped with an AMD Ryzen 5 3600 processor, an Nvidia GeForce GTX 1650 graphic card and 16 GB of RAM is used. The srsRAN software containing the srsENB and srsEPC modules is used to emulate the com-

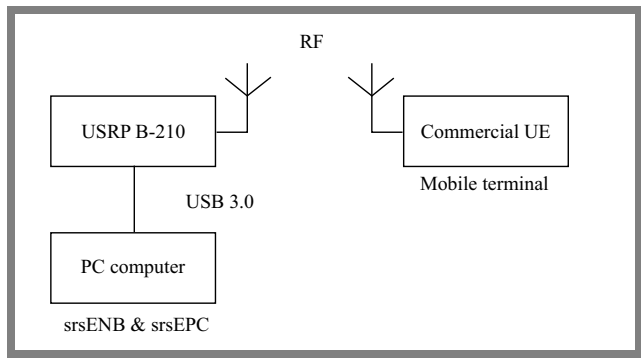


**Fig. 2.** eNodeB and commercial UE terminal configuration.

plete functionality of the base station. The srsUE software is launched on a separate PC with the same hardware configuration and the USRP B210 radio module. In addition, system performance tests with the commercial UE are carried out. The UE used in the described system supports both 4G and 5G solutions. Communication of the radio module with the PC computer is set up via the USB 3.0 port.

### 4.2. System Performance Measurement Procedure

The transmission rate and power received in the downlink and uplink channels are measured at a distance of 3 m from the base station for the UE and USRP B210, and 10 m for the UE only. System performance tests are carried out with the channel bandwidth of 10 MHz and 20 MHz, in the following transmission modes: 1 – SISO mode, 2 – broadcast diversity mode ( $2 \times 2$  MIMO), 3 – spatial multiplexing mode with cyclic delay diversity (CDD), and 4 – closed-loop spatial multiplexing mode. In addition – for transmission mode 4 – performance measurements are made with active automatic gain control (AGC) in the RX path. In that case, 80 dB, 100 dB or 120 dB transmit gain levels (*TX\_gain*) are considered. The transmission takes place in the 2600 MHz frequency band, combining the downlink channel at the frequency of 2680 MHz and the uplink channel at the frequency of 2560 MHz. Performance of the system is verified using the iPerf tool [4] which generates network traffic of a specified intensity, using both TCP and UDP protocols. The tool is launched on both srsUE mobile terminals (USRP B210), commercial UE, and the eNodeB base station (USRP B210).

Figure 1 shows the connection between the USRP B210 modules and PC-class computers using the srsUE client, while Fig. 2 illustrates the connection between the eNodeB station and the commercial UE.

## 5. Installation and Configuration of the 4G Station

### 5.1. Additional Software Packages

In order to properly install the srsRAN package, it is necessary to install the *libuhd\_dev* radio driver package and the *uhd\_host* package first. This allows to control the state of the radio module in Linux.

### 5.2. srsGUI Suite Description and Functionality

The srsRAN environment is integrated with the srsGUI graphical package. This enables the observation of the constellation of the transmitted signal in the shared uplink physical channel (PUSCH) and the PUCCH control channel. The graphical package is launched by running the srsENB module. Similarly, a graphical environment may be installed for the srsUE package. This allows one to view the transmitted signal constellation for the downlink shared channel (PDSCH) and the downlink control channel (PDCCH). The srsGUI desktop environment must be installed prior to the main installation of the srsRAN package to ensure proper operation of the latter.

### 5.3. Software Module Configuration Options

After installing the srsRAN environment and the auxiliary components, it is recommended to move the configuration files to the *etc/srsran* system folder. The complete system configuration is stored in files with the *.conf* extension. This enables the parameters to be edited in a text editor, so that the configuration of the srsEPC, srsENB and srsUE modules can be easily modified. *enb.conf*, *sib.conf*, *rr.conf* and *rb.conf* files configure the following operating parameters of the srsENB module: transmission mode, transmitted power (*enb.conf*), resource allocation (*rr.conf*), and band aggregation (*rb.conf*). The operating parameters of the srsEPC module are controlled by variables stored in *epc.conf* and *user\_db.csv* files. In the *epc.conf* file, IP addresses of the srsENB module (*mme\_bind\_addr* field) and an external DNS server are to be defined. This allows Internet access via the UE interface (to be specified in the *dns\_addr* field). Authentication data of UEs, required to establish a connection with the station, are stored in the *user\_db.csv* file; the type of the identification algorithm (XOR or Milenage), and the IMSI number of the SIM card of the UE can be set. The path to the *user\_db.csv* file is defined in the *epc.conf* file in the *db\_file* field, which enables access to the authentication data. This file is read during the procedure of connecting the UE to the network.

The srsUE module also has the *ue.conf* configuration file which stores various terminal configuration details, including the transmission mode and uplink transmit power. With that borne in mind, neither the eNodeB station nor the srsUE terminal feature a transmission power control mechanism, so the transmit power must be set before turning on the srsENB and srsUE modules.

The srsEPC, srsENB and srsUE modules on a separate machine are launched in individual terminal windows.

In the case of the srsENB module, it is also possible to specify the connection status display in real time. This function displays aggregated control information concerning downlink and uplink bit rate, the channel quality indicator reported by the mobile terminal (the *cqi* parameter) or the modulation and coding scheme (*mcs* parameter).

Additionally, the UHD controller is initialized automatically when starting the srsENB or srsUE modules. Therefore,

communication with the programmable radio module is set up automatically.

## 6. 4G Network Performance Measurements

### 6.1. Testbench Configuration

In the first phase of the performance measurement process, a connection between the srsENB and srsEPC programs and srsUE modules was established. For this purpose, two B210 radio modules were configured, one as the eNodeB and the other as the UE. Each of these modules was connected to a PC running Ubuntu 22.04 LTS with the srsRAN package installed. The iPerf program was used to carry out the tests.

The srsRAN software enables the collection of diagnostic data from all srsRAN software modules which were saved in the *tmp* directory in the *.log* and *.pcap* file formats. Thanks to that, one can analyze the transmitted data frames e.g. in the Wireshark app. Complete files were created when all modules of the srsRAN package ended their operation.

The configuration details are as follows:

**Resource blocks and bandwidth configuration.** The tests were carried out using channel bandwidths of 10 MHz and 20 MHz, which corresponded to using 50 and 100 resource blocks (*n\_prb* is set to either 50 or 100 in the *enb.conf* file).

**MIMO mode configuration.** Depending on the  $2 \times 2$  MIMO antenna configuration selected, it was necessary to set the following parameters accordingly:  $t_m = 2$  (transmit diversity),  $t_m = 3$  (transmission using the CCD cyclic prefix shift),  $t_m = 4$  (multiplexing with closed feedback loop). For multi-antenna transmission modes, the number of active

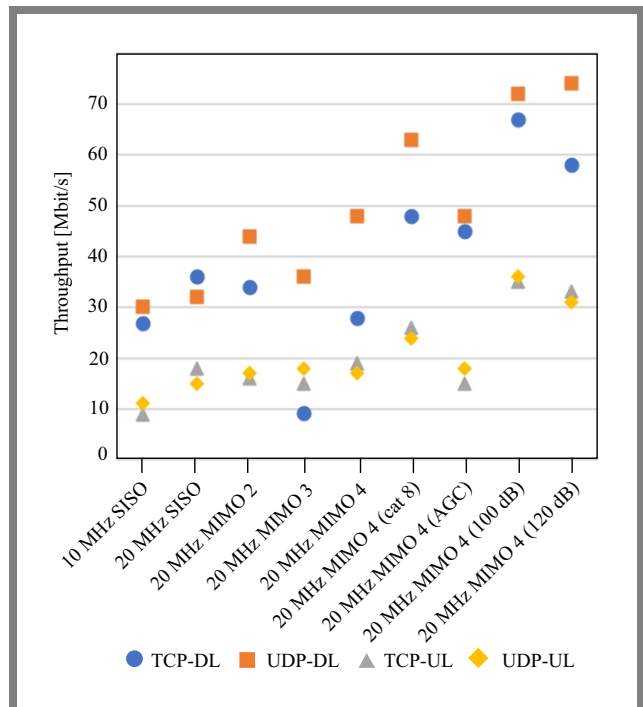


Fig. 3. Iperf performance test – srsUE.

TX ports should be set to 2 (*nof\_ports=2* command in the *enb.conf* file).

**Transmission frequency configuration.** The transmission frequency was determined by the *dl\_eurfcn=3350* parameter which corresponds to the setting of the downlink channel frequency (2680 MHz). The uplink channel frequency was then set automatically at 2560 MHz.

**Transmission gain setup.** The transmit gain (*tx\_gain* parameter) was set at 80 dB, then 100 dB and 120 dB.

**AGC mechanism configuration.** The tests were initially carried out with a constant RX channel gain of 40 dB (*rx\_gain=40* command). In the subsequent measurements, the AGC mechanism was activated by commenting out respective configuration line *#rx\_gain=40*.

**Miscellaneous radio module settings.** In order to improve the efficiency of the USRP B210 radio module, additional arguments, such as *device\_args=(. . .) num\_send\_frames=64* and *num\_recv\_frames=64*, were set. These allowed to increase the number of RX and TX buffers for the transmitted frames from 32 to 64, thereby improving overall performance of the system. In the case of UE emulated by the srsUE software, the same modifications regarding the channel gains were made. Note that the AGC mode was available for RX channels only. In the case of the MIMO mode, the *nof\_antennas* parameter had to be set to 2 in the *ue.conf* file. The arguments passed to the UHD driver in the *device\_args* field were defined in the same way as in the *enb.conf* file.

Performance tests were conducted for category 4 and category 8 terminals. In the latter case, 64-QAM modulation was used in the uplink. The current mobile terminal category was defined in the *ue\_category* field.

**Configuration of srsUE virtual network interface.** To provide the Internet connection to the mobile terminal emulation station using the srsUE software, it was necessary to add a default IP traffic routing gateway on the UE client computer. Then, the default gateway address was linked with the virtual network interface emulated by the srsUE program called *tun\_srsue*. In this case, the IP address was the address of the S/P GW subsystem which acted as a virtual network gateway. It also enabled the routing of packets to UEs connected to the eNodeB station. It is important to note that the *tun\_srsue* virtual interface was active only when the srsUE module was running.

**Downlink channel performance test.** In order to check the downlink throughput at the client station (UE), the iPerf service server was started on a separate terminal. TCP and UDP traffic was generated using the iPerf client.

In the case of TCP traffic, the maximum throughput was not defined by the user, while for UDP traffic, the value of the maximum traffic throughput was set arbitrarily to 100 Mbit/s. Note that the value defined above exceeded the maximum system throughput obtained.

**Uplink channel performance test.** The uplink channel throughput test was carried out in a similar way, but the iPerf service server listening to the network traffic was launched on

the eNodeB station. Simultaneously, a client sending queries was run on the station emulating a UE.

## 6.2. Radio Channel Performance Test Results – srsUE

Radio link throughput results are presented in Fig. 3. Most of the measurements were carried out with the transmit gain of 80 dB. Only for the “20 MHz – MIMO – 4 (100 dB)” and “20 MHz – MIMO – 4 (120 dB)” setups, the gains of 100 and 120 dB were applied, respectively. The AGC was also set active and category 8 UE was selected. The distance between the antennas of the USRP stations was constant throughout all of measurement tests – it was 3.2 m in line of sight (LOS) conditions. The higher data rates observed in the downlink resulted from the use of a higher-order modulation than that used in the uplink.

## 6.3. srsENB Configuration – Commercial UE Terminal

The second series of measurement tests was carried out using commercial off-the-shelf UE. This device was configured to work with the base station by programming accurate SIM card parameters in the *user\_db.csv* file. The following properties have been filled in: Name – the name of the UE defined by the user, Auth – authentication algorithm mode selected – XOR or Milenage, IMSI – identification number of the SIM card installed in the UE, Key – authorization key in hexadecimal format, OP\_Type – type of the used operator code, OP/OPc – value of operator code assigned to the specified SIM card, AMF – authentication management field, SQN – UE sequential number, used when restarting the identification procedure), QCI – a field specifying the identifier associated with the management class of the quality of service (QoS) mechanism, IP\_alloc – the method of assigning the IP address to the UE – static or dynamic.

Having made changes to the configuration in the *user\_db.csv* file, it was necessary to specify the configuration of the APN. For this purpose, a point called default has been defined in the UE settings field (APN tab). This setting was identical to the value of the *apn=default* field contained in the *epc.conf* file. Determining the correct APN point in the UE was necessary to get access to data transmission services via the 4G network, and then connect to the Internet.

The remaining parameters were configured similarly to the process of setting up a connection with a virtual UE, emulated by the srsUE program. The base station maintained the same configuration as in the previous tests i.e. using the srsUE client.

## 6.4. Channel Performance Test – Commercial UE Terminal

Connection tests were carried out using the iPerf tool (with the use of the UDP protocol) in the same way as relied upon for the previous measurement scenario. However, the UE was located at a distance of either 3.2 m or 7.8 m from the base station antennas. The aim of running an additional series of measurements over a distance of 7.8 m was to evaluate the effectiveness of the power control mechanism in the uplink. Unlike the USRP B210, the commercial UE was equipped

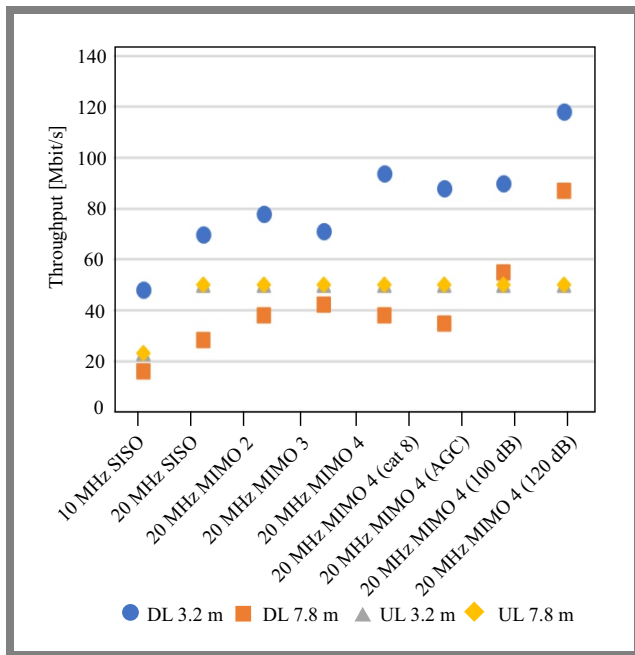


Fig. 4. Iperf performance test – commercial UE.

with a transmission power control mechanism. The implementation of the iPerf server and client on UE was carried out using the iPerf2 for Android application.

Radio link throughput results are shown in Fig. 4. The measurements were carried out with three levels of transmit gain at the 2680 MHz carrier in the downlink. In the case of measurements with TX gain of 100 dB and 120 dB, the AGC mechanism remained active. The other measurements were carried out with a transmit gain of 80 dB. In this case, the “20 MHz – MIMO – 4 (AGC)” measurement was made using the automatic gain control mode and MIMO multi-antenna transmission mode 4 (closed loop spatial diversity).

## 7. Radio Channel Power Measurements

In addition to radio link throughput measurements, the received and transmitted power in the uplink and downlink channels was measured as well. For the srsUE, a Rohde & Schwarz FSH4 spectrum analyzer with an Ettus Research VERT 2450 antenna was used. Similar antennas were used in the USRP B210 modules operating as an eNodeB station and an UE client. Downlink channel power measurements were made at three measurement points – 10 cm from the base station, 10 cm from the srsUE mobile terminal station, and 30 cm from the base station (1.2 m from the UE station). The measurements were repeated with active data transmission and in with the UE in idle mode (no active transmission). The uplink channel power was measured during active packet transmission at a distance of 10 cm from the base station (eNodeB). The received power values have been averaged. A separate series of power measurements was conducted for commercial UE. This kind of terminal features an internal microstrip antenna. Power measurements were carried out for two UE positions – it was located at a distance of 3.2 and

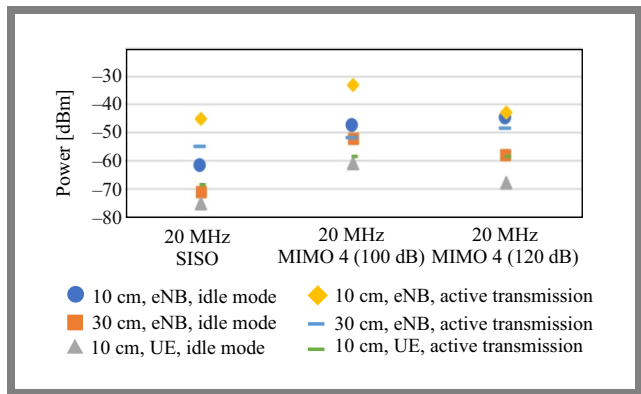


Fig. 5. Power measurements: downlink – srsUE.

7.8 m from the base station. The measurements were made at a distance of 10 cm from the base station, 30 cm from the base station and 10 cm from the UE.

### 7.1. Results for srsUE

The results of power received in the downlink and in the uplink for the srsUE mobile terminal are shown in Fig. 5 and Fig. 6, respectively.

Figure 7 presents the power spectral density (PSD) of the downlink signal in the idle mode at a distance of 10 cm from the base station, assuming a 20 MHz bandwidth and a 100 dB transmit gain.

Figure 8 shows the PSD of the downlink signal in the active transmission mode at a distance of 10 cm from the base station, with a 20 MHz bandwidth and a 100 dB transmit gain

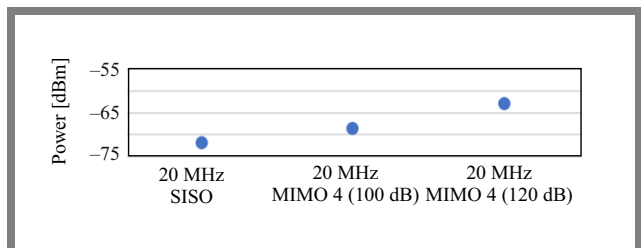


Fig. 6. Power measurements: uplink – srsUE.

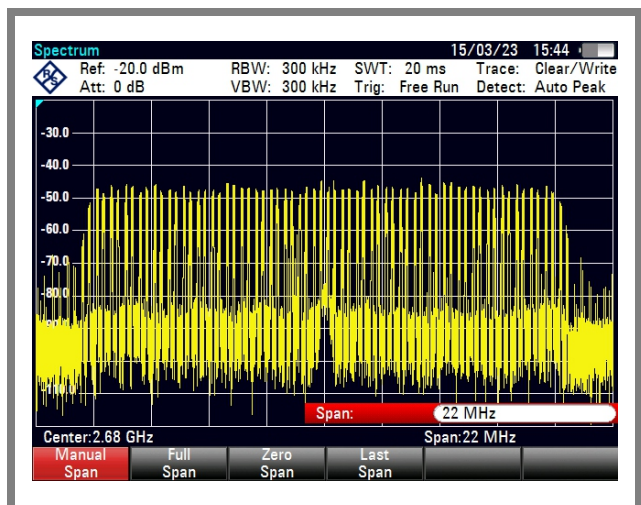


Fig. 7. Spectrum measurements: downlink – 10 cm – idle mode.

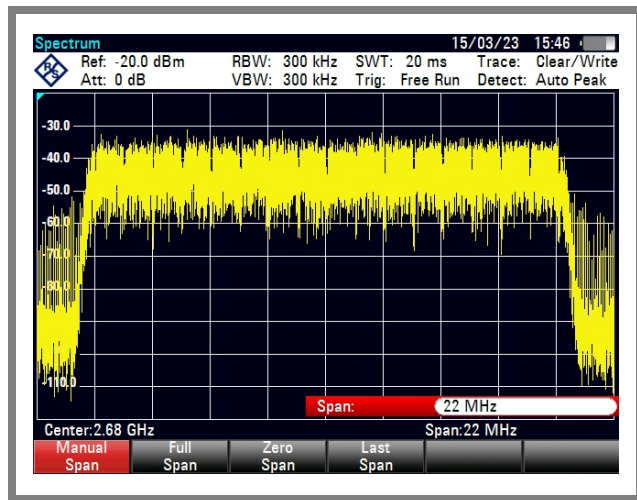


Fig. 8. Spectrum measurements: downlink – 10 cm – active transmission.

assumed again. Figure 9 illustrates the PSD of the uplink signal in the active transmission mode at a distance of 10 cm from the base station for a 20 MHz bandwidth and a 100 dB transmit gain.

## 7.2. Commercial UE Terminal

Figure 10 shows the power transmitted in the downlink for the terminal located at a distance of 3.2 m from the base station. Figure 11 presents the values of the power transmitted in the downlink for the UE located at a distance of 7.8 m from the base station, while Fig. 12 shows the values of the power transmitted in the uplink for the UE located at a distance of 3.2 m and 7.8 m from the base station, i.e. during active data transmission.

Figures 13 and 14 present the PSD of the uplink signal at the base station input situated 3.2 m and 7.8 m from the UE, respectively, for a 20 MHz bandwidth and a 100 dB downlink TX gain.

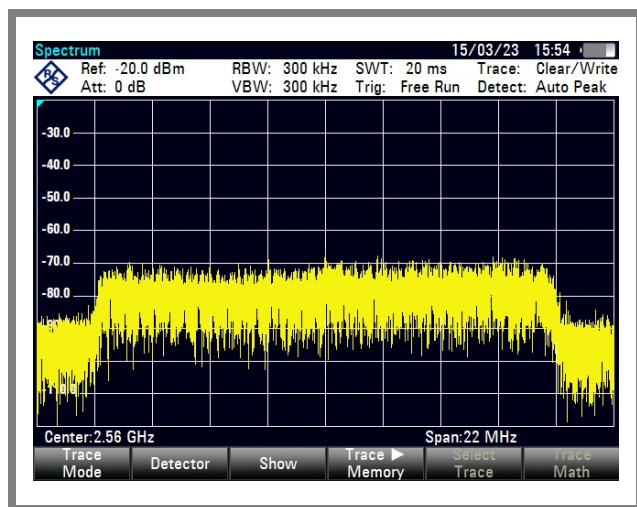


Fig. 9. Spectrum measurements: uplink – active transmission – 20 MHz bandwidth and 100 dB transmit gain.

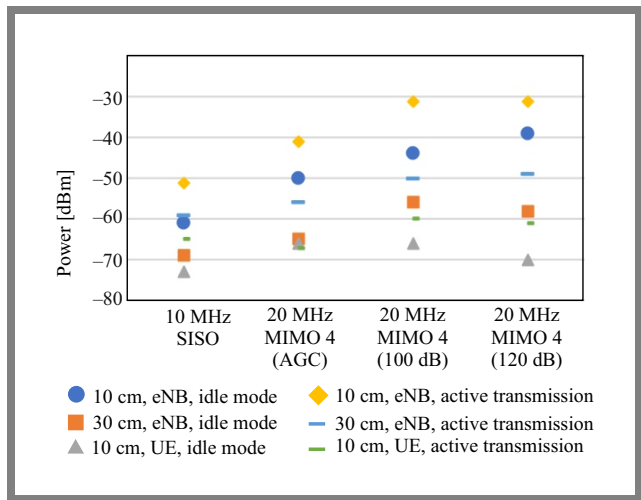


Fig. 10. Power measurements: downlink – 3.2 m – commercial UE.

## 8. Conclusions

In the case of radio link throughput tests using the USRP B210 module acting in the capacity of UE, a radio channel bandwidth increase from 10 to 20 MHz has resulted in boosting the link throughput by approx. 25%. Activation of the multi-antenna transmission mode failed to offer a significant gain in the link throughput, but increased its stability. Upgrading the UE category to level 8 resulted in the link throughput

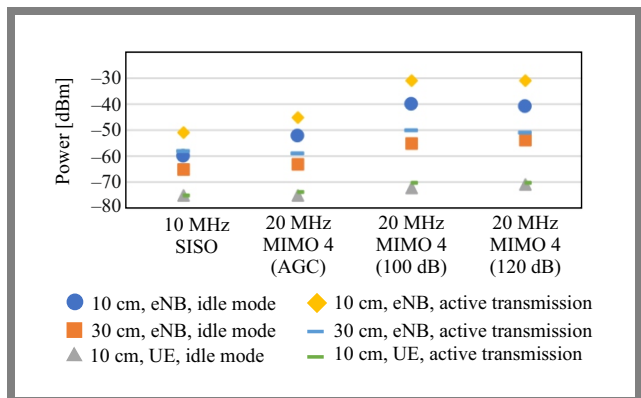


Fig. 11. Power measurements: downlink – 7.8 m – commercial UE.

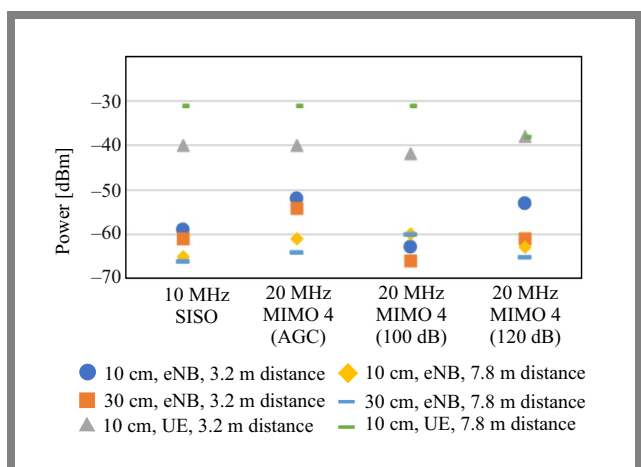


Fig. 12. Power measurements: uplink – commercial UE.

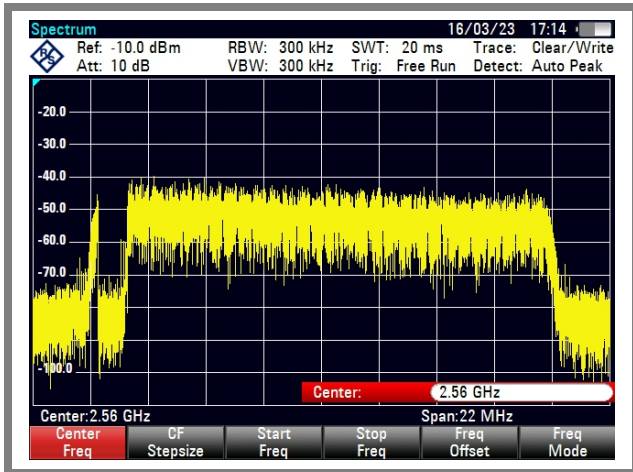


Fig. 13. Spectrum measurements: uplink – distance 3.2 m – 20 MHz bandwidth.

being increased by approx. 50 % (using the UDP test protocol). This is due to the change of the modulation scheme to 64-QAM on the UE side. Activation of the AGC mode caused a drop in link throughput. This resulted from the observed RX gain being lower than in the case of the static setting, which diminished SNR. On the other hand, a significant increase in the link throughput (up to 74 Mbit/s) was observed when the TX gain in both the base station and the UE increased to 120 dB. Any increment in the TX gain boosted the radiated power which, in turn, improved SNR. At a higher SNR range, it was possible to use a higher-order modulation and coding scheme. Exceeding the gain value of 100 dB failed to offer a significant increase in the link throughput. This is caused mainly by the saturation of the radio module power stages. No significant increase in the carrier power level was recorded.

In the case of radio channel throughput tests using commercial grade UE (a phone), the results were similar. The activation of the AGC mode, once gain degraded the link throughput. The decreased link throughput was noticeable when the terminal had moved away from the base station – a higher distance entailed a higher free-space attenuation and an SNR decrement. On the other hand, a drop in SNR caused the station to use lower-order modulation and coding schemes. This mechanism decreased the link throughput with the aim to limit the bit error rate (BER).

During the power measurements, the carrier power increased proportionally to the value of the defined transmit gain. The difference in the downlink carrier power between the idle mode of the UE and active transmission was 10 dB, on average. It is important that the downlink OFDM subcarriers were transmitted continuously and their power increased in the case of active data transmission. The uplink power was measured only during active transmission. With no active user data transmission from the UE, control data were transmitted using the selected pilot subcarriers of the uplink channel bandwidth only. This was unlike with active data transmission, where all subcarriers were used. In the idle mode i.e., without active data transmission on user plane, pilot subcarriers were transmitted only occasionally. As a consequence,

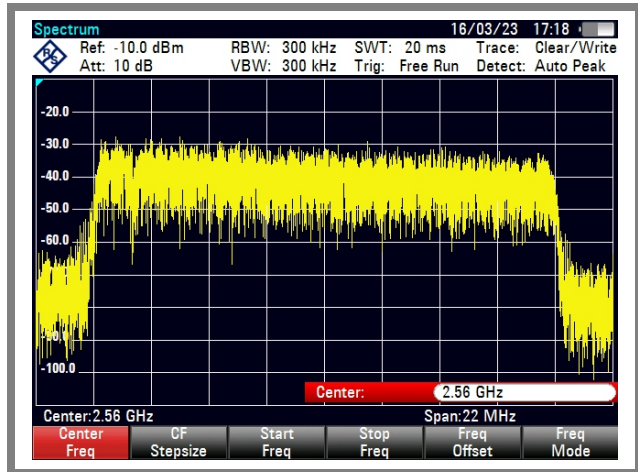


Fig. 14. Spectrum measurements: uplink – distance 7.8 m – 20 MHz bandwidth.

signal spectrum measurements required long averaging times. Commercial-grade off-the-shelf UE, unlike the srsUE virtual terminal, was equipped with a transmission power control mechanism. The effects of this mechanism were visible when the UE was being moved away from the base station. The transmission power on the uplink channel varied between  $-42$  and  $-38$  dBm when the UE was located 3.2 m from the base station. If the distance between the UE and the base station increased to 7.8 m, the transmitted power increased to  $-31$  dBm measured at a distance of 10 cm from the UE.

The cellular system setup considered in this publication enables the development of an open-source radio access network. This is useful for analyzing the operation of the protocol layers of the LTE network standard. The system can be used for propagation tests in the evolving radio environment at a desired location. This is possible thanks to the high degree of hardware mobility and flexibility of software configuration, as well as the fact that a wide range of radio modules and various antenna types are supported. On the other hand, high scalability of this system allows to optimize the configuration using a small-scale network. Finally, it is possible to reproduce it in a large, commercial RAN system.

## Acknowledgments

The work has been performed in 2023 under the project no. 0312/SBAD/8163 titled “New methods of signal transmission and reception, design, and testing of radio communication systems”.

## References

- [1] B. Maqsood, “Implementation and performance analysis of software defined radio (SDR) based LTE platform for truck connectivity application”, Master Thesis, KTH Royal Institute of Technology, Stockholm, Sweden 2019 (<http://www.diva-portal.org/smash/record.jsf?pid=diva2:1413149>).
- [2] A. Kaszuba, R. Chęciński, and J. Łopatka “Capture information about GSM users using software defined radio platform USRP”, *Bulletin of the Military University of Technology*, vol. 62, no. 3, pp. 27–36, 2013

- (<http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-8a873ca8-67c3-4ec6-bad0-e802b8d08741/c/Kaszuba.pdf>) (in Polish).
- [3] P. Sutton *et al.*, “Open Source RAN”, Software Radio System RS RAN Project, 2023 (<https://www.srslte.com>)
  - [4] “The TCP, UDP and SCTP network bandwidth measurement tool”, iPerf, 2019 (<https://iperf.fr>).
  - [5] “Software Radio Systems” – srsRAN Documentation, Release 22.10, 24 Nov 2022 ([https://docs.srsran.com/\\_downloads/4g/en/rfsoc/pdf](https://docs.srsran.com/_downloads/4g/en/rfsoc/pdf)).
  - [6] “Knowledge Base”, Ettus Research, 2023 ([https://kb.ettus.com/Knowledge\\_Base](https://kb.ettus.com/Knowledge_Base)).
  - [7] B. Schulz, “LTE Transmission Modes and Beamforming – White Paper”, Rohde & Schwarz Company, 2015 ([http://cdn.rohde-schwarz.com/pws/dl\\_downloads/dl\\_application/application\\_notes/1ma186/1MA186\\_2e\\_LTE\\_TMs\\_and\\_beamforming.pdf](http://cdn.rohde-schwarz.com/pws/dl_downloads/dl_application/application_notes/1ma186/1MA186_2e_LTE_TMs_and_beamforming.pdf)).
  - [8] R. Hu and Y. Qian, “*Heterogeneous Cellular Networks*”, Wiley 2013 (ISBN: 9781119999126).
  - [9] “5G Core Network” Whitepaper”, Telcoma Global Whitepaper, (<https://telcomaglobal.com/courses/6g-5g-4g-white-papers/lectures/34560661>).
  - [10] J. Sungho, “LTE and Network Evolution”, *ITU-T Workshop on Bridging the Standardization Gap and Interactive Training Session*, Nadi, Fiji, 2011 ([https://www.itu.int/dms\\_pub/itu-t/oth/06/4D/T064D0000020072PDFE.pdf](https://www.itu.int/dms_pub/itu-t/oth/06/4D/T064D0000020072PDFE.pdf)).
  - [11] Recommendation ITU M.2150, “Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications 2020 (IMT-2020)” – ITU-R, 2022 (<https://www.itu.int/rec/R-REC-M.2150/en>).
- 

**Wojciech Flakowski, M.Sc.**

Institute of Radiocommunications, Faculty of Computing and Telecommunications

 <https://orcid.org/0000-0001-6915-3687>

E-mail: wojciech.flakowski@doctorate.put.poznan.pl

Poznan University of Technology, Poznan, Poland

<https://cat.put.poznan.pl>

**Maciej Krasicki, Ph.D., Associate Professor**

Institute of Radiocommunications, Faculty of Computing and Telecommunications

 <https://orcid.org/0000-0001-7726-3114>

E-mail: maciej.krasicki@put.poznan.pl

Poznan University of Technology, Poznan, Poland

<https://cat.put.poznan.pl>

**Rafał Krenz, Ph.D., Associate Professor**

Institute of Radiocommunications, Faculty of Computing and Telecommunications

 <https://orcid.org/0000-0001-5354-9812>

E-mail: rafal.krenz@put.poznan.pl

Poznan University of Technology, Poznan, Poland

<https://cat.put.poznan.pl>

# An Efficient Cooperative Spectrum Sensing Method Using Renyi Entropy Weighted Optimal Likelihood Ratio for CRN

Sandip B. Shrote and Sadhana D. Poshattiwar

Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

<https://doi.org/10.26636/jtit.2023.3.1360>

**Abstract** — The main concept behind employing cognitive radio is to enable secondary users (SUs) or unlicensed users to utilize the available spectrum. Spectrum sensing methods detect the existence of primary users (PUs) and have become the main topic of research in the CRN industry and in academia. This paper proposes a new framework based on the Adam gradient descent (Adam GD) algorithm to develop a spectrum sensing mechanism used in CRNs and detecting the availability of free channels. The signal's components are extracted from the received signal and the spectrum is searched for availability which is detected through a fusion center using the proposed algorithm. The proposed Adam GD algorithm attains the maximum detection probability rate and the minimum false alarm probability of 0.71 and 0.39, respectively, for a Rayleigh channel.

**Keywords** — cognitive radio, eigen statistics, optimal weight, signal energy, spectrum sensing

## 1. Introduction

The conventional static spectrum allocation mechanism allows certain primary users (PU) to access the spectrum, whereas secondary users (SU) are restricted [1], [2]. To increase spectrum efficiency, the cognitive radio technology is introduced to allow SU to utilize the licensed spectrum band during the period of PU inactivity [3]–[5]. CR is a solution that may potentially solve the issue of spectrum scarcity by accessing the spectrum in a dynamic manner and serves as a key 5G enabler [2]. CRNs adjust the radio framework based on their own knowledge. They also efficiently and opportunistically access the licensed spectrum thus increasing its usage rate [5]. The CR technology is considered to be the primary solution for spectrum underusage, as it provides access to spectrum bands that are not free from PUs, without any interference. Therefore, a reliable and accurate spectrum sensing method is required on the SU side before communication may be established.

A key issue in spectrum sensing is to model test statistics in order to achieve higher probability of detection (PD) [4], [8]–[10]. The following solutions are used for this purpose: cyclostationary based method [11], energy detection model [12], matched filter approach [13], eigenvalue-based approach [14], and likelihood-based scheme [15]. The cyclostationary based detection mechanism offers improved performance, but it is

sensitive to synchronization and faces computational complexity issues [16]. It is capable of efficiently detecting spectrum holes, but it requires longer sensing times [6]. The spectrum sensing method based on the machine learning approach is designed to increase the performance of the process [17]. A spectrum sensing approach based on convolutional neural networks was developed in [4]. In [18], deep learning is utilized for spectrum sensing purposes. In recent decades, various deep learning-based detection models have been developed for spectrum sensing. In [19], a long short-term memory (LSTM) network model is designed for detecting PUs in a spectrum sensing scenario. A convolutional neural network (CNN)-based model is used to learn the energy correlation textures of PU signals in [20].

This paper proposes a spectrum sensing framework based on the Adam GD algorithm and relying on the extraction of signal components. At the receiver end, the signal is analyzed for energy by using test statistics relying on the generalized likelihood ratio test (GLRT), Renyi entropy and eigen statistics. The extracted signal components are then fused by weight parameters. The value of each weight parameter is determined using the Adam GD algorithm, i.e. a combination of Adam optimization with the gradient descent (GD) algorithm. Finally, spectrum availability is detected using the proposed algorithm in order to allocate the free channels to unlicensed users.

The remaining portion of the paper is organized as follows. Section 2 reviews different conventional schemes, while Section 3 portrays the model of a CRN. Section 4 describes the spectrum availability sensing approach developed by the authors. Section 5 discusses the results achieved, while Section 6 presents the conclusions.

## 2. Related Works

Kaur *et al.* [21] designed a decentralized multi-agent reinforcement learning approach relied upon in CRNs to efficiently utilize the spectrum. It reduces the operating cost and offers good flexibility. It increases performance in terms of outage probability, network capacity and convergence speed, but it fails to consider different application-specific requirements of CR in the context of cooperative networks. Eappen *et al.* [6] designed a multi-objective modified grey wolf opti-

mization (MOMGWO) algorithm for solving the optimization issues in CRN. It obtained better detection threshold, sensing time, and transmission power values by compromising energy efficiency and interference immunity. However, only a single PU was considered in this method. Zhang *et al.* designed, in [16], a kernel-based sensing approach for CRNs. It maps the input signal matrix to the higher dimensional of features through a non-linear Gaussian function and generates the test statistics in the feature space. It achieved better performance when faced with interference and Gaussian noise, but it proved sensitive to the level of Gaussian noise. Liu *et al.* [4] developed a deep neural network (DNN) based model for exploring test statistics in CRN. It used a covariance matrix as the CNN input for the sensing process, to increase performance. It achieved higher detection and false alarm probability rates, but faced greater computational complexity issues.

Eappen *et al.* [7] designed a hybrid particle swarm optimization with gravitational search algorithm (PSO-GSA) for detecting spectrum holes by analyzing energy usage. It effectively detects the holes by finding optimal values of transmission power, sensing bandwidth, and power spectral density. It increased energy efficiency, but its higher population value increases computational complexity.

Pan *et al.* [22] designed a spectrum sensing model with deep learning and the cycle spectrum mechanism in CRN. It analyzes cyclic autocorrelation of the OFDM signal by relying on the cyclic spectrum captured through the time domain fast Fourier transform (FFT) algorithm. Cyclic spectrum was normalized to the gray scale dispensation in order to form the gray scale image. Deep features were learned through CNN and the sensing process was performed in the cyclic spectrum. Performance was increased in the presence of noise, but different issues were faced with single feature inputs.

Jothiraj *et al.* [23] modeled an optimized sensing model using the dragonfly algorithm. It increased the usage of spectrum by allocating the free spectrum to unlicensed or SUs in the network. It used the adaptive threshold factor to detect spectrum holes. This model outperforms other techniques in terms of efficiency and detection accuracy. Reddy *et al.* [25] designed an improved whale optimization algorithm for spectrum sensing in CRN. The weight function for CNN was computed to analyze its accuracy performance. It increased the performance in terms of such measures as energy consumption, delivery ratio, and delay.

Some of the issues faced by the conventional approaches include the following:

- The use of machine learning in the resource allocation mechanism results in outstanding performance, better than that of other simple convex algorithms relied upon in the dual decomposition model, with the consideration of imperfect channel state information (CSI). However, considering different application-related requirements in cooperative CRN settings, the significant difficulties continue to be faced [21].
- The use of machine learning in the resource allocation mechanism results in outstanding performance, better than

that of other simple convex algorithms relied upon in the dual decomposition model, with the consideration of imperfect channel state information (CSI). However, considering different application-related requirements in cooperative CRN settings, the significant difficulties continue to be faced [21].

- MOMGWO is an efficient method for defining multi-objective functions with the best Pareto front. Nevertheless, it continues to face problems related to the computational time required [6].
- Kernel-based detection is not effective in terms of certain factors concerning Gaussian mixture noise, but its efficiency was improved by increasing the number of receiver antennas, samples and by modifying the signal-to-noise-ratio (SNR) [16].
- DNN achieved higher efficiency, but it faced issues related to signal-to-noise-ratio complexity [4].
- The extreme population size increases computational complexity of optimization algorithms [7].

### 3. System Description

The cooperative spectrum sensing approach used for detecting PUs in multiple-input multiple-output (MIMO) systems is modeled as a binary hypothesis corresponding to the absence and presence of PUs [29]. The CRN contains a fusion center with  $X_a$  receiver antennas and  $X$  cooperating SUs with  $X_b$  antennas for every cooperating user. Let us consider  $p_{d,m}$  as probability of detection, and  $p_{f,m}$  as probability of a false alarm of a local decision rule of the  $m$ -th user. When the  $m$ -th secondary node made the decision, the node transmits the set of  $A$  vectors  $G_m(h) \in \mathbb{R}^{X_b \times 1}$ ,  $1 \leq h \leq A$  corresponding to the local decision concerning the absence or presence of the PU, to the fusion center. Matrix  $G_m$  is generated by stacking all sets of  $A$  vectors  $g_m(h)$ ,  $1 \leq h \leq A$  corresponding to the decision, as:

$$G_m(h) = [g_m(1), g_m(2), \dots, g_m(A)]^B \in \mathbb{R}^{A \times X_b}.$$

Next, matrix  $G_m$  is sent towards the fusion center to get the values of  $G_m = P_0$  or  $G_m = P_1$  corresponding to the decision regarding the absence  $\aleph_0$  or presence  $\aleph_1$  of PUs, respectively. Under a non-antipodal signal with  $A = 2$  and  $X_b = 2$ , matrix  $P_0 = 0_{2 \times 2}$  and  $P_1$  is selected as an orthogonal matrix  $P = 0_{2 \times 2}$  and  $P_1$  is selected as an orthogonal matrix  $[1 \ 1 \ 1 \ -1]_{2 \times 2}$  with  $P_1 P_1^B = AC_2$ . Accordingly, local performance for the  $m$ -th secondary user, i.e.  $p_{d,m}$  and  $p_{f,m}$ , is:

$$p_{d,m} = \text{pr}(G_m = P_1 | \aleph_1), \quad (1)$$

$$1 - p_{d,m} = \text{pr}(G_m = P_0 | \aleph_1), \quad (2)$$

$$p_{f,m} = \text{pr}(G_m = P_1 | \aleph_0), \quad (3)$$

$$1 - p_{d,m} = \text{pr}(G_m = P_0 | \aleph_0). \quad (4)$$

It is assumed that the number of antennas located at the PU, as well as detection mechanism utilized at every cooperating SU are chosen arbitrarily. Accordingly, signal  $J_m(h) \in \mathbb{R}^{X_a \times 1}$  received at the fusion center and corresponding to vector

$g_m(h)$  sent by the  $m$ -th ( $1 \leq m \leq X$ ) SU at the  $m$ -th ( $1 \leq h \leq A$ ) instant over the orthogonal MAC is specified as:

$$J_m(h) = \mathbf{B}_m g_m(h) + q_m(h), \quad (5)$$

where  $\mathbf{B}_m \in \mathbb{R}^{X_a \times X_b}$  denotes the MIMO channel matrix and every element  $k_m(x, y)$  of matrix  $\mathbf{B}_m$  shows a fading coefficient between the  $y$ -th transmitting antenna of the  $m$ -th SU and the  $x$ -th receiving antenna of the fusion center. The  $l$ -th row of channel matrix  $\mathbf{B}_m = [k_{m,1}, k_{m,2}, \dots, k_{m,X_a}]^B$  is defined by  $k_{m,l}^B \in \mathbb{R}^{1 \times X_b}$  and vector  $q_m(h) \in \mathbb{R}^{X_a \times 1}$  denotes circularly symmetric additive white Gaussian noise that is showed as identically separated and independent with the value of zero mean, such that the covariance matrix is represented as  $\mathbf{S}_q = E\{q_m(h)q_m^B(h)\} = \lambda^2 C_{X_a}$ . The received signal  $v_{m,l}(h)$  at the  $l$ -th receive antenna of the fusion center ( $1 \leq l \leq X_a$ ) is modeled as:

$$v_{m,l}(h) = k_{m,l}^B g_m(h) + q_{m,l}(h), \quad (6)$$

where vector  $J_m(h)$  in Eq. (5) is represented as  $J_m(h) = [v_{m,1}(h), v_{m,2}(h), \dots, v_{m,X_a}(h)]^T \in \mathbb{R}^{X_a \times 1}$  and the term  $q_{m,l}(h)$  implies the  $l$ -th element of noise vector. Figure 1 illustrates the model of a cooperative MIMO CRN.

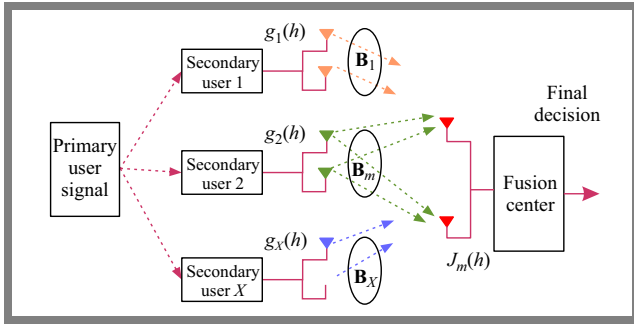


Fig. 1. Model of a cooperative MIMO cognitive radio network.

## 4. Proposed Adam GD Algorithm for Spectrum Sensing

To enable detecting availability of the free spectrum and to allocate the licensed bands to SUs, a spectrum sensing framework based on the Adam GD algorithm has been designed, relying on the fusion center mechanism. At first, few signal components are collected and next they are fuse based on the weights assigned to specific signal components by the Adam GD algorithm. A block diagram of the proposed model is presented in Fig. 2.

The first step is to define the signal received at the antenna with the use of a mathematical expression in order to further extract signal components needed to perform the fusion strategy required for making the final decision. The signal received at the  $l$ -th antenna of the fusion center for a set of  $A$  concatenated transmitted vectors that corresponds to the decision of the  $m$ -th SU is expressed as:

$$J_{m,l}(h) = G_m k_{m,l} + q_{m,l}, \quad (7)$$

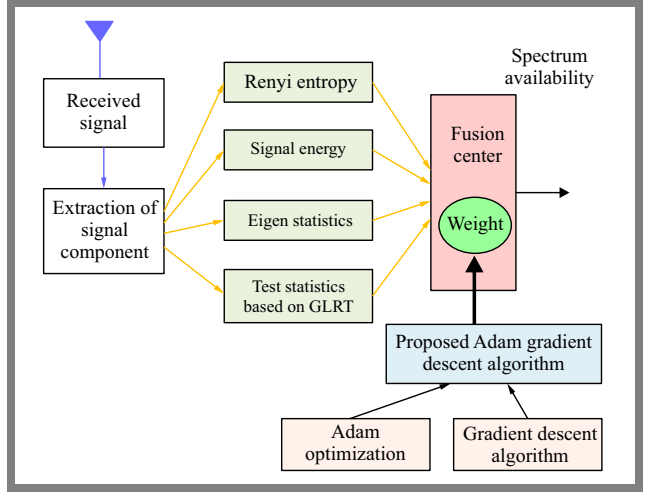


Fig. 2. Block diagram of the proposed, fusion center-based method.

where  $J_{m,l} = [v_{m,l}(1), v_{m,l}(2), \dots, v_{m,l}(A)]^B \in \mathbb{R}^{A \times 1}$ . The stacked noise vector  $q_{m,l} \in \mathbb{R}^{A \times 1}$  is generated as  $q_{m,l} = [u_{m,l}(1), u_{m,l}(2), \dots, u_{m,l}(A)]^B$ .

### 4.1. Extraction of Signal Components

After receiving signal  $J_{m,l}$ , other signal components, such as Renyi entropy, signal energy, eigen statistics, and test statistics based on GLRT need to be extracted. Let us assume a CRN scenario with  $r$  transmitters and  $s$  receivers or sensors, such that each transmitter is engaged in communication within the CRN radio network. To enable communication between the transmitter and the sensor, there must exist a free channel in the CR environment. The proposed framework for determining channel availability relies on the following measures.

**Renyi entropy measure** for extracting a signal component is defined as:

$$RE = \frac{1}{1 - J_{m,l}} \ln \sum_{m=1}^X \sum_{l=1}^{X_a} J_{m,l}, \quad (8)$$

where denotes the Renyi entropy.

**Signal energy and eigen statistics** – this measure uses information obtained by receivers from transmitters [28] to generate data matrix  $\mathbf{D}$  as:

$$D = [D_{cn}]_{M \times N}, \quad 1 \leq c \leq M, \quad 1 \leq n \leq N, \quad (9)$$

where  $D_{cn}$  implies  $n$ -th data received by  $c$ -th receiver in the CRN and  $N$  denotes the total number of signals received by sensors. Accordingly, matrix  $\mathbf{D}$  is made up of the transmitted signals, thermal noise, and channel gain. A signal matrix  $\mathbf{S}$  is formed by the signals that are sent by the transmitter and is specified as:

$$S = [S_{zn}]_{r \times N}, \quad (10)$$

where  $r$  implies the total number of transmitters, and  $N$  shows all data samples,  $S_{zn}$  represents the signal obtained from the  $z$ -th transmitter. Hence, communication is established through the radio channel and is processed with the use of an unoccupied channel. The data sample matrix is represented as:

$$D = [R * Q] + L, \quad (11)$$

where  $R$  is channel gain,  $L$  indicates noise matrix, and  $Q$  represents desired data. The channel matrix is estimated as:

$$R = [R_{cz}]_{s \times r}, \quad (12)$$

where  $R_{cz}$  denotes channel gain between the  $c$ -th receiver and the  $z$ -th transmitter. In general, signals transmitted by radio channels in CRN are degraded by noise, such that noise  $L$  is assumed while creating channel matrix  $D$  and thermal noise matrix is specified as:

$$K = [K_{cn}]_{M \times N}, \quad (13)$$

where  $K_{cn}$  is the thermal noise of the channel used to receive information from the  $c$ -th sensor. Next, the data sample matrix is utilized for calculating the signal energy and eigen statistics. The covariance matrix  $H$  is calculated from  $D$ , in which the binary hypothesis test is considered. The ensemble covariance matrix of the received signal is represented as:

$$H = \text{expected}[D D^+], \quad (14)$$

where  $\text{expected}[\cdot]$  is the expected operator value and  $D^+$  implies the conjugate or the transpose of  $D$ . At first, the ensemble covariance matrix is not known. Therefore, it is replaced with the maximum likelihood estimate (MLE) for specifying the sample as:

$$H = \frac{1}{p} [D D^+]. \quad (15)$$

By considering covariance matrix  $H$ , signal energy and Renyi entropy, eigen statistics and test statistics based on GLRT are used for estimating spectrum availability. The eigen values of  $H$  are estimated as:

$$\{\gamma_1 \geq \gamma_2 \geq \dots \gamma_r\}. \quad (16)$$

The energy statistics are:

$$\gamma = \frac{\gamma_1}{\frac{1}{s} \sum_{c=1}^M \gamma_c}, \quad (17)$$

where  $\gamma_c$  shows the eigen value of the  $c$ -th sensor. The eigen statistics are:

$$I = \frac{1}{r \times \eta^2} \sum_{c=1}^M \gamma_c, \quad (18)$$

where  $\eta$  indicates the thermal noise factor.

GLRT-based test statistics [29] for local detection of the respective users are:

$$TS = \sum_{m=1}^X \sum_{l=1}^{X_a} J_{m,l}^B \alpha J_{m,l}, \quad (19)$$

where  $TS$  denotes the test statistics and the parameter  $\alpha$  is given as:

$$\alpha = C - \mu, \quad (20)$$

where  $\mu$  shows the projection matrix.

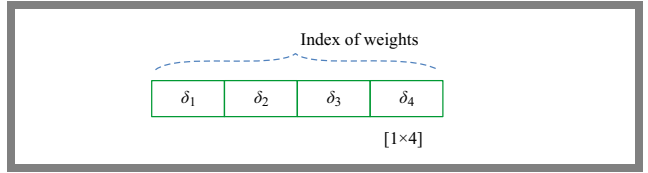


Fig. 3. Solution encoding.

#### 4.2. Fusion Center

After extracting individual components from the received signal, they need to be fused. The fusion mechanism relies on combining all signal components with applicable weight values assigned thereto:

$$F = \delta_1 \gamma + \delta_2 I + \delta_3 TS + \delta_4 RE, \quad (21)$$

where  $F$  denotes probability of detection,  $\delta_1 \dots \delta_4$ , are the weight factors,  $RE$  shows signal component obtained through the Renyi entropy measure,  $I$  shows the eigen statistics value,  $\gamma$  is the signal energy and  $TS$  represents test statistics based on GLRT.

#### 4.3. Proposed Adam GD Algorithm

The optimal weight for the fusion process is determined using the Adam GD algorithm derived through the incorporation of the GD algorithm from [27] with the Adam optimization technique from [26]. The GD algorithm is a popular optimization algorithm, also referred to as the black box optimizer, since its strengths and weaknesses are hard to explain. It is mainly used to determine the cost function based on the  $\theta$  parameter for a training set. Each update process requires that a gradient for the entire dataset be computed. When the dataset does not fit to memory, it becomes slow and it is also intractable. The Adam algorithm is an effective stochastic optimization method that needs a first order gradient and is characterized by limited memory usage. It calculates the adaptive learning rate for different factors by estimating the first and the second moments of gradients. The benefits of Adam optimization are that it updates the magnitude factor such that it is invariant to gradient rescaling and it works well with sparse gradients.

The solutions are encoded in the form of a vector, as  $\delta_1, \delta_2, \delta_3$ , and  $\delta_4$  with the dimension of  $[1 \times 4]$  (Fig. 3). Four different weight factors are determined using the proposed algorithm.

The proposed Adam GD algorithm is divided into the following steps:

**Initialization.** It defines the parameters for the entire training set in order to perform the updating process.

**Fitness measure.** Fitness is measured based on the area under the receiver operating characteristic (ROC) curve, i.e. a quantitative metric used for comparing various detectors. It is an expression used to determine optimal solution during iteration [?] and is expressed as:

$$F = \int_0^1 P_d(t) dp_f(t) dt, \quad (22)$$

where  $t$  represents the specific detector,  $p_d$  is the probability of detection, and  $p_f$  signifies the probability of a false alarm.

**Update solution.** The update vector of the GD algorithm is

formulated as:

$$Y_w = \beta Y_{w-1} + \chi \nabla_{\theta} Z(\theta), \quad (23)$$

The standard expression of Adam optimization is specified as:

$$Y_w = Y_{w-1} - \frac{\tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}}, \quad (24)$$

$$Y_{w-1} = Y_w + \frac{\tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}}, \quad (25)$$

By substituting Eq. (25) to Eq. (23) we get:

$$Y_w = \beta \left( Y_w + \frac{\tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}} \right) + \chi \nabla_{\theta} Z(\theta), \quad (26)$$

$$Y_w = \beta Y_w + \frac{\beta \cdot \tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}} + \chi \nabla_{\theta} Z(\theta), \quad (27)$$

$$Y_w - \beta Y_w = \frac{\beta \cdot \tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}} + \chi \nabla_{\theta} Z(\theta), \quad (28)$$

$$Y_w(1 - \beta) = \frac{\beta \cdot \tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}} + \chi \nabla_{\theta} Z(\theta), \quad (29)$$

$$Y_w = \frac{1}{1 - \beta} \left[ \frac{\beta \cdot \tilde{\omega} \cdot \hat{i}_w}{\sqrt{\hat{j}_w + \varepsilon}} + \chi \nabla_{\theta} Z(\theta) \right], \quad (30)$$

$$Y_w = \frac{\beta \cdot \tilde{\omega} \cdot \hat{i}_w + \chi \nabla_{\theta} Z(\theta) \sqrt{\hat{j}_w + \varepsilon}}{\sqrt{\hat{j}_w + \varepsilon} (1 - \beta)}, \quad (31)$$

where  $\varepsilon = 10^8$ ,  $\beta$  represents the momentum term and equals 0.9,  $\chi$  specifies the size of the steps taken to reach the minimum,  $Z(\theta)$  signifies the objective function,  $w$  indicates the time step and  $\hat{i}_t = \frac{i_t}{1 - \rho_t}$ . Here  $\rho_t$  denotes the decay rate within the 0...1 range.

**Evaluating feasibility.** The fitness of each solution is computed based on the fitness measure function such that the optimal solution shows the best value for specific weight factors.

**Termination.** The above steps are repeated until the optimal solution is attained. Algorithm 1 summarizes the proposed method.

**Algorithm 1.** Pseudo code of proposed model.

- 1: Input: Momentum term, decay rate, and step size.
- 2: Output:  $Y_w$ .
- 3: Initialize the factors.
- 4: Compute fitness measure.
- 5: Derive update solution of  $Y_w$  using Eq. (31).
- 6: Evaluate feasibility.
- 7: Return best solution.

## 5. Results and Discussion

The proposed model is evaluated using Matlab and is compared with existing approaches, such as kernel-based spectrum sensing [16], DNN [4], cooperative reinforcement learning [22], and deep learning [23] to show the extent of the improvements achieved.

Figure 4 analyzes the detection probability for a Rayleigh channel. At SNR of 5 dB, detection probability achieved by kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.47, 0.39, 0.33, 0.53, and 0.71, respectively (Fig. 4a). A false alarm probability analysis is shown in Fig. 4b. For SNR of 5 dB, the probability measure for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.76, 0.58, 0.52, 0.44, and 0.39, respectively. Figure 4c shows an analysis of ROC. For the probability of detection equaling 0.6, the false alarm probability rate obtained by kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and Adam GD algorithm is 0.47, 0.42, 0.32, 0.18, and 0.50, respectively. Figure 5 shows the results of a detection probability analysis for a Rician channel. For SNR of 10 dB, the detection probability of kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.93, 0.86, 0.95, 0.78, and 0.96, respectively (Fig. 5a). Results of the false alarm probability are shown in Fig. 5b. For SNR of 5 dB, the probability measure for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.79, 0.51, 0.58, 0.46, and 0.40, respectively. Analysis of ROC for a detection value of 0.7 gives a false alarm probability for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm of 0.27, 0.53, 0.55, 0.66, and 0.78, respectively (Fig. 5c).

Figure 6 presents the results of a detection probability analysis for a Gaussian channel. When SNR=10 dB, the detection value for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.78, 0.67, 0.86, 0.89, and 0.91, respectively. The false alarm probability shown in Fig. 6b for SNR of 10 dB for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.92, 0.83, 0.94, 0.72, and 0.64, respectively. Figure 6c shows the ROC. For a detection value of 0.8, the false alarm probability for kernel-based spectrum sensing, DNN, cooperative reinforcement learning, deep learning, and the Adam GD algorithm is 0.64, 0.55, 0.69, 0.80, and 0.87, respectively. Table 1 summarizes the results of all analyses performed.

## 6. Conclusion

Spectrum sensing is a key technology used in CRNs. The objective of this research is to determine the availability of free spectrum using the proposed Adam GD algorithm. Here, the signal received at the receiver end is used to extract individual signal components, such as energy, Renyi entropy, test statistics, and eigen statistics. Accordingly, the signal's components are fused at the fusion center by applying weight factors determined using the Adam GD algorithm. The proposed approach achieves the maximum probability of detection and the minimum probability of a false alarm equaling 0.7099 and 0.3892, respectively, for a Rayleigh channel.

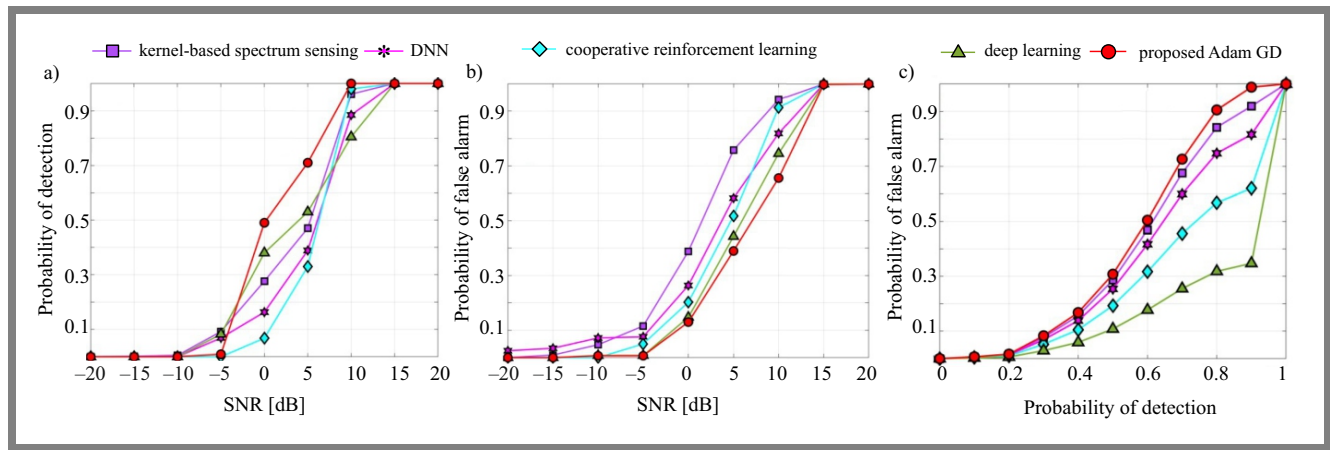


**Fig. 4.** Rayleigh channel analysis results: a) probability of detection, b) probability of false alarm, and c) ROC.

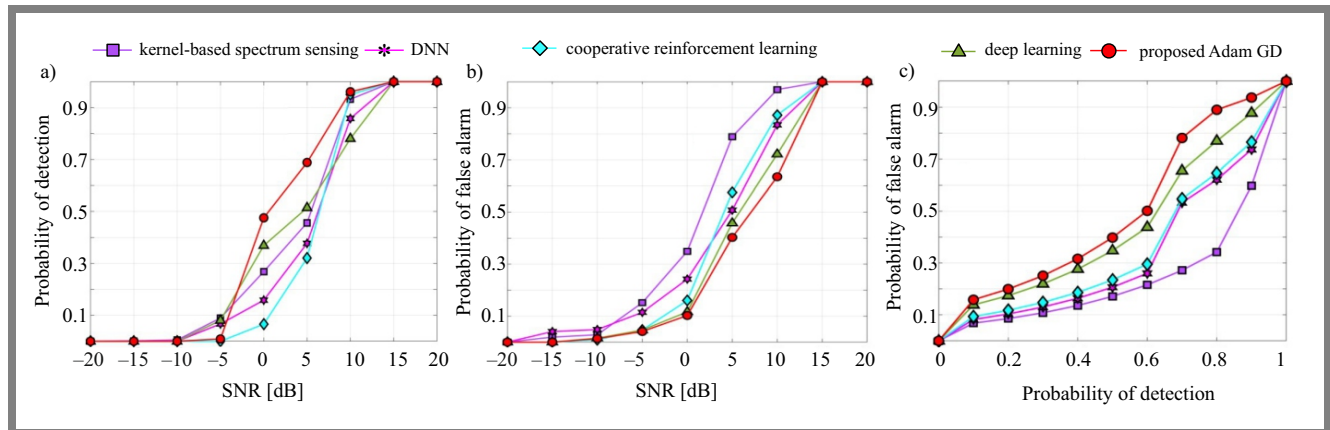


**Fig. 5.** Rician channel analysis results: a) probability of detection, b) probability of false alarm, and c) ROC.

**Tab. 1.** Analysis summary for SNR = 5 dB.

| Channel type | Metrics                    | Kernel-based spectrum sensing | DNN    | Cooperative reinforcement learning | Deep learning | Proposed Adam GD algorithm |
|--------------|----------------------------|-------------------------------|--------|------------------------------------|---------------|----------------------------|
| Rayleigh     | Probability of detection   | 0.4704                        | 0.3889 | 0.3303                             | 0.5300        | 0.7099                     |
|              | Probability of false alarm | 0.7578                        | 0.5820 | 0.5173                             | 0.4423        | 0.3892                     |
| Rician       | Probability of detection   | 0.4563                        | 0.3772 | 0.3204                             | 0.5141        | 0.6886                     |
|              | Probability of false alarm | 0.7888                        | 0.5077 | 0.5759                             | 0.4578        | 0.4029                     |
| Gaussian     | Probability of detection   | 0.5967                        | 0.5105 | 0.5319                             | 0.6693        | 0.6909                     |
|              | Probability of false alarm | 0.7781                        | 0.4570 | 0.5723                             | 0.4656        | 0.4097                     |

## References

- [1] M. Amjad, M.H. Rehmani, and S. Mao, "Wireless Multimedia Cognitive Radio Networks: A Comprehensive Survey", *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1056–1103, 2018 (<https://doi.org/10.1109/COMST.2018.2794358>).
- [2] Y. Chen and H.S. Oh, "A Survey of Measurement-based Spectrum Occupancy Modeling for Cognitive Radios", *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 848–859, 2014 (<https://doi.org/10.1109/COMST.2014.2364316>).
- [3] J. Lunden, V. Koivunen, and H.V. Poor, "Spectrum Exploration and Exploitation for Cognitive Radio: Recent Advances", *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 123–140, 2015 (<https://doi.org/10.1109/MSP.2014.2338894>).
- [4] C. Liu, J. Wang, X. Liu, and Y.C. Liang, "Deep CM-CNN for Spectrum Sensing in Cognitive Radio", *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2306–2321, 2019 (<https://doi.org/10.1109/JSAC.2019.2933892>).
- [5] Y. Ma, Y. Gao, Y.C. Liang, and S. Cui, "Reliable and Efficient Sub-Nyquist Wideband Spectrum Sensing in Cooperative Cognitive Radio Networks", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2750–2762, 2016 (<https://doi.org/10.1109/JSAC.2016.2605998>).
- [6] G. Eappen and T. Shankar, "Multi-Objective Modified Grey Wolf Optimization Algorithm for Efficient Spectrum Sensing in the Cognitive Radio Network", *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3115–3145, 2021 (<https://doi.org/10.1007/s13369-020-05084-3>).

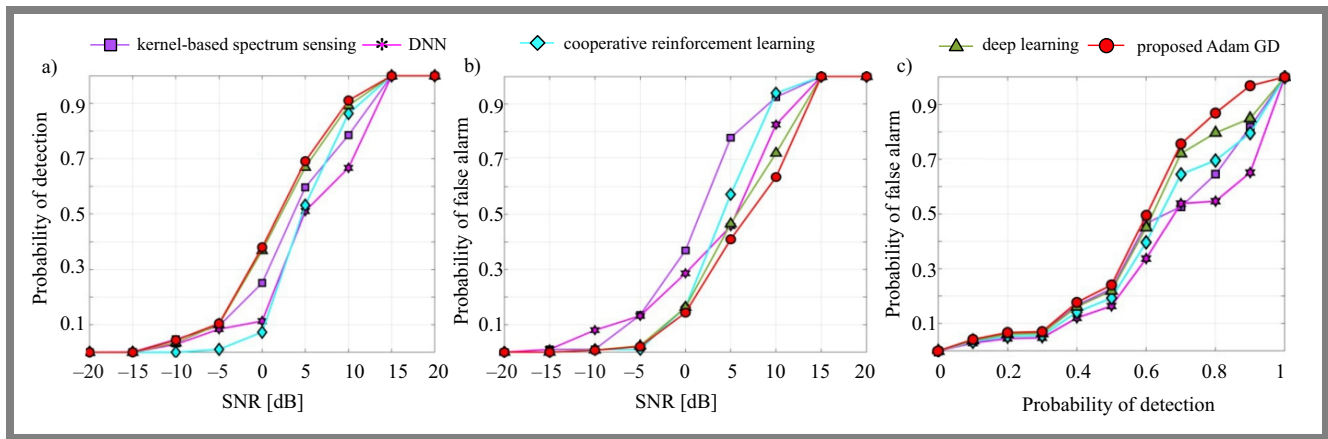


**Fig. 6.** Gaussian channel analysis: a) probability of detection, b) probability of false alarm, and c) ROC.

- [7] G. Eappen and T. Shankar, "Hybrid PSO-GSA for energy efficient spectrum sensing in cognitive radio network", *Physical Communication*, vol. 40, art. no. 101091, 2020 (<https://doi.org/10.1016/j.phycom.2020.101091>).
- [8] H.A. Shah and I. Koo, "Reliable machine learning based spectrum sensing in cognitive radio networks", *Wireless Communications and Mobile Computing*, vol. 2018, art. no. 5906097, 2018 (<https://doi.org/10.1155/2018/5906097>).
- [9] H. He and H. Jiang, "Deep Learning Based Energy Efficiency Optimization for Distributed Cooperative Spectrum Sensing", *IEEE Wireless Communications*, vol. 26, no. 3, pp. 32–39, 2019 (<https://doi.org/10.1109/WMC.2019.1800397>).
- [10] Y. Arjouni and N. Kaabouch, "A Comprehensive Survey on Spectrum Sensing in Cognitive Radio Networks: Recent Advances, New Challenges, and Future Research Directions", *Sensors*, vol. 19, no. 1, art. no. 126, 2019 (<https://doi.org/10.3390/s19010126>).
- [11] Y. Zeng and Y.C. Liang, "Robustness of the Cyclostationary Detection to Cyclic Frequency Mismatch", in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Istanbul, Turkey, pp. 2704–2709, 2010 (<https://doi.org/10.1109/PIMRC.2010.5671799>).
- [12] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple Antenna Spectrum Sensing in Cognitive Radios", *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, 2010 (<https://doi.org/10.1109/TWC.2009.02.090385>).
- [13] H.S. Chen, W. Gao, and D.G. Daut, "Signature Based Spectrum Sensing Algorithms for IEEE 802.22 WRAN", in *IEEE International Conference on Communications*, Glasgow, UK, pp. 6487–6492, 2007 (<https://doi.org/10.1109/ICC.2007.1073>).
- [14] C. Liu, J. Wang, X. Liu, and Y.C. Liang, "Maximum Eigenvalue-Based Goodness-of-Fit Detection for Spectrum Sensing in Cognitive Radio", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7747–7760, 2019 (<https://doi.org/10.1109/TVT.2019.2923648>).
- [15] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Spectrum Sensing over MIMO Channels Using Generalized Likelihood Ratio Tests", *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 439–442, 2013 (<https://doi.org/10.1109/LSP.2013.2250499>).
- [16] J. Zhang *et al.*, "MIMO Spectrum Sensing for Cognitive Radio-Based Internet of Things", *IEEE IoT Journal*, vol. 7, no. 9, pp. 8874–8885, 2020 (<https://doi.org/10.1109/JIOT.2020.2997707>).
- [17] F. Azmat, Y. Chen, and N. Stocks, "Analysis of Spectrum Occupancy Using Machine Learning Algorithms", *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 6853–6860, 2015 (<https://doi.org/10.1109/TVT.2015.2487047>).
- [18] S. Zheng *et al.*, "Spectrum Sensing Based on Deep Learning Classification for Cognitive Radios", *China Comm.*, vol. 17, no. 2, pp. 138–148, 2020 (<https://doi.org/10.48550/arXiv.1909.06020>).
- [19] W.M. Lees *et al.*, "Deep Learning Classification of 3.5-GHz Band Spectrograms with Applications to Spectrum Sensing", *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 2, pp. 224–236, 2019 (<https://doi.org/10.1109/TCCN.2019.2899871>).
- [20] J. Xie, J. Fang, C. Liu, and X. Li, "Deep Learning-Based Spectrum Sensing in Cognitive Radio: A CNN-LSTM Approach", *IEEE Communications Letters*, vol. 24, no. 10, pp. 2196–2200, 2020 (<https://doi.org/10.1109/LCOMM.2020.3002073>).
- [21] A. Kaur and K. Kumar, "Imperfect CSI Based Intelligent Dynamic Spectrum Management Using Cooperative Reinforcement Learning Framework in Cognitive Radio Networks", *IEEE Transactions on Mobile Computing*, vol. 27, no. 5, pp. 1672–1683, 2020 (<https://doi.org/10.1109/TMC.2020.3026415>).
- [22] G. Pan, J. Li, and F. Lin, "A Cognitive Radio Spectrum Sensing Method for an OFDM Signal Based on Deep Learning and Cycle Spectrum", *International Journal of Digital Multimedia Broadcasting*, art. no. 5069021, 2020 (<https://doi.org/10.1155/2020/5069021>).
- [23] S. Jothiraj, S. Balu, and N. Rangaraj, "An efficient adaptive threshold-based dragonfly optimization model for cooperative spectrum sensing in cognitive radio networks", *International Journal of Communication Systems*, vol. 34, no. 10, art. no. e4829, 2021 (<https://doi.org/10.1002/dac.4829>).
- [24] S.S. Reddy and M.S.G. Prasad, "Improved Whale Optimization Algorithm and Convolutional Neural Network Based Cooperative Spectrum Sensing in Cognitive Radio Networks", *Information Security Journal: A Global Perspective*, vol. 30, no. 3, pp. 160–172, 2021 (<https://doi.org/10.1080/19393555.2020.1825882>).
- [25] A. Patel, H. Ram, A.K. Jagannatham, and P.K. Varshney, "Robust Cooperative Spectrum Sensing for MIMO Cognitive Radio Networks under CSI Uncertainty", *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 18–33, 2017 (<https://doi.org/10.1109/TSP.2017.2759084>).
- [26] K.U. Chowdary and B.P. Rao, "Hybrid mixture model based on a hybrid optimization for spectrum sensing to improve the performance of MIMO-OFDM systems", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 7, art. no. 2058008, 2020 (<https://doi.org/10.1142/S0218001420580082>).
- [27] S. Ruder, "An overview of gradient descent optimization algorithms", 2016 (<https://doi.org/10.48550/arXiv.1609.04747>).
- [28] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *3rd Int. Conference for Learning Representations*, San Diego, USA, 2014 (<https://doi.org/10.48550/arXiv.1609.04747>).
- [29] S. Ali, G. Seco-Granados, and J.A. López-Salcedo, "Spectrum Sensing with Spatial Signatures in the Presence of Noise Uncertainty and Shadowing", *EURASIP Journal on Wireless Comm. and Netw.*, art. no. 150, pp. 1–16, 2013 (<https://doi.org/10.1186/1687-1499-2013-150>).

**Sandip B. Shrote, Research Scholar**

E-mail: sandip.shrote@gmail.com

Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

<https://www.sggs.ac.in>

**Sadhana D. Poshattiwar, Professor**

Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

<https://www.sggs.ac.in>

# Location-based Power Control Mechanism for D2D Communication Underlying a Cellular System

Marcin Rodziewicz

*Poznan University of Technology, Poznan, Poland*

<https://doi.org/10.26636/jtit.2023.3.1361>

**Abstract** — The paper presents a location-based approach to controlling the power of device-to-device (D2D) underlay of a frequency reuse-1 cellular system. The system allows for direct communication to share uplink resources with cellular users. As a result, both D2D and cellular users are experiencing additional interferences in the system. By controlling the output power of the devices, these interferences can be mitigated and the performance of the network can be improved in terms of better spectral and energy efficiency. The proposed location-based target signal-to-interference ratio power control scheme for D2D communications utilizes information about users' locations to estimate the interference level experienced by the receiver of the direct link. Based on this estimation, an appropriate transmit power can be determined. The performance of the proposed power control solution is investigated via system level simulations.

**Keywords** — context-awareness, device-to-device, frequency reuse-1, power control

## 1. Introduction

In recent years, device-to-device (D2D) communication operating as an underlay in cellular networks has attracted a lot of attention within the research community. It is believed that it can ensure better spectral efficiency of the system and is capable of reducing power consumption, thus improving energy efficiency. Moreover, such a communication solution can enable new types of multimedia services [1]–[3]. Working as an underlay, D2D communications may use the same radio resources as those relied upon by other cellular users. Such an approach allows to increase the frequency reuse factor (FRF) even above the level achieved by the reuse-1 scheme [3]. However, enabling D2D communications that share the spectrum with a cellular telecommunication system poses some new challenges. For instance, new interference patterns emerge which have to be managed, typically with the help of power control or resource allocation mechanisms. This leads to improved performance of the network in terms of better spectral and energy efficiency.

The remainder of the paper is organized as follows. In Section 2, a short review of other works is presented. Section 3 describes the system model under consideration. In Section 4, the proposed power control solution is described, with an

analysis of its performance given in Section 5. The paper is concluded in Section 6.

## 2. Related Work

Numerous studies have been conducted on mitigating interference in D2D communications, i.e. [1], [2], [4]–[9]. The most commonly used approaches involve power control and resource allocation solutions. For example, in [4] a D2D power reduction method was introduced to control interference affecting cellular users. Interference mitigation solutions based on resource allocation often exploit slow-scale parameters, such as path-loss or shadowing, to perform interference-aware resource allocation, as proposed in [7].

Some papers explore the possibility of using location information for resource allocation and for selecting users sharing those resources, i.e. [10]–[13]. In [10] and [11], an interference limited area (ILA) control strategy was proposed. It is employed in addition to a power control mechanism to ensure that the outage probability of D2D communications, caused by interference from cellular users, is lower than a predetermined threshold. Meanwhile, in [12], an interference limited area control method is applied to restrict interference before the resource allocation process. In [13], the distance-constrained resource-sharing criterion (DRC) was introduced to limit the set of cellular users that can share resources with D2D users, resulting in a significant reduction in the D2D link's outage probability. Furthermore, DRC does not require cellular users to reduce their transmission power, thereby avoiding degradation of the cellular link's performance.

## 3. System Model

In this paper, we consider an FRF-1 cellular system in which a D2D underlay is operational, as presented in Fig. 1. The system allows for D2D communications to share uplink resources with co-channel cellular users (CUEs). Consequently, additional interferences are introduced in the system. Interference affecting the receiving D2D device (DUE) is caused when CUEs transmit relying the same resource. On the other hand, from the point of view of CUEs, interference is caused

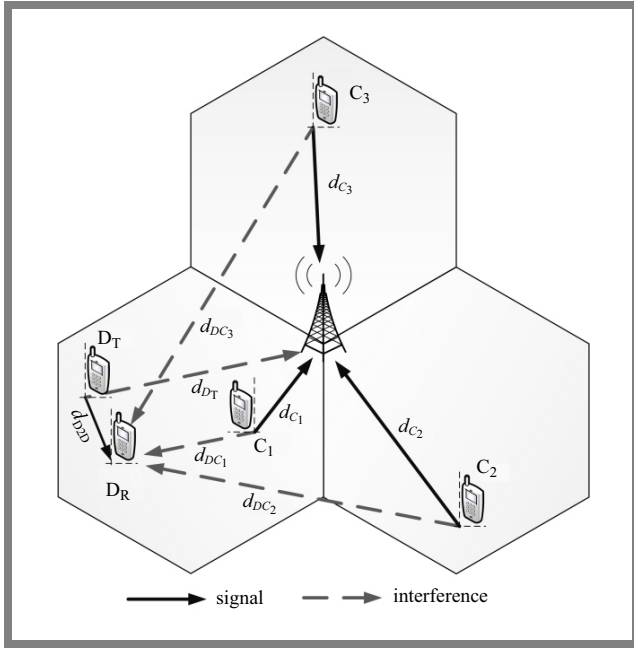


Fig. 1. System model.

by transmitting DUEs and is experienced by the serving base station. The signal-to-noise and interference ratio (SINR) for the DUE receiver  $\gamma_D$  and the BS for a cellular user  $k$  ( $\gamma_{C_k}$ ) are given by:

$$\gamma_D = \frac{h_D(d)P_D}{\sum_{i=1}^N h_{DC_i}(d)P_{C_i} + N_0} \quad (1)$$

and

$$\gamma_{C_k} = \frac{h_{C_k}(d)P_{C_k}}{\sum_{i=1, i \neq k}^N h_{C_i}(d)P_{C_i} + h_D(d)P_D + N_0} \quad (2)$$

where  $N$  is the number of adjacent cells using the same frequency (that includes the cell where the D2D pair is located).  $h_D(d)$  and  $h_{C_i}(d)$  are the distance dependent losses on the path between D2D users and between the DUE receiver and the CUE transmitter  $i$ , respectively,  $h_{C_k}(d)$ ,  $h_{C_i}(d)$  and  $h_D(d)$  are the losses on the path between cellular transmitters  $k$  and  $i$  and the base station and on the path between the DUE transmitter and the base station, respectively. The additive white Gaussian noise (AWGN) variance is denoted by  $N_0$ .  $P_D$  defines the transmit power and the transmitted signal of DUE and  $P_{C_i}$  stands for the transmit power and transmitted signal of the  $i$ -th CUE transmitter.

In the considered system an open-loop power control (OLPC) mechanism is utilized for the CUE's power settings. The transmit power of the cellular users is given by:

$$P_C = \min(P_0 + A \cdot h(d), P_{max}) \quad (3)$$

where  $P_0$  is the initial power level of the UE,  $A$  is a path-loss compensation factor and  $h(d)$  is the path-loss between the BS and the CUE. The maximum transmit power is limited by  $P_{max}$ . In the system model, we assume that the base station has knowledge about the location of the devices it serves. With the increasing popularity of location-based services (LBS),

the acquisition of UEs' location is becoming less troublesome and applicable data may be obtained in various ways, e.g., via the satellite-based GNSS positioning system.

## 4. Location-based Power Control for D2D

In this paper, we propose a location-based target signal-to-interference ratio (SIR) power control (LTSIPC) scheme for D2D communications. The mechanism utilizes information about users' locations to estimate the interference experienced by the D2D receiver. The proposed LTSIPC is a centralized approach, where the base station is acting as the central unit which, through its own power control mechanism, such as OLPC, obtains information about the transmitting powers of cellular user devices that interfere with the D2D receiver. Using the positions of those devices, the host entity can estimate the distance between them and, consequently, the interference caused to the D2D receiver. Since the proposed LTSIPC is based solely on distance and location, it requires no knowledge of channel coefficients, which allows to reduce the number of channel quality reports in the system. The SIR of the D2D receiver can be expressed as:

$$\zeta_D = \frac{|h_{D2D}|^2 P_D d_{D2D}^{-\alpha}}{\sum_{i=1}^N |h_{DC_i}|^2 P_{C_i} d_{DC_i}^{-\alpha}} \quad (4)$$

where  $h_{D2D}$  and  $h_{DC_i}$  are channel coefficients of the D2D and D2D-CUE links considered,  $d_{D2D}$  and  $d_{DC_i}$  are the distances between D2D devices and the D2D receiver and the cellular user  $i$ , respectively. The path-loss exponent is denoted with  $\alpha$  and the transmit powers of the D2D transmitter and the  $i$ -th CUE are given by  $P_D$  and  $P_{C_i}$ , respectively. Based on Eq. (4), the transmit power of the D2D transmitter can be derived by setting a SIR target  $\zeta_0$ :

$$P_{DTX} = \frac{\zeta_0 \sum_{i=1}^N |h_{DC_i}|^2 P_{C_i} d_{DC_i}^{-\alpha}}{|h_{D2D}|^2 d_{D2D}^{-\alpha}} \quad (5)$$

As mentioned beforehand, we are only using location information to determine the interference. The knowledge of instantaneous channel coefficients  $h_{D2D}$  and  $h_{DC_i}$  is not available to the power control mechanism. To simplify, we assume that the mean value of both coefficients ( $h_{D2D}$  and  $h_{DC_i}$ ) is equal to 1. Taking this assumption into account, and setting the upper limit on transmit power  $P_{max}$ , we obtain:

$$P_D = \min(P_{max}, \zeta_0 \cdot d_{D2D}^{\alpha} \sum_{i=1}^N P_{C_i} d_{DC_i}^{-\alpha}) \quad (6)$$

## 5. Simulations and Results

### 5.1. Scenarios and Parameters

The performance of the proposed LTSIPC power control solution is investigated via system level simulations performed with the use of a multi-cell OFDMA-based frequency reuse-1 network. The tool used in the simulations was co-created

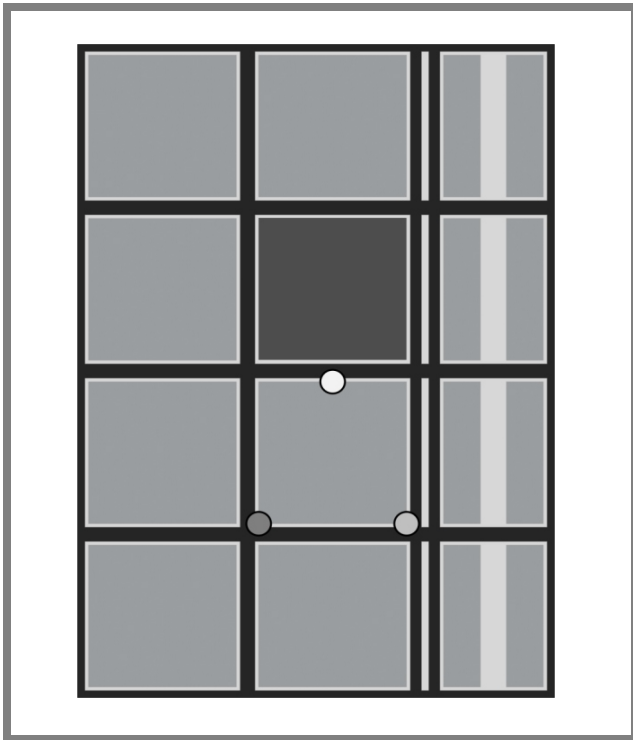


**Fig. 2.** In the Madrid grid model, locations of base station antennas are marked with circles.

by the author and it follows the guidelines established in the METIS project [14]. More details on the tool can be found in [15].

Channel models defined by METIS [16] are used for the cellular users. These models use, contrary to the most commonly used models, 3D map-based real-time methods for evaluating sight conditions between the individual nodes. For D2D users, a modified version of the D2D model defined by ITU-R [17] is used. The modification relies on a map-based approach for evaluating line of sight conditions, instead of the statistical approach that was defined in ITU-R documentation.

The solution under consideration is deployed with the help of the Madrid grid model [14] (Fig. 2) and consists of a macro base station operating in the frequency division duplex (FDD) mode and 3 sectors (each sector operates in the same frequency band). In the evaluation scenario, 400 users are uniformly placed outside the buildings, either on pavements (120 CUEs) or in cars (200 CUEs). Among these 400 users, 40 D2D pairs are deployed (80 DUEs) with the distance between each DUE and D2D pair ranging uniformly from 0 to 100 m. Additionally, 15 of the D2D pairs are pedestrian users, and the remaining 25 D2D pairs are deployed in cars (the distance between the users in cars is limited by the assumed dimensions of the vehicle). The mobility of the users, including cars, is also modelled in accordance with METIS guidelines. The process of allocating resources to cellular users is based on round robin mechanism. In the case of D2D users, resource blocks to be shared for communication are selected randomly. Cellular users are using the OLPC mechanism for determining the transmit power, whereas D2D users use the proposed LTSIPC mechanism with an SIR target of 20 dB. Benchmark scenarios in which D2D users used OLPC and in which no

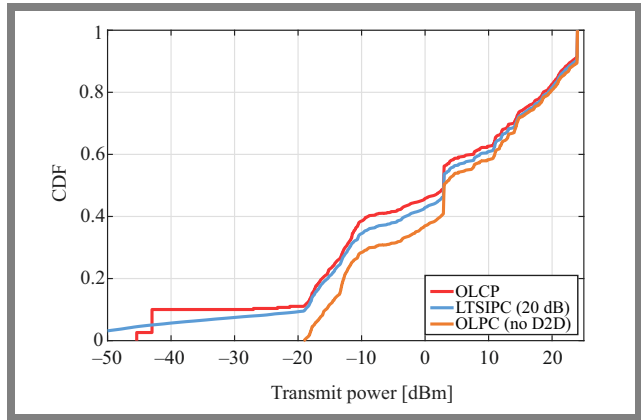


**Fig. 3.** CDFs of UE transmit power (for both CUEs and DUEs).

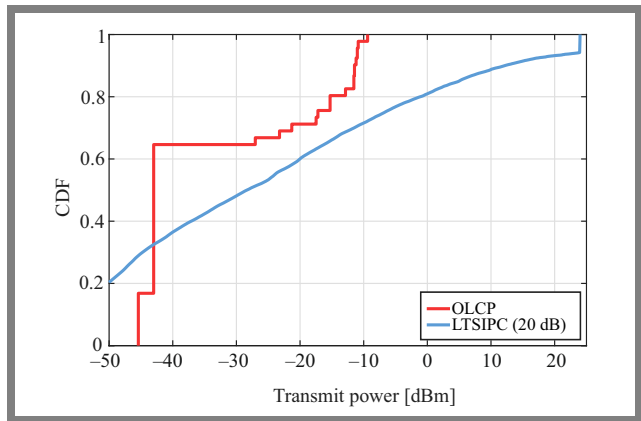


**Fig. 4.** CDFs of DUE transmit power.

D2D communication was enabled, were considered as well. Various system performance statistics are collected in the simulation tests, with the most important of them being:

- cumulative distribution function (CDF) of the transmit power of all users and active D2D users,
- spectral efficiency of the system (expressed in bits/s/Hz) for downlink (DL) and uplink (UL) transmissions and for active D2D users.

## 5.2. Results

The first set of results presented in Figs. 3 and 4 shows the CDFs of transmit powers for all UEs in the system and for D2D users, respectively. The introduction of D2D communication allows to reduce the transmit power of the UEs, regardless of the power control mechanism deployed. This is of course due to the assumption that the distance between devices communicating directly is lower than the distance between the individual devices and the base station.

Another observation is that the OLPC method achieves greater reduction in the transmit power than the proposed LTSIPC approach. This is due to the fact that LTSIPC aims at achieving the SIR target at the receiver, which in some cases may lead to an increased transmit power requirement in order to mitigate interference generated by cellular transmissions. Looking at the transmit power curves for D2D devices only (Fig. 4), we can observe that the LTSIPC performs slightly worse

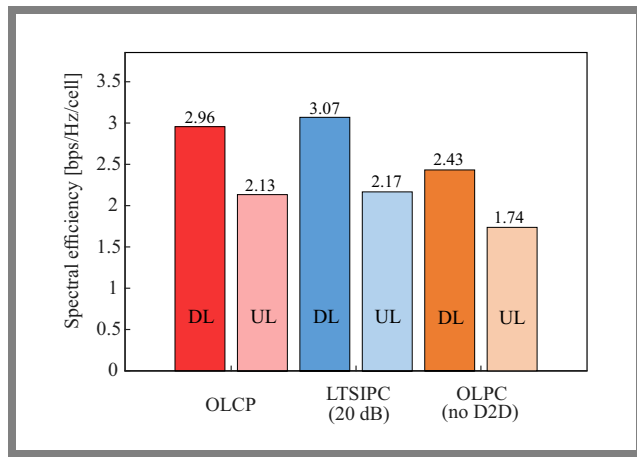


Fig. 5. Spectral efficiency of the system.

than OLPC in almost the entire transmit power range under consideration. However, the proposed LTSIPC approach offers better granularity of transmit power settings and allows for lower transmit power settings for D2D devices that are very close to each other, for instance in cars.

The second set of results presents the spectral efficiency of the simulated scenario. Figure 5 presents the spectral efficiency for downlink and uplink transmissions within the system. We can notice that the introduction of D2D may lead to an increase in spectral efficiency in both directions, regardless of the power control mechanism deployed. Taking into account the previous results related to the transmit power, this means that energy efficiency has increased as well. Better spectral efficiency has been achieved thanks to the spectrum sharing aspect of the D2D underlay. By allowing some devices to communicate directly and to share the spectrum with cellular users, we reduce the number of devices that have to be served by the base station and thus extend the network reuse factor above one.

Another observation that may be made based on spectral efficiency plots is that the proposed LTSIPC mechanism outperforms the OLPC approach, but as mentioned previously, at the cost of slightly increased transmit power demands. This can also be noticed from the spectral efficiency plot for the D2D devices only, as shown in Fig. 6.

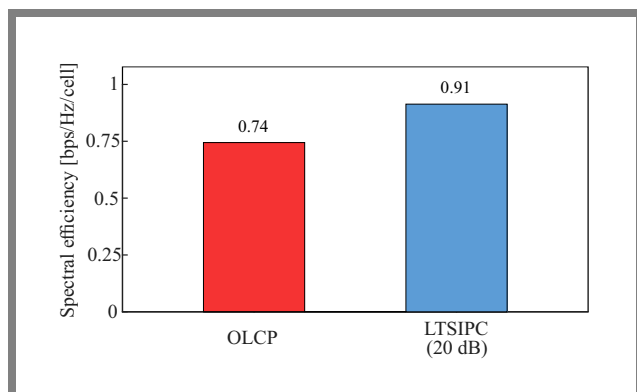


Fig. 6. D2D communication spectral efficiency.

## 6. Conclusion

The results show that the introduction of D2D communications, exploiting the proposed location-based power control method that shares UL resources with CUEs, may bring about benefits in terms of decreased transmit power levels, increased spectral efficiency and better energy efficiency. However, when comparing LTSIPC with OLPC, one may notice that the former offers slightly higher transmit power outputs almost in the entire transmit power range under consideration. However, the proposed LTSIPC approach allows for better granularity of transmit power settings and enables lower transmit power levels for D2D devices that are very close to each other, for instance in cars. Moreover, the LTSIPC approach achieves better spectral efficiency than the OLPC mechanism, both for the entire network and for D2D users only. It is also worth mentioning that the location-based approach does not require additional reports and channel measurements, needing to track the user's location only.

## References

- [1] P. Janis *et al.*, "Device-to-Device Communication Underlying Cellular Communications Systems", *International Journal of Communications, Network and System Sciences*, vol. 2, no. 3, 2009 (<https://doi.org/10.4236/ijcns.2009.23019>).
- [2] J. Seppala, T. Koskela, T. Chen, and S. Hakola, "Network controlled Device-to-Device (D2D) and cluster multicast concept for LTE and LTE-A networks", in: *2011 IEEE Wireless Communications and Networking Conference (WCNC)*, Cancun, Mexico, pp. 986–991, 2011 (<https://doi.org/10.1109/WCNC.2011.5779270>).
- [3] K. Doppler, M. Rinne, C. Wijting, C. Riberio, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks", *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, 2009 (<https://doi.org/10.1109/MCOM.2009.5350367>).
- [4] C.-H. Yu, K. Doppler, C. Riberio, and O. Tirkkonen, "On the performance of device-to-device underlay communication with simple power control", in: *VTC Spring '09-IEEE 69th Vehicular Technology Conference*, Barcelona, Spain, 2009 (<https://doi.org/10.1109/VETECS.2009.5073734>).
- [5] C.-H. Yu, K. Doppler, C. Riberio, and O. Tirkkonen, "Power optimization of device-to-device communication underlying cellular communication systems", in: *IEEE International Conference on Communications*, Dresden, Germany, 2009 (<https://doi.org/10.1109/ICC.2009.5199353>).
- [6] J. Gu, S.J. Bae, B.-G. Choi and M.Y. Chung, "Dynamic Power Control Mechanism for Interference Coordination of Device-to-Device Communication in Cellular Networks", in: *2011 Third International Conference on Ubiquitous and Future Networks (ICUFN'11)*, Dalian, China, pp. 71–75, 2011 (<https://doi.org/10.1109/ICUFN.2011.5949138>).
- [7] P. Janis *et al.*, "Interference-aware resource allocation for device-to-device radio underlying cellular networks", in: *VTC Spring '09-IEEE 69th Vehicular Technology Conference*, Barcelona, Spain, pp. 1–5, 2009 (<https://doi.org/10.1109/VETECS.2009.5073611>).
- [8] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlying LTE network", in: *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, Niagara Falls, Canada, 2011 (<https://doi.org/10.1109/WIMOB.2010.5645039>).
- [9] N. Reider and G. Fodor, "A distributed power control and mode selection algorithm for D2D communications", *EURASIP Journal on Wireless Communication and Networking*, art. no. 266, 2012 (<https://doi.org/10.1186/1687-1499-2012-266>).
- [10] P. Bao and G. Yu, "An interference management strategy for device-to-device underlying cellular networks with partial location infor-

- mation”, in: *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, Sydney, Australia, pp. 465–470, 2012 (<https://doi.org/10.1109/PIMRC.2012.6362830>).
- [11] H. Min, J. Lee, S. Park, and D. Hong, “Capacity enhancement using an interference limited area for device-to-device uplink underlying cellular networks”, *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 3995–4000, 2011 (<https://doi.org/10.1109/TWC.2011.100611.101684>).
- [12] X. Chen, L. Chen, M. Zeng, X. Zhang, and D. Yang, “Downlink resource allocation for Device-to-Device communication underlying cellular networks”, in: *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, Sydney, Australia, pp. 232–237, 2012 (<https://doi.org/10.1109/PIMRC.2012.6362746>).
- [13] H. Wang and X. Chu, “Distance-constrained resource-sharing criteria for device-to-device communications underlying cellular networks”, *Electronics Letters*, vol. 48, no. 9, pp. 528–530, 2012 (<https://doi.org/10.1049/el.2012.0451>).
- [14] P. Agyapong *et al.*, “Simulation guidelines” METIS Deliverable D6.1 [Online]. Available: ([https://www.metis2020.com/wp-content/uploads/deliverables/METIS\\_D6.1\\_v1.pdf](https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D6.1_v1.pdf)).
- [15] K. Bakowski, K. Wesolowski and M. Rodziewicz, “Simulation Tools for the Evaluation of Radio Interface Technologies for IMT-Advanced and Beyond”, in: *Simulation Technologies in Networking and Communications: Selecting the Best Tool for the Test*, CRC Press, pp. 365–391, 2014 (<https://doi.org/10.1201/b17650-17>).
- [16] V. Nurmela *et al.*, “Initial channel models based on measurements”, METIS deliverable D1.2, 2013. [Online]. Available: ([https://www.metis2020.com/wp-content/uploads/deliverables/METIS\\_D1.2\\_v1.pdf](https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.2_v1.pdf)).
- [17] Recommendation ITU P.1411-7, “Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz”, 2013. [Online]. Available: <https://www.itu.int/rec/R-REC-P.1411-7-201309-S/en>

# Marcin Rodziewicz, Ph.D.

Institute of Radiocommunications, Faculty of Computing and Telecommunications

 <https://orcid.org/0000-0002-0487-1204>

E-mail: marcin.rodziewicz@put.poznan.pl

Poznan University of Technology, Poznan, Poland

<https://cat.put.poznan.pl>

# A Cross-layer Method for Scheduling and Routing Real-time Traffic Flow in Industrial IoT

Nikta Azimian, Reza Mohammadi, and Mohammad Nassiri

*Bu-Ali Sina University, Hamedan, Iran*

<https://doi.org/10.26636/jtit.2023.3.1362>

**Abstract** — In the Industrial Internet of Things, a wide variety of sensors are distributed all over the environment to monitor data collection, thereby allowing industrial processes to be monitored more efficiently. One of the fundamental goals of IIoT is to provide the highest level of reliability while simultaneously increasing network lifetime, reducing power consumption, and preventing delays. 6TiSCH is a popular communication standard relied upon in IIoT. The aim of the present study is to propose an inter-layer method that simultaneously considers network scheduling and routing processes based on TSCH and RPL approaches in multi-sink environments. The proposed method is intended to address the limitations of IIoT and meet the requirements of field-specific applications.

**Keywords** — *Internet of Things, Industrial IoT, multi sink, TSCH standard*

## 1. Introduction

The growing popularity of smart gadgets, industrial automation systems, and smart buildings has resulted in a tremendous growth of wireless sensor networks (WSNs) and the Internet of Things (IoT). The underlying aim of the Internet of Things is to connect all devices to the Internet. In the Internet of Things, a wide variety of sensors are distributed all over the environment to monitor and report various conditions and states, thereby allowing for smart decisions to be made based on the collected information [1]–[3]. In recent years, the Industrial Internet of Things (IIoT) has been proposed as a specialized IoT variety that is compatible with industrial environments and manufacturing sites. This idea is aimed at increasing flexibility and productivity while simultaneously decreasing production costs in industrial environments.

Although different devices, including entire systems, sensors, etc. were only connected to form clusters in the past, they are now connected to a single macro-structure that is accessed by the user. With the help of IIoT, industries may be connected to a single integrated network [2]–[5]. IIoT consists of a set of nodes, including a processor, a battery (power supply), a radio transceiver for communication and data exchange, and a number of sensors (monitoring temperature, pressure, humidity, etc.). These nodes exchange their data with one or more receiver stations that may be either of the stationary or mobile variety. Nodes working in IoT envi-

ronments face various limitations related to their processing power, memory capacity, and energy consumptions [1]–[4]. In IIoT applications, the most important criteria include also reliability and delay (lag). In order to ensure optimum power consumption in IIoT settings, it is necessary to develop suitable sleep/wake mechanisms and to use the Carrier Sense Multiple Access (CSMA) approach in the Medium Access Control (MAC) sublayer [1], [2]. To achieve this goal, such standards as IEEE 802.15.4 [6] and IEEE 802.15.4e [7] define low power MAC and physical layers capable of meeting IIoT requirements. Focusing on the centralized coordination of nodes, the IEEE 802.15.4 standard controls the operation of CSMA among the nodes. It coordinates the transmissions by sending small packets called beacons. TSCH is one of the varieties of IEEE 802.15.4e. It is a combination of time synchronization, frequency division, and channel hopping that reduces the chance of a collision between different transmitters. The TSCH mechanism enables each node to synchronize its data transmission on a single channel and in a specific time slot. However, the standard does not determine how the channel/slot must be allocated.

In WSNs, addresses are allocated using IPv6. To activate IPv6 in TSCH, the IETF 6TiSCH work group runs a protocol stack based on IIoT-specific standards, such as RPL, 6LoWPAN, and CoAP routing protocols, using 6LoWPAN for address mapping. In addition to designing the MAC and physical layers, the current structure of the routing layer was modified in IEEE 802.15.4e by means of a routing protocol called RPL that is suitable for low-power networks. In order to use the multi-parent feature of this protocol, its DODAG structure can be utilized for convergent traffic in IIoT solutions [1]. As mentioned above, numerous scheduling methods

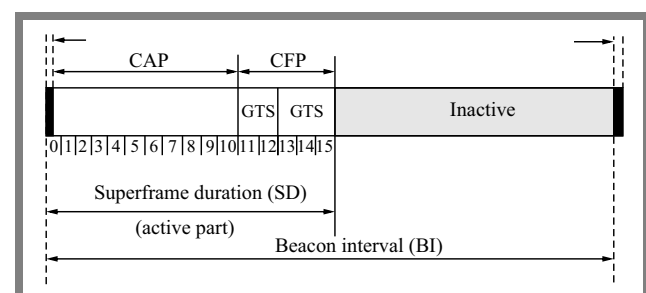


**Fig. 1.** The superframe structure in IEEE 802.15.4.

have so far been proposed for 6TiSCH networks, but few studies have addressed the problem of time scheduling in multi-sink networks. Therefore, the aim of this study is to meet the needs of industrial applications and remove some of the existing limitations by implementing several sinks in the network and utilizing a scheduling algorithm. In what follows, we shall investigate the possibility of creating multiple sinks. The reason for using multiple sinks in IIoT is to reduce power consumption and, ultimately, increase network lifetime. Using a 6TiSCH simulator, we evaluate the performance of the implemented method and compare it to other proposals.

Section 2 presents a review of the communication standards used in industrial IoT. In Section 3, we define the basic concepts and, in Section 4, we briefly review the existing literature. Section 5 describes the simulations performed and evaluates the results. Finally, Section 6 sums up the results and draws the conclusions.

## 2. Communication Standards in IIoT

An implementation of an IIoT environment requires a variety of standards and protocols to be complied with. This section examines several standards for the data link layer and the network layer in IIoT applications.

A wireless sensor network (WSN) is usually used in IIoT for purposes that contribute to mitigating overload conditions caused by intra-network processing, data integration, and decision-making processes. The fundamental challenges in these networks are to decrease delay, reduce power consumption, ensure fault tolerance, and increase the lifetime of nodes.

IEEE 802.15.4 implements the physical layer and the media access control layer for low-power WSNs. It transmits packets via the data link layer, using a shared channel, in an integrated manner relying on the synchronization of nodes. The physical layer of this standard uses one of three bands: 868 to 868.6 MHz, 902 to 928 MHz, and 2400 to 2483 MHz. The majority of the implementations so far have relied upon the 2.4 GHz band with 16 separate channels and a transmission rate of 250 kbps. The data link layer is designed to reduce the probability of data collision and data transfer using this layer is either of the beaconless or beacon-based variety. In the beaconless mode, the classic CSMA-CA approach is implemented, in which the sender listens to the channel. If it finds the channel empty, it sends the data. Otherwise, it waits a random time interval before sending the data. The network coordinator must stay awake to receive the frame from the nodes, hence it

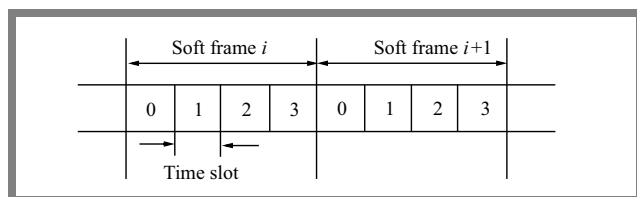


**Fig. 2.** A slot frame with four time slots in the TSCH mode of IEEE 802.15.4e.

cannot save energy. This increases the overload and greatly affects the end-to-end delay. Since the nodes keep consuming energy even when idle, this state is not appropriate for multi-hop modes which suffer from energy limitations [1]. In the beacon-based mode provided for in IEEE 802.15.4, both the sender and the receiver use a superframe structure (Fig. 1) for exchanging data. In this mode, a coordinator periodically sends beacons to determine the boundaries of its superframe. A superframe has an active and an inactive period. By determining the duration of each of the periods, the work cycle can be set dynamically. An active period begins with a beacon frame which is sent by the coordinator in regular time intervals. A superframe is composed of CAP and CFP. CAP is the part in which competition takes place and packets are transmitted by means of the CSMA/CA mechanism. In CFP, with no competition, data packets are sent and GTS cells are allocated to data transmission. In the beacon-based mode, nodes are awakened by the coordinator just before sending the beacon, in order to save energy, and they fall asleep again after receiving data. This allows more energy to be saved, because the nodes are not required to stay awake permanently. The IEEE 802.15.4e standard, based on IEEE 802.15.4 used in low-power short-range wireless networks, has become popular for communicating in industrial environments [7]. The IEEE 802.15.4 standard has certain limitations, such as high power consumption of middle nodes that must always remain operable in order to transfer information to other nodes, or using a single channel for information transfers, which causes interference. IEEE 802.15.4 uses some of the existing methods applied in WSNs, including time slots, specific and shared cells, multiple channels, as well as frequency hopping. It relies upon different modes, including LLDN, BLINK, DSME, AMCA, and TSCH. As the TSCH mechanism is a solution that reconciles all other modes and meets as many needs as possible, it is the most complex and appealing solution and is considered to be the core mechanism of IEEE 802.15.4e [1], [6]. TSCH is a combination of time synchronization, frequency division, and channel hopping. Due to its optimal use of time and frequency, this technique is a suitable choice for networks with multi-hop communications.

In TSCH, time is divided into specific spans called slot frames or superframes. Each slot frame is divided into a certain number of equal time slots and repeats periodically. The duration of a time slot is equal to the duration in which the sender transmits a packet to its single-hop neighbor and receives the verification of the packet's second layer from the neighbor. All nodes within the network share the same time structure. Its ability to support multiple channels by channel hopping is a characteristic feature of TSCH. TSCH has 16 different channels, each with the bandwidth of 5 MHz. In this mechanism, a direct communication link is created between sender and receiver nodes in a given time slot in a specific channel. Each link that is selected for communication between two nodes is specified by two indexes, i.e., time slot number and channel number. Figure 2 shows a slot frame with four time slots. In each time slot, a node can send the largest amount of data possible [1]. The RPL standard [8] has recently emerged as

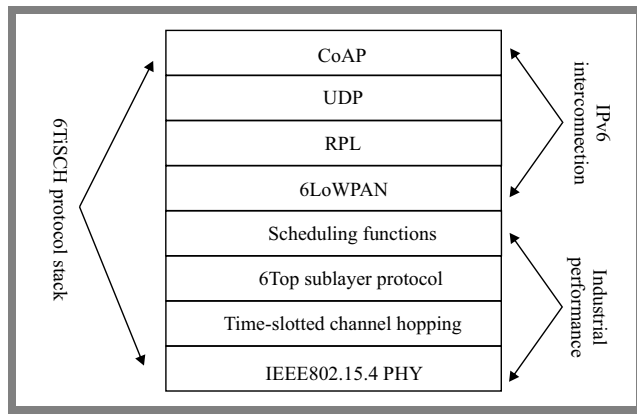


Fig. 3. 6TiSCH protocol stack.

a solution for routing protocol in WSNs. It defines a structure for efficient and energy-saving transmission of packets. The RPL protocol creates a destination-oriented directed acyclic graph (DODAG). In a DODAG, the edges do not form a loop and are laid out so that they face towards the root node. The sink node is the coordinator that acts as the graph root and is the destination for all packets. To prevent cycles in DODAG, a target function is defined to determine the rank of the nodes. A node's rank is a numerical value that shows the place of the node within the DODAG. The closer a given node is to the sink, the lower its rank. In fact, the rank indicates the relative position of each node to the root. The nodes regularly inform each other about their ranks via periodic messages of DODAG information objects (DIO) [1].

### 3. 6TiSCH Network

6TiSCH brings IPv6 to low-power industrial wireless networks and establishes communication between time-slotted channel hopping (TSCH) networks and 6LoWPAN networks [9]. 6TiSCH combines industrial functions with IPv6 and provides a mechanism by which the link layer topology is adapted to the routing topology [10]. Figure 3 shows the protocol stack of 6TiSCH, using different protocols and standards in each layer [11], i.e., IEEE 802.15.4 in the physical layer, TSCH in the media access control (MAC) layer, SF as a scheduling algorithm and the 6Top protocol. In the network layer, the RPL protocol is used for routing and constructing the topology, and 6LoWPAN is used for IPv6 address mapping for the nodes and for compressing IPv6 packet headers. The UDP protocol is used in the transmission layer, when error detection and correction are not necessary. It is also used in running programs to avoid excessive process overheads. UDP is used for time sensitive applications. The CoAP protocol [12] is also used in the application layer as one of the most widely used protocols in the IoT. This protocol is standardized by targeting limited nodes in low-power networks. This protocol gives devices the ability to connect to a wider range of networks.

## 4. Centralized Scheduling Algorithm

In this type of scheduling, it is the central node that assumes the role of the host and is responsible for the formation and maintenance of network scheduling. All the network nodes periodically inform the host about the list of their neighbors and the data they generate. Using the gathered information, the host creates a network bonding graph and allocates an appropriate number of time slots to the bonding links based on the information concerning the amount of the data produced by each node. Once the scheduling has been completed, the admin node provides each node with its schedule. In real-world environments, networks have a main node called as a sink. In centralized scheduling, the host node is usually in charge of this task. As this node has knowledge about the entire network, the centralized solution offers an optimal schedule. In scenarios in which the topology is changing and there is a need for network reconstruction, the distributed solution is a more suitable choice [1].

## 5. Literature Review

Palattella *et al.* [13] proposed a basic method called TASA which is a centralized scheduling algorithm based on network topology and traffic load. Unfortunately, the TASA algorithm was less efficient than other algorithms. Jin *et al.* [14] proposed the adaptive multi-hop scheduling (AMUS) approach. This method executes three functions: scheduling, experimental cell allocation, and End-of-Q informing mechanism. Choi *et al.* in [15] proposed an algorithm for centralized scheduling, called CLS. Using the minimum number of control messages, CLS creates an efficient schedule multi-hop schedule, as it allocates and moves the slots without re-scheduling the entire application. Paper [16] used a routing structure based on multiple paths in which every router has several further hops ahead of it to increase fault tolerance. The authors also proposed a scheduling algorithm in which several senders that are connected to a single receiver are scheduled within a cell. This method is useful in multipath applications. Sebastian *et al.* [17] developed a method for implementing multiple sinks based on DODAG. In this method, the topology contains multiple DODAGs which have separate sink nodes. There is no connection between these nodes, and such an approach reduces delay. This method can be suitable for considering the multi-sink approach in a network. Only few studies have been performed recently focusing on scheduling in networks with the possibility of creating multiple sinks. Some of the main reasons behind using multiple sinks in a network include the following: increasing network lifetime, reducing power consumption, and improving reliability. Therefore, the aim of this study is to implement multiple sinks in 6TiSCH networks as separate DODAGs and utilize a centralized scheduling algorithm, adopted from [15], to improve the use of multiple sinks in a 6TiSCH simulator [18], as well as to enhance power consumption, network lifetime, packet delivery rate, and delay.

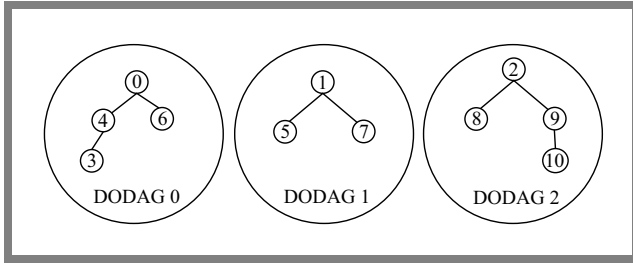


Fig. 4. Topology with three sinks.

## 6. The Proposed Multi-sink CLS

In the proposed method, several sinks are implemented in a 6TiSCH simulator [18], [19]. We increase the number of sinks to three (Fig. 4) and we select nodes 0, 1, and 2 as sinks. Given this, formation about network topology changes and the network initially contains three sinks that are connected to each other. Through these connections, the nodes can send packets to other sinks. This will increase the delay. According to [17], we use the idea of Multi-DODAGs and apply a limitation by removing the connection between the sinks. Therefore, the topology now contains three DODAGs with three separate sinks, and delay is improved.

We use the CLS algorithm [15] in a multi-sink mode. The proposed method is named MCLS. This algorithm has two main features. It is a multi-hop scheduling algorithm that considers the order of scheduling the links, simultaneously taking into account interference during the scheduling process. Thank to such an approach, the receiver node cannot receive data simultaneously from several sender nodes, because the signals may interfere with each other. Given these two features, the process of scheduling is as follows.

First, a scheduling matrix should be formed that is available to all nodes. The nodes can analyze the information in the matrix before scheduling and then cell allocation is performed. Cell allocation refers to the attempt to find the appropriate time and channel for the intended node. In TSCH, the number of the slot number is called slot offset and the channel number is called channel offset. Slot offset and channel offset are also determined by the number of matrix rows and columns, respectively. The strategy of scheduling and allocating suitable times and channels is based on interference level. Scheduling is performed for a given node using the distance of each node from its parent, i.e., hop distance or node rank. Thus, for nodes which are at a distance of one hop from the sink node, the slot offset from the first time slot is set, equaling 1 [15]. The first time slot for scheduling is set as 1, because time slot 0 is used to send initial packets, such as beacon and RPLDIO. The channel offset value is calculated as [15]:

$$\text{ChannelOffset} = \text{rank}(\text{hopdistance}) - 1, \quad (1)$$

where rank is the node's distance from its parent.

When the nodes are at a distance of several hops from the sink node, the slot offset value can be calculated according to [15]:

$$\text{SlotOffset} = \text{SlotOffset}(\text{parent}) + 1. \quad (2)$$

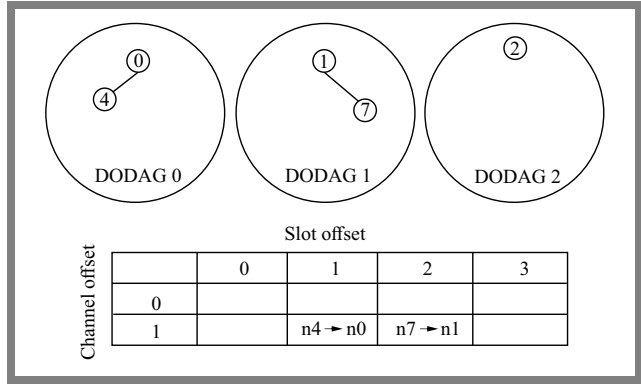


Fig. 5. Scheduling for a one-hop scenario.

In Eq. (2), the value of slot offset (parent) is calculated with reference to the slot offset of the parent which has connected to the network beforehand and for which cell allocation has already been completed. The value of channel offset can be calculated by Eq. (1). Scheduling is performed based on a dynamic strategy. This means that when the simulator is run, scheduling and cell allocation are performed for a node at the moment when it connects to the network. Figures 5–6 show the scheduling and cell allocation processes for one-hop and multi-hop scenarios. Figure 7 shows the pseudocode of the MCLS allocation scheduling algorithm for cell allocation. This algorithm is executed for each node that joins DODAG.

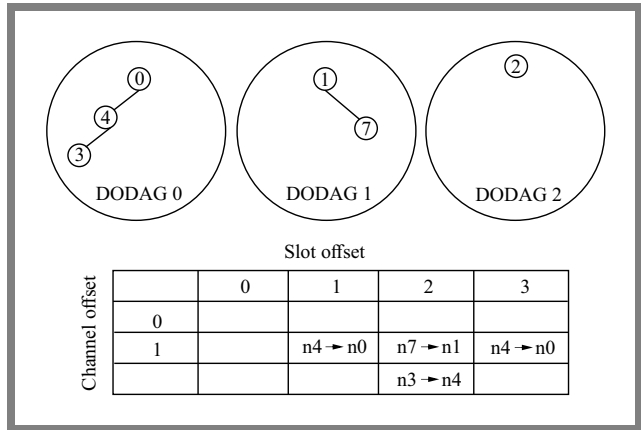


Fig. 6. Scheduling for a multi-hop scenario.

## 7. Simulation and Evaluation

The proposed MCLS method was simulated using 6TiSCH [18], [19]. It is a Python based open-source tool which has been developed by the IETF 6TiSCH work group. This simulator measures different criteria, such as end-to-end delay, end-to-end reliability, power consumption, and network lifetime. Using this simulator, we compare the results of MCLS with the following methods from the literature:

- the CLS [15] centralized scheduling algorithm that assumes a single sink node in the network,
- Multi-DODAG from [17] which implemented multiple sinks in separate DODAGs in a network, lacking a scheduling algorithm for the nodes.

```

Algorithm: CELL Allocation Processing
1: Input : u: the local node
   Parent(u) : the parent of u
   DODAGrank(u) : DODAGrank of u
   matrixSchedual : matrix of scheduling
   timeSlot = 0 : the initialize of timeSlot
   channel = 0 : the initialize of channel
   SlotOffset(u) = 0 : the initialize of SlotOffset
   ChannelOffset(u) = 0 : the initialize of ChannelOffset
2: Output : SlotOffset, ChannelOffset
3: Function :
   for u that join the network
     DODAGrank(u) = DODAGrank(u) + 1
     if DODAGrank(u) == 2:
       timeSlot = 1
       Channel = DODAGrank(u) - 1
     if DODAGrank(u) >= 3:
       timeSlot = SlotOffset(parent(u)) - 1
       Channel = DODAGrank(u) - 1
     if matrixSchedual[timeSlot][Channel] == 0
       SlotOffset(u) = timeSlot
       ChannelOffset(u) = channel

```

**Fig. 7.** MCLS scheduling algorithm.**Tab. 1.** Simulation parameters for the first scenario.

| Parameter                  | Value                  |
|----------------------------|------------------------|
| Number of nodes            | 10, 12, 14, 16, 18, 20 |
| Duration of simulation     | 260 s                  |
| Routing algorithm          | RPL                    |
| Packet size                | 90 bytes               |
| Packet transmission period | 60 s                   |

**Tab. 2.** Simulation parameters for the second scenario.

| Parameter                  | Value                |
|----------------------------|----------------------|
| Number of nodes            | 20                   |
| Duration of simulation     | 330 s                |
| Routing algorithm          | RPL                  |
| Packet size                | 90 bytes             |
| Packet transmission period | 20, 30, 40, 50, 60 s |

The criteria for evaluation in each scenario include network lifetime, reliability, packet transmission rate, and delay. In order to make more accurate interpretations of the results, we will keep to a confidence level of 99%. Moreover, the location of each node is randomly arranged around the sink and in a  $2 \times 2$  square space.

In the first scenario, the performance of MCLS is examined in terms of an increase in the number of nodes as well as scalability. The simulation parameters of the first scenario are listed in Tab. 1. Figure 8a shows that the average lifetime of the network is greater in the proposed method. The reason is that reduced power consumption and traffic loads enhance network lifetime. Figure 8b shows the average end-to-end delay. As one may notice, in some cases the delay in MCLS is lower than in CLS, especially where the num-

**Tab. 3.** Simulation parameters for the third scenario.

| Parameter                  | Value           |
|----------------------------|-----------------|
| Number of nodes            | 20              |
| Duration of simulation     | 254 s           |
| Routing algorithm          | RPL             |
| Packet size                | 90 bytes        |
| Packet transmission period | 20, 30, 40      |
| Number of sink nodes       | 1, 3, 5, 7, 9 s |

**Tab. 4.** Simulation parameters for the fourth scenario.

| Parameter                  | Value         |
|----------------------------|---------------|
| Number of nodes            | 20            |
| Duration of simulation     | 254 s         |
| Routing algorithm          | RPL           |
| Packet size                | 90 bytes      |
| Packet transmission period | 1, 5, 7, 10 s |

ber of nodes is greater. The reason is that, in multiple sinks, the distance between the nodes and the sinks is reduced and packets arrive at the sinks with a shorter delay. It may be noticed that MCLS has significantly decreased the delay of the Multi-DODAG method. As shown in Fig. 8c, the packet delivery rate (expressed in percentages) in MCLS and CLS is better than in Multi-DODAG, because, as mentioned earlier, the Multi-DODAG method does not perform scheduling for nodes. Moreover, as MCLS uses multiple sinks and reduces the network's traffic load, it decreases the probability of packet loss and increases the packet delivery rate.

In the second simulation scenario, the performance of MCLS is examined in terms of the number of packets sent and traffic load. The simulation parameters are listed in Tab. 2. As shown in Fig. 9a, MCLS performs better than the two other methods at higher traffic rates. If the nodes send a packet every 20 s, the nodes' lifetime will still remain longer in MCLS. This is explained by the presence of the sink nodes in the network, as these can manage large traffic loads. Figure 9b shows the changes in the average end-to-end delay for different packet transmission durations. The average delay in MCLS has decreased in comparison with the two remaining methods, due to the smaller number of hops required to arrive at the destination. Figure 9c shows the changes in the average packet delivery rate for different packet transmission durations. In MCLS, due to the existence of multiple sinks in the network, the network's traffic load for different packet delivery durations is distributed among the sink nodes. Therefore, the packet delivery rate of the sink nodes is greater than in the two other methods, even when the packets are sent every 20 s and the network is very crowded.

In the third scenario, the performance of MCLS is examined based on the number of sink nodes (Tab. 3).

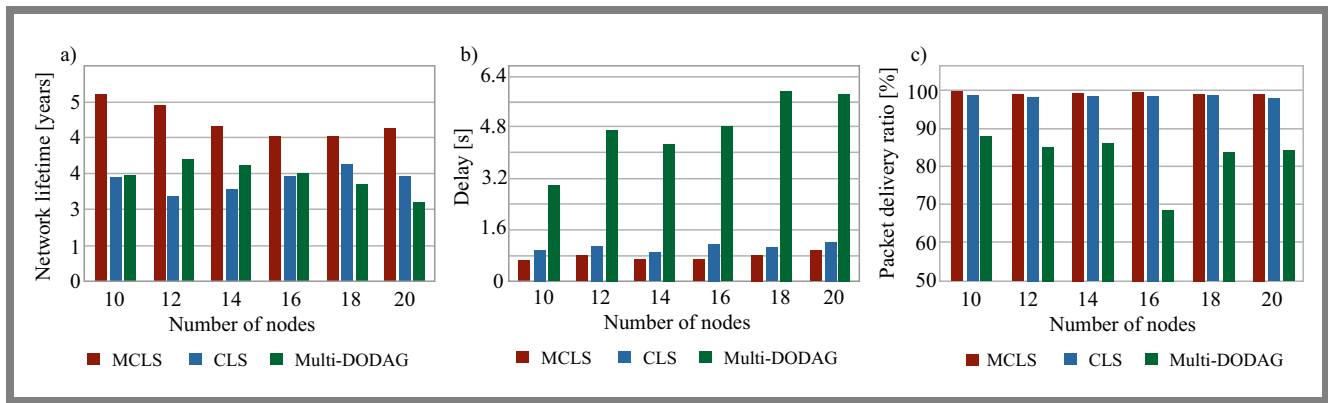


**Fig. 8.** Simulation results for scenario 1: a) average network lifetime, b) average delay, and c) average packet delivery rate.

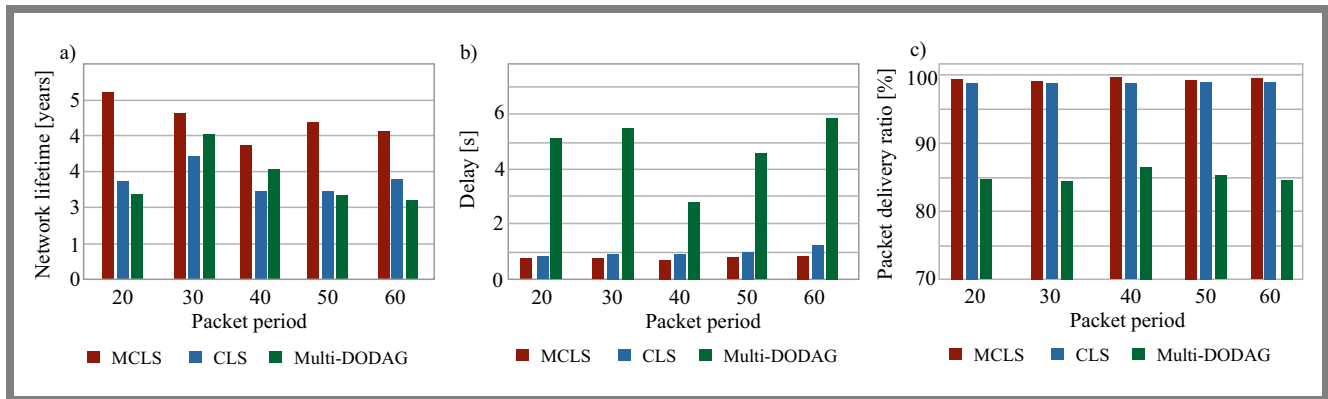


**Fig. 9.** Comparison of: a) average lifetime, b) average end-to-end delay, and c) average packet delivery ratio.

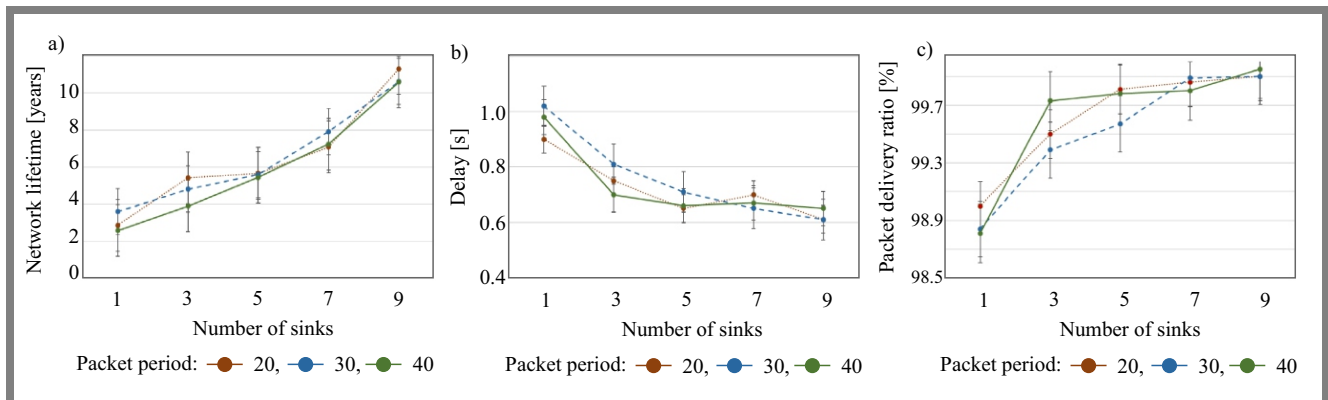


**Fig. 10.** Comparison of average: a) lifetime of networks, b) delay, and c) packet delivery ratio with different number of sink nodes.

As one may see in Fig. 10, as the number of sink nodes in the network increases and for different packet transmission durations, such parameters as network lifetime, delay, and packet delivery rate improve as well. The reason is that the nodes consume less power for packet transmission and, therefore, have a longer lifetime.

On the other hand, packets are delivered with smaller delays due to the decrease in the number of hops from the source to the destination node. Also, due to the existence of multiple sinks, the network's traffic load for different packet delivery durations is distributed among the individual sink nodes. This causes an increase in packet delivery rate in the sink nodes. As can be seen in the comparison with Multi-DODAG and

CLS, MCLS has improved network lifetime and packet delivery rate, simultaneously decreasing delay.

In the fourth scenario, the network's congestion control capabilities are examined. For this purpose, by adding the number of packets sent per second, network lifetime parameters, packet reception and delay are analyzed (Tab. 4). As shown in Fig. 11, along with the growing network congestion, the proposed method works better and the parameters of lifetime, latency and packet delivery are improved.

In the proposed method, due to the presence of several sink nodes, when the number of packets sent from the nodes increases, the network can manage the congestion more effectively. Therefore, the packet loss rate is reduced and the

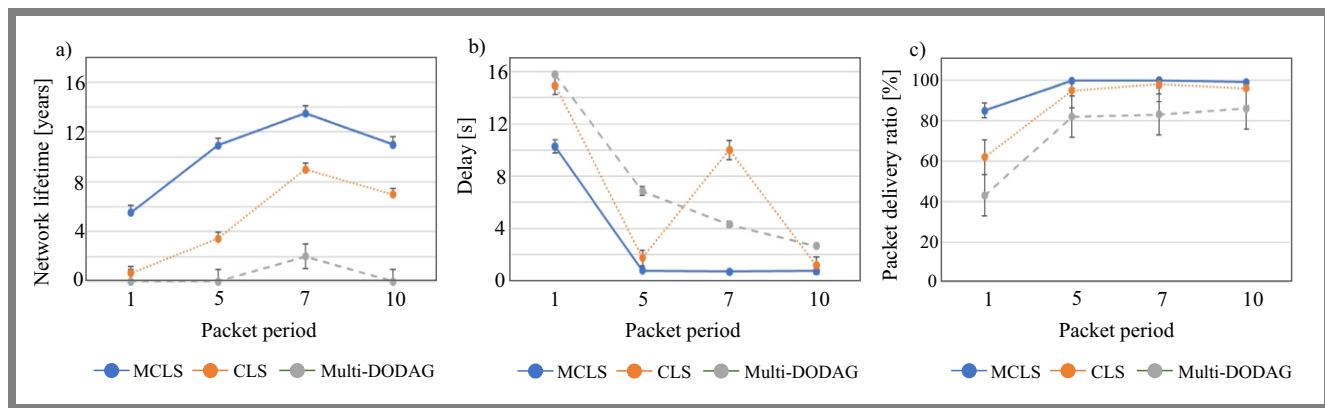


Fig. 11. Comparison of average: a) delay, b) packet delivery ratio, and c) network lifetime for different packet period.

nodes do not need to resend them. They consume less energy and, finally, the overall lifetime of the network increases compared to the two other methods.

On the other hand, due to the presence of several sink nodes and the reduction of the number of steps between the source and the destination, the packet sending delay is reduced.

## 8. Conclusion

In this paper, we implemented multiple sinks in the 6TiSCH simulator and use a centralized scheduling algorithm. To implement multiple sinks, we used the Multi-DODAG method which lacks a scheduling component. For the purpose of scheduling the nodes, we used the CLS method, i.e. a centralized scheduling algorithm that contains one sink node only. The proposed method is called MCLS and uses both multiple sinks, just like Multi-DODAG, and a centralized scheduling method, just like the CLS method. According to the simulation results, MCLS has overcome the limitation of using one sink in CLS and has outperformed Multi-DODAG by adopting the centralized scheduling algorithm of CLS. As discussed in the evaluation section, the criteria of network lifetime, end-to-end delay, and packet delivery rate in different scenarios were improved for each of the methods.

## References

- [1] S. Scanzio *et al.*, "Wireless Sensor Networks and TSCH: A Compromise Between Reliability, Power Consumption, and Latency", *IEEE Access*, vol. 8, pp. 167042–167058, 2020 (<https://doi.org/10.1109/ACCESS.2020.3022434>).
- [2] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The Industrial Internet of Things (IIoT): An Analysis Framework", *Computers in Industry*, vol. 101, pp. 1–12, 2018 (<https://doi.org/10.1016/j.compind.2018.04.015>).
- [3] K. Gulati *et al.*, "A Review Paper on Wireless Sensor Network Techniques in Internet of Things (IoT)", *Materials Today: Proceedings*, vol. 51, no. 1, pp. 161–165, 2022 (<https://doi.org/10.1016/j.matpr.2021.05.067>).
- [4] C.-Y. Chong and S. P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges", *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003 (<https://doi.org/10.1109/JPROC.2003.814918>).
- [5] A. Sari, A. Lekidis, and I. Butun, "Industrial Networks and IIoT: Now and Future Trends", in *Industrial IoT, Challenges, Design Principles, Applications, and Security*, pp. 3–55, Springer, 2020 ([https://doi.org/10.1007/978-3-030-42500-5\\_1](https://doi.org/10.1007/978-3-030-42500-5_1)).
- [6] A. Gezer and S. Okdem, "Improving IEEE 802.15.4 channel access performance for IoT and WSN devices", *Computers & Electrical Engineering*, vol. 87, art. no. 106745, 2020 (<https://doi.org/10.1016/j.compeleceng.2020.106745>).
- [7] D. De Guglielmo, S. Brienza, and G. Anastasi, "IEEE 802.15.4e: A survey", *Computer Communications*, vol. 88, pp. 1–24, 2016 (<https://doi.org/10.1016/j.comcom.2016.05.004>).
- [8] H. Kharrufa, H.A.A. Al-Kashoash, and A.H. Kemp, "RPL-based Routing Protocols in IoT Applications: a Review", *IEEE Sensors Journal*, vol. 19, no. 15, pp. 5952–5967, 2019 (<https://doi.org/10.1109/JSEN.2019.2910881>).
- [9] M.R. Palattella *et al.*, "Standardized Protocol Stack for the Internet of (Important) Things", *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1389–1406, 2012 (<https://doi.org/10.1109/SURV.2012.111412.00158>).
- [10] X. Vilajosana *et al.*, "IETF 6TiSCH: A Tutorial", *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 595–615, 2019 (<https://doi.org/10.1109/COMST.2019.2939407>).
- [11] T. Chang *et al.*, "6TiSCH on SCμM: Running a Synchronized Protocol Stack without Crystals", *Sensors*, vol. 20, no. 7, art. no. 1912, 2020 (<https://doi.org/10.3390/s20071912>).
- [12] A. Bhattacharjya, X. Zhong, J. Wang, and X. Li, "CoAP-Application Layer Connection-less Lightweight Protocol for the Internet of Things (IoT) and CoAP-IPSEC Security with DTLS Supporting CoAP", in: *M. Farsi, A. Daneshkhah, A. Hosseinian-Far, H. Jahankhani (Eds.) Digital Twin Technologies and Smart Cities. Internet of Things*, pp. 151–175, Springer Cham, 2020 ([https://doi.org/10.1007/978-3-030-18732-3\\_9](https://doi.org/10.1007/978-3-030-18732-3_9)).
- [13] M.R. Palattella, N. Accettura, M. Dohler, L.A. Grieco, and G. Boggia, "Traffic Aware Scheduling Algorithm for Reliable Low-power Multi-hop IEEE 802.15.4e Networks", in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 327–332, 2012 (<https://doi.org/10.1109/PIMRC.2012.6362805>).
- [14] Y. Jin, P. Kulkarni, J. Wilcox, and M. Sooriyabandara, "A Centralized Scheduling Algorithm for IEEE 802.15.4e TSCH Based Industrial Low Power Wireless Networks", in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6, 2016, (<https://doi.org/10.1109/WCNC.2016.7565002>).
- [15] K.-H. Choi and S.-H. Chung, "A New Centralized Link Scheduling for 6TiSCH Wireless Industrial Networks", in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, Springer, 2016, pp. 360–371 ([https://doi.org/10.1007/978-3-319-46301-8\\_30](https://doi.org/10.1007/978-3-319-46301-8_30)).
- [16] E.M. Ahrar, M. Nassiri, and F. Theoleyre, "Multipath Aware Scheduling for High Reliability and Fault Tolerance in Low Power Industrial Networks", *J. of Network and Comp. Applications*, vol. 142, pp. 25–36, 2019 (<https://doi.org/10.1016/j.jnca.2019.05.013>).

- [17] A. Sebastian and S. Sivagurunathan, "Multi DODAGs in RPL for Reliable Smart City IoT", *Journal of Cyber Security and Mobility*, vol. 7, no. 1, pp. 69–86, 2018 (<https://doi.org/10.13052/jcs-m2245-1439.716>).
- [18] E. Municio *et al.*, "Simulating 6TiSCH Networks", *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 3, art. no. e3494, 2019 (<https://doi.org/10.1002/ett.3494>).
- [19] T. Chang, M. Vučinić, X.V. Guillén, D. Dujovne, and T. Watteyne, "6TiSCH Minimal Scheduling Function: Performance Evaluation", *Internet Technology Letters*, vol. 3, no. 4, art. no. e170, 2020 (<https://doi.org/10.1002/itl2.170>).

---

**Nikta Azimian, M.Sc. Student**  
Department of Computer Engineering  
E-mail: nikta76azimian@gmail.com

Bu-Ali Sina University, Hamedan, Iran  
<https://basu.ac.ir>

**Reza Mohammadi, Ph.D., Assistant Professor**  
Department of Computer Engineering  
 <https://orcid.org/0000-0002-2139-4518>  
E-mail: r.mohammadi@basu.ac.ir  
Bu-Ali Sina University, Hamedan, Iran  
<https://basu.ac.ir>

**Mohammad Nassiri, Ph.D., Associate Professor**  
Department of Computer Engineering  
 <https://orcid.org/0000-0002-0703-5949>  
E-mail: m.nassiri@basu.ac.ir  
Bu-Ali Sina University, Hamedan, Iran  
<https://basu.ac.ir>

# Characterization of a Split Circle Element for Microstrip Reflectarrays

Khalil H. Sayidmarie<sup>1</sup> and Likaa S. Yahya<sup>2</sup>

<sup>1</sup>Ninevah University, Mosul, Iraq,

<sup>2</sup>Northern Technical University, Mosul, Iraq

<https://doi.org/10.26636/jtit.2023.3.1363>

**Abstract** — A split circular element is proposed as a unit cell for reflectarray antennas. The unit cell is derived from a circle divided into four equal sectors. The radius of two oppositely located sectors is then scaled by a certain factor to form the proposed shape. The CST Microwave Studio Suite software simulator was used to investigate the performance of the proposed unit cell, which was evaluated using Floquet port excitation. The designed element's reflection phase range was compared to that of a conventional circular patch. Four scenarios of varied substrate characteristics are investigated for the antenna to establish the best performance parameters. The simulations showed that a basic substrate with a thickness of 0.16 mm and a dielectric constant of 3.2, backed by a 3 mm foam with a dielectric constant of 1.05 and a scaling factor of 0.72 offers a wide phase range of 601.3°. The obtained phase slope is 76.37°/mm or 134°/GHz.

**Keywords** — microstrip antenna, reflectarray antenna, split circular, X band

## 1. Introduction

A microstrip reflectarray is a flat array of microstrip patches or dipoles printed on a thin dielectric substrate. A feed antenna illuminates the array, where each element is designed to scatter the incident field with the proper phase to make a planar phase surface in the presence of the aperture [1]. Reflectarrays thus combine the features of flat arrays and reflector antennas, and microstrip reflectarrays have emerged as a prospective solution that offers high gain and higher efficiency levels at a lower cost [2], [3].

Several resonant element designs using sophisticated approaches have been presented in the literature. These include double concentric rings [4], a double-cut ring with varying sizes [5], an elliptical dipole, slot lines, as well as a ground plane [6]. Dual band and dual circular polarizations were employed in the design of a reflectarray antenna for space communications in [7]. In this paper, a unit cell comprises two rings, an inner ring adjusted at the Ka band and a larger ring suited for the K band. The phase and amplitude of the reflecting field are regulated by stubs loaded with rings [7]. Patch elements with a cross loop and a square ring slot loaded patch were proposed in [8]. Many varied forms of fundamental elements [9]–[10], structures with multiple layers [11]–[13], and rotation methods [14], [15] were also presented in various research papers.

For the design of reflectarray unit cells, some fractal patch topologies with low losses, small sizes and appropriate phase ranges were proposed in [16], [17]. The self-similarity feature characterizes fractal geometries, where a specific shape is scaled and repeated several times to generate the final fractal shape. In practice, a fractal shape is created using an iterative technique that includes an initial shape, also known as the fractal generator, a scaling factor  $s$  used to shape the dimensions to be scaled, and the repetition number [18]. The fractal reflectarray antenna is a technique that is widely used in antenna design, for example [16], [17] describe variable-sized fractal reflectarray units. These investigations aim to attain a phase range greater than 360° and a lower gradient in the reflection phase graph, as a function of element size [17], [18]. The use of the fractal geometries offers a wider bandwidth, resulting from the various dimensions offered by the fractal shapes, as seen in paper [18]. To further enhance the solution's performance, a multilayer design and fractal geometry with aperture coupling were also employed in [19].

In [20], an asymmetric rectangular patch reflectarray antenna, comprising a single layer, is presented for 5G communication at 26 GHz. The idea is to produce two active lengths leading to two different frequencies, thus widening the bandwidth.

This paper aims to design a reflectarray unit cell element offering more than one resonance response on a single layer. The many resonance responses are then used to obtain an extension in the reflection phase range and lower slope. The circular patch providing smaller phase slopes [21] is utilized here to propose a split circle element that can achieve a double resonance response at a lower slope. The split circle microstrip unit element is investigated for the X band reflectarray design.

The paper is organized in the following manner. Section 2 presents the design of the proposed unit cell. The results of the investigations are presented in Section 3. The performance of the proposed unit cell is compared with that of other published elements in Section 4, while conclusions are drawn and described in Section 5.

## 2. Design of the Unit Cell Element

For reflectarrays, the unit cell element should provide a phase range in excess of  $2\pi$  radians and should be characterized by low sensitivity to size and frequency changes. An inadequate

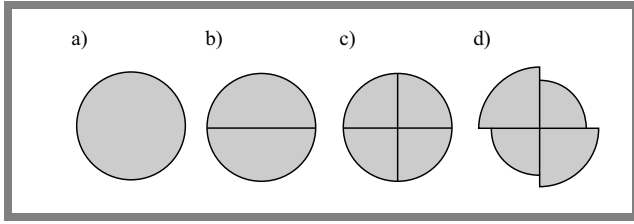


**Fig. 1.** Evolution of the split circle: a) circle, b) split into two sectors, c) split into four sectors and d) opposite sectors are scaled.

phase range might cause phase errors on the array, resulting in poor radiation pattern performance [21]. To extend the range of the phase response up to the full swing of  $360^\circ$ , a unit cell split circle is proposed here to enhance the array's performance.

The evolution of the proposed split-circle element is depicted in Fig. 1. This unit cell is derived from a circle of diameter  $d$  that is divided into two halves by the diagonal. Then, each half is further divided into two halves along the other perpendicular diagonal, resulting in four adjacent sectors. The radius ( $r = \frac{d}{2}$ ) of two oppositely located sectors is then scaled by a certain factor, as shown in Fig. 1. The four sectors are kept in contact forming a single conducting element placed on one side of a grounded substrate. The multi-dimension features in this proposed element generate more than one resonance frequency and thus increase the bandwidth of the unit cell.

The research begins with a unit cell element designed for an X band microstrip reflectarray with a central frequency of 10 GHz. Taconic TLY-5 substrate with relative permittivity of 2.2 and loss tangent  $\tan \delta$  of 0.0009 is used as a separator between the ground plane and the reflecting patch. The unit cell size is  $18.1 \times 18.1$  mm, which corresponds to  $0.6 \times 0.6 \lambda_0$ , where  $\lambda_0$  is free space wavelength at an operating frequency of 10 GHz.

### 3. Simulation Results

The proposed unit cell was investigated using the CST Microwave Studio simulator. The frequency solver was employed to determine the unit cell's characteristics. As illustrated in Fig. 2, unit cell boundary conditions are used in the transverse  $x$  and  $y$  directions, while Floquet port boundary conditions are used in the positive  $z$  direction. The incident wave is assumed to propagate in the negative  $z$  direction.

#### 3.1. Phase Range Performance

The simulations comprised four cases. In case 1, the element is mounted on top of a 3.175 mm thick Taconic TLY-5 substrate ( $\epsilon_r = 2.2$  and  $\tan \delta = 0.0009$ ) which is backed by a conducting ground plane with a thickness of 0.035 mm. Case 2 uses identical properties, except for the fact that the thickness of the substrate is altered to 1.57 mm. In the third scenario (case 3),  $\epsilon_r$  is set to 3.2, while the remaining substrate parameters remain unchanged. In case 4, the active element is printed on a 0.16 mm thick Taconic TLY-5 substrate supported by a 3 mm thick foam layer with  $\epsilon_r = 1.05$ .

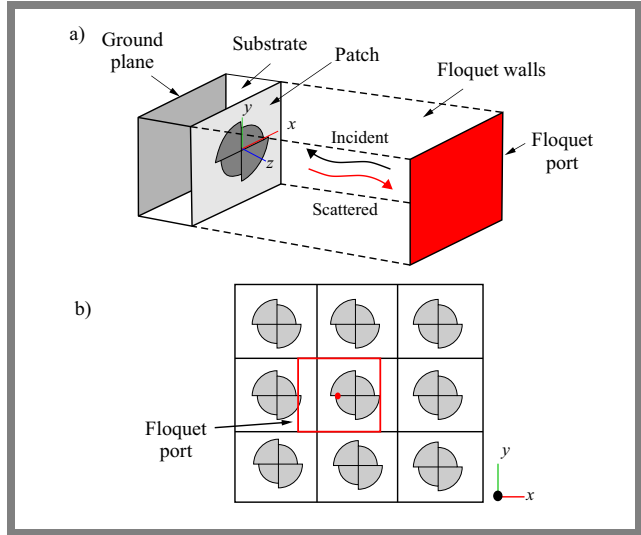


**Fig. 2.** The Floquet technique: a) boundary conditions of a unit cell and b) CST MWS model of an infinite periodic array.

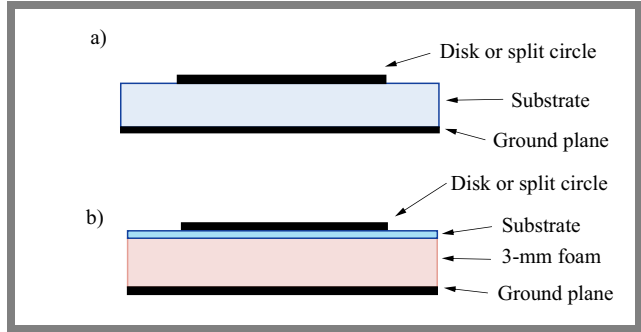


**Fig. 3.** Details of the antenna substrate configurations for: a) cases 1–3 and b) case 4.

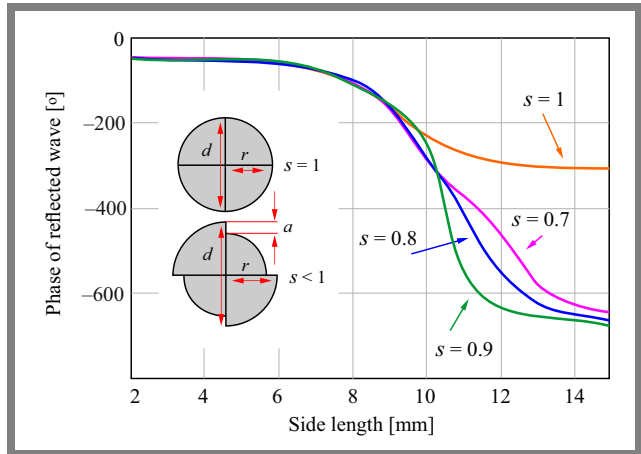


**Fig. 4.** Case 1 phase responses of the disk and split circle structures.

Figure 3 shows the antenna substrate configurations for all four cases. The phase responses for a unit cell including the disk or split circular patch on a single-layer substrate are given in each case. Outer diameter  $d$  determines the phase and magnitude of the reflected wave. Scale factor  $s$  is given for four alternative values in each scenario (1, 0.9, 0.8, 0.7). Figures 4, 5, 6, and 7 demonstrate the phases of a unit cell's reflected wave as a function of its diameter. The results were

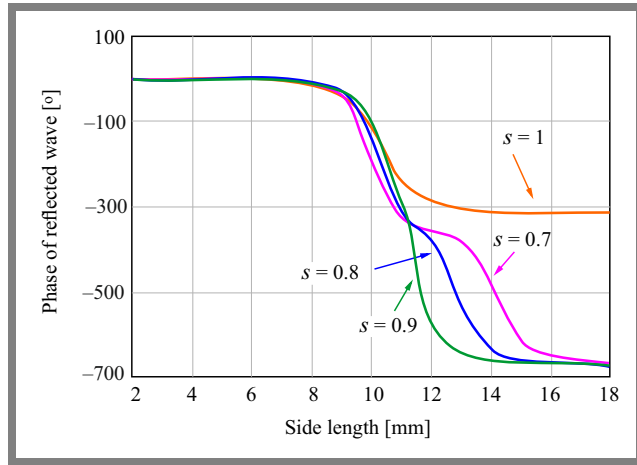


**Fig. 5.** Case 2 phase responses of the disk and split circle structures.

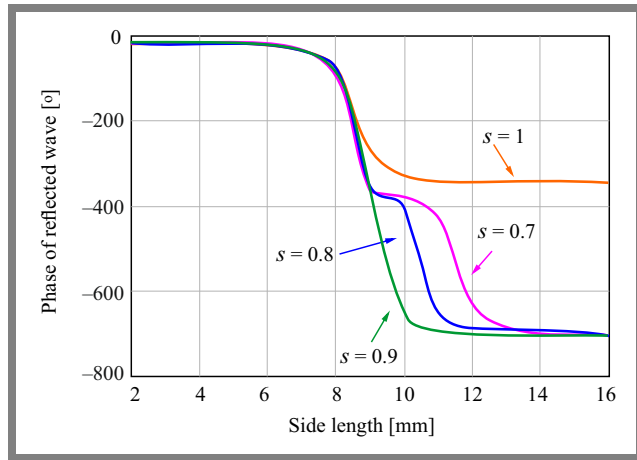


**Fig. 6.** Phase responses of the disk and split circle in case 3.

arranged in four cases, as described below. Figure 4 shows the simulation results for case 1. One can see that the disk element ( $s = 1$ ) has a phase range of  $257.2^\circ$ , which is less than the required value of  $360^\circ$ . This result is congruent with the finding presented in [21]. By decreasing  $s$  from 0.9 to 0.7 (split circle), the phase range is expanded to approx.  $615^\circ$  by altering  $d$  from 2 to 15 mm. It is also demonstrated that with the growing the value of  $s$ , the phase gradient decreases. For case 2, the results are displayed in Fig. 5. The phase range was greater than in the previous case, and reached approx.  $663^\circ$ . As  $s$  was decreased from 0.9 to 0.7, the slope becomes less linear than in case 1, particularly at  $s = 0.7$ . Figure 6 depicts the results for case 3. The obtained phase range is  $686^\circ$ , which is better than in the two previous cases. However, the linearity of the phase response deteriorates with the growing  $s$  parameter.

Figure 7 depicts the effect of attaching a 3 mm foam substrate with a dielectric constant of 1.05 to the element with a thickness of 0.16 mm and a dielectric constant of 3.2 (case 4). While lowering the  $s$  parameter from 0.9 to 0.7, the linearity remains good as compared with cases 2–3, while the phase range is  $592^\circ$ , which is less than in the preceding simulations. The value is sufficient, however, to develop a reflectarray antenna.

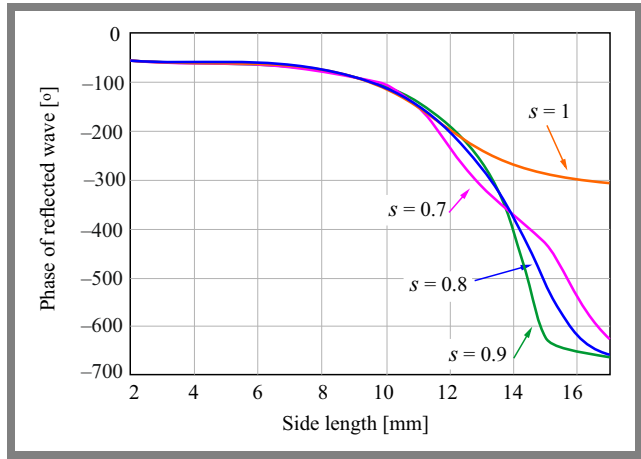


**Fig. 7.** Phase responses of the disk and split circle in case 4.

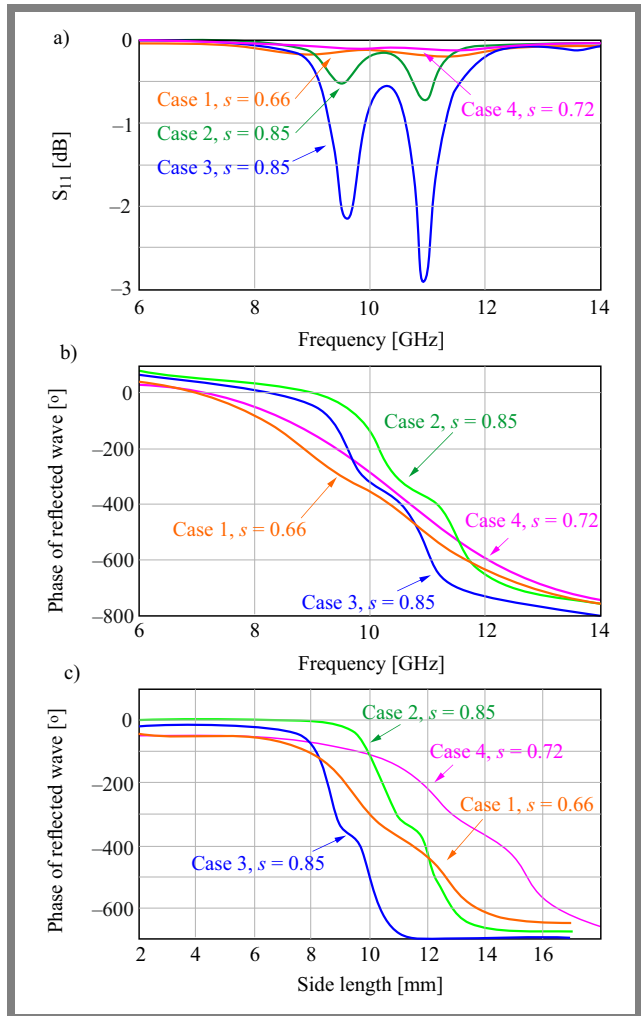


**Fig. 8.** Reflection coefficient features for the four simulated cases: a) magnitude responses versus frequency, b) phase responses versus frequency, and c) phase responses with diameter  $d$ .

### 3.2. Reflection Loss and Phase Slope of Reflection

The amplitude and phase responses of the reflection coefficients that were obtained as best values of the four cases are shown in Fig. 8. The results were gathered for  $d = 10.8, 11.0, 9.0$ , and  $12.7$  mm for the four cases, respectively. As seen in

Fig. 8a, the presence of a split circle results in two resonances in the magnitude of  $s$  parameters. The resonances are caused by the unit cell's small and large sectors, and the reflection amplitude varies between  $-2.9$  dB in case 3 to  $-0.12$  dB in case 4. It is worth noting that when the magnitude of the reflection coefficient approaches zero, then the reflection tends to be perfect or the unit cell is experiencing smaller losses. To assess the phase range performance of different reflection phase curves, a figure of merit (FoM) related to the slope of the phase curves has been created. The FoM is provided by [2]:

$$FOM = \frac{\Delta\theta}{\Delta f}, \quad (1)$$

or a phase slope that can be expressed by [17]:

$$Phase\ slope = \frac{\Delta\theta}{\Delta L}, \quad (2)$$

where  $\frac{\Delta\theta}{\Delta f}$  signifies the phase variation with the frequency determined at the phase center and  $\frac{\Delta\theta}{\Delta L}$  denotes the change in the reflection phase with size calculated at the phase center. Therefore, FoM is calculated in degrees per GHz and phase sensitivity is computed in degrees per millimeter. To improve bandwidth performance, the FoM, or slope of the phase curve around the resonance of a reflectarray antenna, must be as low as possible [17], [21]. Figures 8b and 8c illustrate the simulated reflection phases versus frequency and size for the best values of four cases. The thicker substrate (case 1 and case 4) produces a linear phase and slow fluctuation as compared to cases 2–3, as indicated by the phase responses presented in Figs. 8b and 8c and the computed values of their properties provided in Tab. 1. The smallest slopes in terms of size and frequency are obtained in case 4, which is equal to  $76.37^\circ$  per mm and  $134^\circ$  per GHz with a phase range of  $601.3^\circ$ . It is also noted that case 4, with a minimum FoM value of  $134^\circ$  per GHz, has the minimum phase range of approx.  $601.3^\circ$ , but case 3, with a maximum FoM of  $263.1^\circ$  per GHz, has a maximum phase range of  $678.6^\circ$ . It is clear that the phase range is related to the FoM.

In order to calculate the element's bandwidth from phase fluctuation, a maximum limit of  $180^\circ$  for the change in phase when the frequency varied is considered. With such a limit, an oscillating phase swing of  $\pm 90^\circ$  from the value at the center frequency of operation is thought to be sufficient to degrade the array's radiation pattern. Such a phase change limit was taken into account when estimating the reflectarray bandwidth in [22]. Table 2 lists the calculated bandwidths from the responses displayed in Fig. 8b. When compared to the other cases, the element in case 1 has a 13.4% wider bandwidth.

## 4. Comparison to Existing Papers

The performance of the proposed unit cell element is compared to earlier works utilizing various element structures. Table 3 summarizes the comparative performance of unit cell reflectarray antennas. The table shows that the developed design delivers a phase variation of  $601.3^\circ$  throughout a frequency range in the X band (8–12 GHz), which is greater than

**Tab. 1.** Calculated phase response properties, and substrate details for the split circle unit cells, as depicted in Figs. 8b and 8c.

| Parameter                            | Case 1 | Case 2 | Case 3 | Case 4 |
|--------------------------------------|--------|--------|--------|--------|
| Gradient<br>[ $^\circ$ /mm]          | 83.48  | 161    | 196    | 76.37  |
| Gradient<br>[ $^\circ$ /GHz]         | 137.8  | 241.0  | 263.1  | 134    |
| Max.<br>phase [ $^\circ$ ]           | -48.9  | 3.12   | -18.57 | -57.29 |
| Min.<br>phase [ $^\circ$ ]           | -645.2 | -670.6 | -697.2 | -658.6 |
| Phase<br>range [ $^\circ$ ]          | 596.3  | 673.7  | 678.6  | 601.3  |
| $\epsilon_{r1}$                      | 2.2    | 2.2    | 3.2    | 3.2    |
| Substrate<br>thickness<br>$h_1$ [mm] | 3.175  | 1.57   | 1.57   | 0.16   |
| $\epsilon_{r2}$ (foam)               | –      | –      | –      | 1.05   |
| Foam<br>thickness<br>$h_2$ [mm]      | –      | –      | –      | 3      |

**Tab. 2.** Calculated bandwidths for the split circle unit cells depicted in Fig. 8b.

|                    | $\theta$ at $f_r$<br>[ $^\circ$ ] | $f_1$ at<br>$\theta - 90^\circ$<br>[GHz] | $f_2$ at<br>$\theta + 90^\circ$<br>[GHz] | BW<br>$f_2 - f_1$ | Percent<br>BW |
|--------------------|-----------------------------------|------------------------------------------|------------------------------------------|-------------------|---------------|
| Case 1<br>$s=0.66$ | -356.9                            | 10.64                                    | 9.30                                     | 1.34              | 13.4          |
| Case 2<br>$s=0.85$ | -143.6                            | 10.23                                    | 9.61                                     | 0.62              | 6.2           |
| Case 3<br>$s=0.85$ | -326.5                            | 10.60                                    | 9.68                                     | 0.92              | 9.2           |
| Case 4<br>$s=0.72$ | -289.9                            | 10.60                                    | 9.40                                     | 1.20              | 12            |

the phase range achieved in the majority of the remaining papers. Nevertheless, the element size is larger when compared with the majority of other references. The area of the unit cells ranged from  $0.43 \times 0.43$  to  $0.94 \times 0.94$  effective wavelengths, with the proposed unit cell within the range. The smooth phase curve of the suggested unit cell yields a lower maximum phase sensitivity of  $76.37^\circ$  per mm when compared to the values presented in [9], [13] and [26].

## 5. Conclusions

A single-layer microstrip reflectarray element has been proposed and investigated. The element's development is based on a traditional circular patch element. Using the proposed technique, two distinct resonance frequencies can be obtained from a single element. Expansion of the reflection

**Tab. 3.** Comparison of the performance of the proposed unit cell with earlier works.

| Reference,<br>type of<br>element                                           | Phase<br>range<br>[°] | Area [mm <sup>2</sup> ]<br>$\lambda_g \cdot \lambda_g$ | Slope<br>[°/mm] | Resonance<br>freq.<br>[GHz] | Substrate<br>type | $\epsilon_{r1}$ | Substrate<br>thickness<br>[mm] | Air, foam<br>or second<br>layer<br>[mm] | tg $\delta$ |
|----------------------------------------------------------------------------|-----------------------|--------------------------------------------------------|-----------------|-----------------------------|-------------------|-----------------|--------------------------------|-----------------------------------------|-------------|
| [4], single<br>layer two<br>concentric<br>rings                            | 450                   | $15 \times 15$<br>( $0.43 \times 0.43$ )               | NA              | 11.5                        | NA                | 2.2             | 0.508                          | 6.35<br>foam                            | 0.0009      |
| [5], single<br>layer double<br>cut ring                                    | $\cong 300$           | $12.5 \times 12.5$<br>( $0.45 \times 0.45$ )           | NA              | 10                          | Arlon<br>TC600    | 6.15            | 0.508                          | 2<br>air                                | NA          |
| [8], single<br>layer cross<br>loop with<br>a square ring<br>slot           | 599                   | $10.7 \times 10.7$<br>( $0.51 \times 0.51$ )           | NA              | 14                          | RT5880            | 2.2             | 0.508                          | 3<br>air                                | NA          |
| [9], single<br>layer circular<br>ring                                      | 345                   | $15 \times 15$<br>( $0.83 \times 0.83$ )               | 525             | 10                          | NA                | 3.2             | 1.57                           | –                                       | 0           |
| [13], double<br>layer circular<br>disk                                     | $\cong 500$           | $17.65 \times$<br>$17.65$<br>( $0.94 \times 0.94$ )    | 88.23           | 8.5                         | FR4               | 4.4             | 3                              | 1.5<br>2nd layer                        | NA          |
| [16], sliced<br>circular<br>fractal                                        | $\cong 550$           | $21 \times 21$<br>( $0.72 \times 0.72$ )               | 45              | 10                          | NA                | 2.2             | 0.1                            | 3<br>foam                               | 0.004       |
| [17], single<br>layer Koch's<br>double-ring<br>fractal                     | 421.5                 | $15 \times 15$<br>( $0.54 \times 0.54$ )               | 52.9            | 10                          | NA                | 2.2             | 1                              | 3<br>foam                               | 0.004       |
| [21], single<br>layer two<br>circular rings                                | 500                   | $15 \times 15$<br>( $0.51 \times 0.51$ )               | 50              | 10                          | Laminate          | 3.4             | 0.16                           | 6<br>foam                               | 0.0009      |
| [23], single<br>layer four<br>concentrated<br>circular rings               | 382                   | $10 \times 10$<br>( $0.52 \times 0.52$ )               | NA              | 15                          | RT5880            | 2.2             | 0.762                          | 3<br>foam                               | 0.0009      |
| [24], single<br>layer<br>rectangular<br>patch and<br>e-shaped<br>structure | 500                   | $25 \times 25$<br>( $0.53 \times 0.53$ )               | NA              | 5.8                         | F4B               | 2.65            | 2                              | 4<br>air                                | 0.002       |
| [25], double<br>layer spider-<br>shaped<br>structure                       | 800                   | $8 \times 8$<br>( $0.52 \times 0.52$ )                 | NA              | 11.725                      | Taconic<br>RF 35  | 3.5             | 0.76                           | 1.52<br>2nd layer                       | 0.0018      |
| [26], single<br>layer flower<br>shape                                      | 360                   | $16 \times 16$<br>( $0.75 \times 0.75$ )               | $\cong 384$     | 12.5                        | Rogers<br>4003    | 3.55            | 0.508                          | 1<br>air                                | 0.0027      |
| This work<br>(case 4)                                                      | 601.3                 | $18.1 \times 18.1$<br>( $0.64 \times 0.64$ )           | 76.37           | 10                          | Taconic<br>TLY-5  | 3.2             | 0.16                           | 3<br>foam                               | 0.0009      |

phase swing beyond  $596^\circ$  was possible due to the formation of the additional resonance frequency. The split circle element analysis reveals that the proposed unit cell element has a sufficient phase variation range, a decent linear reflection phase, a reasonable slope, and a simple design.

## References

- [1] D.M. Pozar, S. D. Targonski, and H.D. Syrigos, "Design of Millimeter Wave Microstrip Reflectarrays", *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 2, 287–296, 1997 (<https://doi.org/10.1109/8.560348>).
- [2] J. Huang and J.A. Encinar, *Reflectarray Antennas*, Wiley & Sons, USA, 2007 (ISBN: 9780470178768).
- [3] P. Nayeri, F. Yang, and A.Z. Elsherbeni, *Reflectarray Antennas: Theory, Designs, and Applications*, Wiley & Sons, USA, 424 p., 2018 (<https://doi.org/10.1002/9781118846728>).
- [4] Y. Li, M.E. Bialkowski, K.H. Sayidmarie, and N.V. Shuley, "81-Element Single-layer Reflectarray with Double-ring Phasing Elements for Wideband Applications", in *IEEE Antennas and Propagation Society International Symposium*, Toronto, Canada, 2010 (<https://doi.org/10.1109/APS.2010.5562103>).
- [5] H. Bodur and S. Cimen, "Reflectarray Antenna Design with Double Cutted Ring Element for X-band Applications", *Microwave and Optical Technology Letters*, vol. 62, no. 10, pp. 3248–3254, 2020 (<https://doi.org/10.1002/mop.32436>).
- [6] L. Zhang *et al.*, "A Single-layer 10-30 GHz Reflectarray Antenna for the Internet of Vehicles", *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1480–1490, 2022 (<https://doi.org/10.1109/TVT.2021.3134836>).
- [7] R.L. Farias, C. Peixeiro, and M.V.T. Heckler, "Single-layer Dual-band Dual-circularly Polarized Reflectarray for Space Communication", *IEEE Transactions on Antennas and Prop.*, vol. 70, no. 7, pp. 5989–5994, 2022 (<https://doi.org/10.1109/TAP.2022.3161552>).
- [8] V. Lingasamy, M.G.N. Alsath, K.T. Selvan, and R. Jyoti, "A Wideband, Single Layer Reflectarray Antenna with Cross Loop and Square Ring Slot Loaded Patch Elements", *International Journal of Microwave and Wireless Technologies*, vol. 11, no. 7, pp. 1–8, 2019 (<https://doi.org/10.1017/S1759078719000187>).
- [9] M.E. Bialkowski and K.H. Sayidmarie, "Phasing Characteristics of a Single Layer Microstrip Reflectarray Employing Various Basic Element Shapes", in *Proc. of 2008 Int. Workshop on Antenna Technology: Small Antennas and Novel Metamaterials*, Chiba, Japan, pp. 79–82, 2008 (<https://doi.org/10.1109/IWAT.2008.4511295>).
- [10] S.A.M. Soliman, A.M. Attiya, and Y.M. Antar, "Low Profile/Single Layer X-Band Circularly Polarized Reflectarray with a Linearly Polarized Feed", *Progress in Electromagnetics Research M*, vol. 113, pp. 87–99, 2022 (<https://dx.doi.org/10.2528/PIERM22070106>).
- [11] C. Zhang *et al.*, "Ka Band Multi-layer Folded Reflectarray using Dual-polarized Slot Antennas as Unit Cells", in *Proc. of European Space Agency Space Antenna Workshop*, Noordwijk, Netherlands, Oct 2012 ([https://artes.esa.int/sites/default/files/03\\_1745\\_Zhang1.pdf](https://artes.esa.int/sites/default/files/03_1745_Zhang1.pdf)).
- [12] R. Florencio *et al.*, "Reflectarray Antennas for Dual Polarization and Broadband Telecom Satellite Applications", *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 4, pp. 1234–1246, 2015 (<https://doi.org/10.1109/TAP.2015.2391279>).
- [13] S.A.M. Soliman, E.M. Eldesouki, and A.M. Attiya, "Analysis and Design of an X-Band Reflectarray Antenna for Remote Sensing Satellite System", *Sensors*, vol. 22, no. 3, art. no. 1166, 2022 (<https://doi.org/10.3390/s22031166>).
- [14] J. Huang, and R.J. Pogorzelski, "A Ka-band Microstrip Reflectarray with Elements Having Variable Rotation Angles", *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 5, pp. 650–656, 1998 (<https://doi.org/10.1109/8.668907>).
- [15] R. Florencio *et al.*, "Design of Ku- and Ka-band Flat Dual Circular Polarized Reflectarrays by Combining Variable Rotation Technique and Element Size Variation", *Electronics*, vol. 9, no. 6, art. no. 985, 2020 (<https://doi.org/10.3390/electronics9060985>).
- [16] K.H. Sayidmarie and M.E. Bialkowski, "Fractal Unit Cells of Increased Phasing Range and Low Slopes for Single-layer Microstrip Reflectarrays", *IET Microwaves, Antennas, and Propagation*, vol. 5, no. 11, pp. 1371–1379, 2010 (<https://doi.org/10.1049/iet-map.2010.0581>).
- [17] K.H. Sayidmarie and A.M. Saleh, "Comparison of Phase Responses of Proposed Element Shapes for Reflectarray Unit Cells", in *International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE)*, Beijing, China, 2011 (<https://doi.org/10.1109/MAPE.2011.6156273>).
- [18] S. Costanzo *et al.*, "Fractal Reflectarray Antennas: State of Art and New Opportunities", *International Journal of Antennas and Propagation*, art. no. 7165143, 2016 (<https://dx.doi.org/10.1155/2016/7165143>).
- [19] E. Ozturk and B. Saka, "Multilayer Minkowski Reflectarray Antenna with Improved Phase Performance", *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 12, pp. 8961–8966, 2021 (<https://doi.org/10.1109/TAP.2021.3090533>).
- [20] M.H. Dahri *et al.*, "A Novel Asymmetric Patch Reflectarray Antenna with Ground Ring Slots for 5G Communication Systems", *Electronics*, vol. 9, no. 9, art. no. 1450, 2020 (<https://doi.org/10.3390/electronics9091450>).
- [21] M.E. Bialkowski and K.H. Sayidmarie, "Investigations into Phase Characteristics of a Single-layer Reflectarray Employing Patch or Ring Elements of Variable Size", *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 11, pp. 3366–3372, 2008 (<https://doi.org/10.1109/TAP.2008.2005470>).
- [22] K.H. Sayidmarie and A.M. Saleh, "Evaluation of Phase Responses of Double Ring Elements for Reflectarray by Simulation and Measurement", in *Fourth International Conference on Computational Intelligence, Communication Systems and Networks*, Phuket, Thailand, 2012 (<https://doi.org/10.1109/CICSyN.2012.30>).
- [23] L. Veluchamy, K.T. Selvan, R. Jyoti, and S.S. Kumar, "Wideband Reflectarray Antennas using Concentric Ring-based Elements for Ku-band Applications", *IETE Journal of Research*, vol. 69, no. 3, pp. 1675–1685, 2023 (<https://doi.org/10.1080/03772063.2021.1874842>).
- [24] H. Zhang *et al.*, "A Single-layer Circularly Polarized Reflectarray Antenna with High Aperture Efficiency for Microwave Power Transmission", *International Journal of Antennas and Propagation*, art. no. 1569710, 2023 (<https://doi.org/10.1155/2023/1569710>).
- [25] M. Elahi, T. Jeong, Y. Yang, K.-Y. Lee, and K.C. Hwang, "A Wideband Reflectarray Antenna for Satellite Application with Low Cross-polarization", *Applied Science*, vol.13, no. 7, art. no. 4545, 2023 (<https://doi.org/10.3390/app13074545>).
- [26] E. Hadian, M-M. Askhiri, S. Fakhte, and D.Ramezani, "Design of a Wideband Flower-like Shape Metamaterial Reflectarray Antenna", *IETE Journal of Research*, 2023 (<https://doi.org/10.1080/03772063.2023.2173671>).

**Khalil H. Sayidmarie, Professor**

College of Electronics Engineering

 <https://orcid.org/0000-0001-6525-0949>

E-mail: kh.sayidmarie@uoninevah.edu.iq

Ninevah University, Mosul, Iraq

<https://uoninevah.edu.iq>

**Likaa S. Yahya, Ph.D.**

Institute of Technology

 <https://orcid.org/0000-0003-2965-158X>

Northern Technical University, Mosul, Iraq

<https://ntu.edu.iq>

# Design and Analysis of a Low-profile Microstrip Antenna for 5G Applications using AI-based PSO Approach

Krishanu Kundu<sup>1</sup>, Ankan Bhattacharya<sup>2</sup>, Firdous H. Mohammed<sup>3</sup>, and Narendra Nath Pathak<sup>4</sup>

<sup>1</sup>G.L. Bajaj Institute of Technology & Management, Greater Noida, India,

<sup>2</sup>Hooghly Engineering & Technology College (HETC), Hooghly, West Bengal, India,

<sup>3</sup>University of the Cumberland's Williamsburg, Kentucky, USA,

<sup>4</sup>Dr. B.C. Roy Engineering College Durgapur, India

<https://doi.org/10.26636/jtit.2023.3.1368>

**Abstract** — Microstrip antennas are high gain aerials for low-profile wireless applications working with frequencies over 100 MHz. This paper presents a study and design of a low cost slotted-type microstrip patch antenna that can be used in 5G millimeter wave applications. This research focuses on the effect of ground slots and patch slots which, in turn, affect different antenna parameters, such as return loss, VSWR, gain, radiation pattern, and axial ratio. The working frequency range varies from 24 to 28 GHz, thus falling within 5G specifications. A subset of artificial intelligence (AI) known as particle swarm optimization (PSO) is used to approximatively solve issues involving maximization and minimization of numerical values, being highly challenging or even impossible to solve in a precise manner. Here, we have designed and analyzed a low-profile printed microstrip antenna for 5G applications using the AI-based PSO approach. The novelty of the research is mainly in the design approach, compactness of size and antenna applicability. The antenna was simulated with the use of HFSS simulation software.

**Keywords** — 5G applications, high gain, low profile, microstrip patch antenna, PSO

The benefits of microstrip patch antennas are in their compact size, simple manufacturing methods, low weight, and easy design. This has led to the replacement of traditional antennas used in mobile devices [1]. Selection of the substrate and determination of the patch proportions depend on the operating frequency and on the specification of the dielectric material used. Significant parameters, such as length and width of the patch and substrate, as well as the location and length of the feed network can be estimated employing equations given in [2]–[5]. The available dielectric materials have their unique conduction characteristics and vary in dielectric constants and other parameters that influence the fringing waves in the antenna patch [2]–[5]. Because of the cost factor, the most popular dielectric materials include bakelite, FR4 glass epoxy, Rogers RO4003, Taconic TLC, and Rogers RT/Duroid [6]–[8]. When feeding the signal to the antenna, many techniques, including proximity-coupled, inset feed, aperture-coupled, as well as coaxial probe feed methods, are used [3]–[4], [9].

## 1. Introduction

A low-profile micro strip or patch antenna (MPA) is created by mounting a metal patch plane over the ground level with a dielectric separator between them. Typically, the feed lines and the radiating patch are manufactured of a dielectric substrate using the PCB process. The patch is often square, rectangular, circular, triangular, or elliptical in shape to meet performance-related requirements. For a rectangular patch, length  $L$  of the patch is usually in the  $0.33\lambda_0 \dots 0.5\lambda_0$  range, where  $\lambda_0$  represents the free-space wavelength. The patch must be very thin, hence the thickness of the copper foil used is less than  $\lambda_0$ . The dielectric substrate's height is in the  $0.003\lambda_0 \dots 0.05\lambda_0$  range, while the dielectric constant ranges between 2.2 and 12.

Due to their small size and low profile, microstrip patch antennas have become increasingly popular in smartphones and other consumer electronics devices relying on wireless com-

## 2. Related Work

Many researchers have used various feeding techniques in recent years to develop MPAs of various shapes and used them in various applications. Sharma proposed, in [10], a small, high gain multiband antenna with a glass-shaped radiating patch and a rectangular ground plane. A unique rectangular-shaped, DGS-based effective multi-band frequency reconfigurable antenna was proposed by Sathikbasha *et al.* in [11]. A portable multi-band MPA with resonances at 23.9, 35.5, and 70.9 GHz, suitable for 5G mobile applications, was demonstrated by Punith *et al.* [12].

A graphene-packed dual band mmWave antenna for 28.1 GHz and 37.4 GHz with a DC bias was proposed by Luo *et al.* [13]. A tiny, portable ultra-wideband microstrip antenna for 5G applications was developed by Araujo *et al.* [14], while a small and dual-polarized triple-band antenna for sub-6 GHz 5G applications was created by Alieldin *et al.* [15].

Naji demonstrated in [16], three fork-shaped monopole-based CPW-fed MPAs: SFA, MFA, and DSFA. Ikram *et al.* [17] presented a MIMO antenna that operates in WLAN (2.45/5.2 GHz), 4G LTE (2.6 GHz), and 5G (24/28 GHz) bands, incorporating a 4-element dual-band and wideband monopole tapered slot system. Parchin *et al.* [18] presented a multi-band slot antenna array for 4G and 5G smartphones, including double element square-ring slot radiators mounted on the PCB corners. Two compact, rectangular, u-shaped parasitic elements-based antennas for 2.6, 6.0, and 8.5 GHz bands have been conceptualized by Asif *et al.* [19]. Such a design is ideal for WiMAX or weather radar applications. A u-shape slot-terminated reconfigurable patch antenna with a low profile and light weight was also shown by Yeole *et al.* [20], while a compact antenna for wideband applications has been proposed by Bhattacharya *et al.* in [21]–[24].

Artificial intelligence (AI) is a field of science in which particle swarm optimization is used [25], [26]. Three main approaches are utilized there [27]: statistical methods, symbolic artificial intelligence, and computational intelligence. Computational intelligence can be implemented using either of the three methods: artificial neural network, fuzzy logic, and evolutionary computation. Swarm intelligence is an active part of evolutionary computation. Just like in the case of ant colony optimization, the working principle of particle swarm optimization (PSO) [28] completely relies on swarm intelligence.

Here, PSO has been applied to optimize the dimensions of a compact sized microstrip antenna. Different iterative structures, namely antennas A–D, have been simulated and optimized using PSO. Finally, we have obtained the D antenna model which is validated by the simulation results and measurements.

### 3. Antenna Design

In this study, the design is realized by cutting rectangular holes at the patch antenna's radiating edge. That is because slot loading antennas perform better – in terms of resonant frequency, return loss, and bandwidth cost – than a traditional rectangular patch aerial. In the 24 GHz band, the antenna's parameters, such as return loss, VSWR, and radiation pattern have been analyzed, as the decreased bandwidth of microstrip patch antennas is their principal drawback.

The equation for the effective dielectric constant for MPA is [2]:

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[ 1 + 12 \frac{h}{w} \right]^{-\frac{1}{2}}. \quad (1)$$

The patch length is stretched on either end by  $\Delta L$  and can be formulated as [2]:

$$\frac{\Delta L}{h} = 0.412 \frac{(\epsilon_{\text{reff}} + 0.3) \left( \frac{w}{h} + 0.264 \right)}{(\epsilon_{\text{reff}} - 0.258) \left( \frac{w}{h} + 0.8 \right)}. \quad (2)$$

The effective length of the patch  $L_{\text{eff}}$  is:

$$L_{\text{eff}} = L + 2\Delta L. \quad (3)$$

VSWR is the ratio between the maximum and minimum RF voltage levels in the antenna. The ratio of incident wave and reflected voltage wave amplitudes can be defined as a reflection coefficient  $\rho$ , retrieved from:

$$r = \frac{Z_{\text{input}} - Z_0}{Z_{\text{input}} + Z_0}, \quad (4)$$

where  $Z_0$  is the antenna's characteristic impedance. Reflection coefficient  $\rho$ , VSWR and characteristic impedance  $Z_0$  are:

$$\rho = |r| = \frac{VSWR - 1}{VSWR + 1}, \quad (5)$$

$$VSWR = \frac{|r| + 1}{|r| - 1}, \quad (6)$$

$$Z_0 = \sqrt{\frac{L}{C}}, \quad (7)$$

where  $L$  signifies inductance and  $C$  is the antenna's capacitance. The return loss parameter  $S_{11}$  is given by:

$$S_{11} = -20 \log \left[ \frac{VSWR + 1}{VSWR - 1} \right], \quad (8)$$

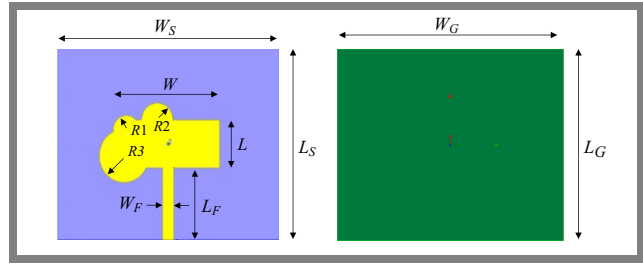


Fig. 1. Reference antenna A a) and ground plane b).

A slot loading antenna outperforms traditional rectangular patch versions in terms of resonance frequency, return loss, and bandwidth. It allows to reduce the aerial's size by more than 60% and lowers resonance frequency by up to 36%. The use of additional slots may contribute to increasing the antenna's bandwidth even further. High gain may be achieved by using a high permittivity substrate and a modified slot shape. This shows that the choice of the substrate and of its thickness is an important design factor. In this research, FR4 substrate is used with its thickness equaling 0.8 mm and its dielectric constant amounting to  $\epsilon_r$  of 4.4.

### 4. PSO in Antenna Design

Particle swarm optimization [29] is a very effective algorithm for solving optimization issues. The given solutions are referred to as “particles”, where the best or most appropriate value is referred to as p-best, while g-best refers to the overall best value. The goal is to acquire the best possible p- and g-values for the researched issues. PSO may

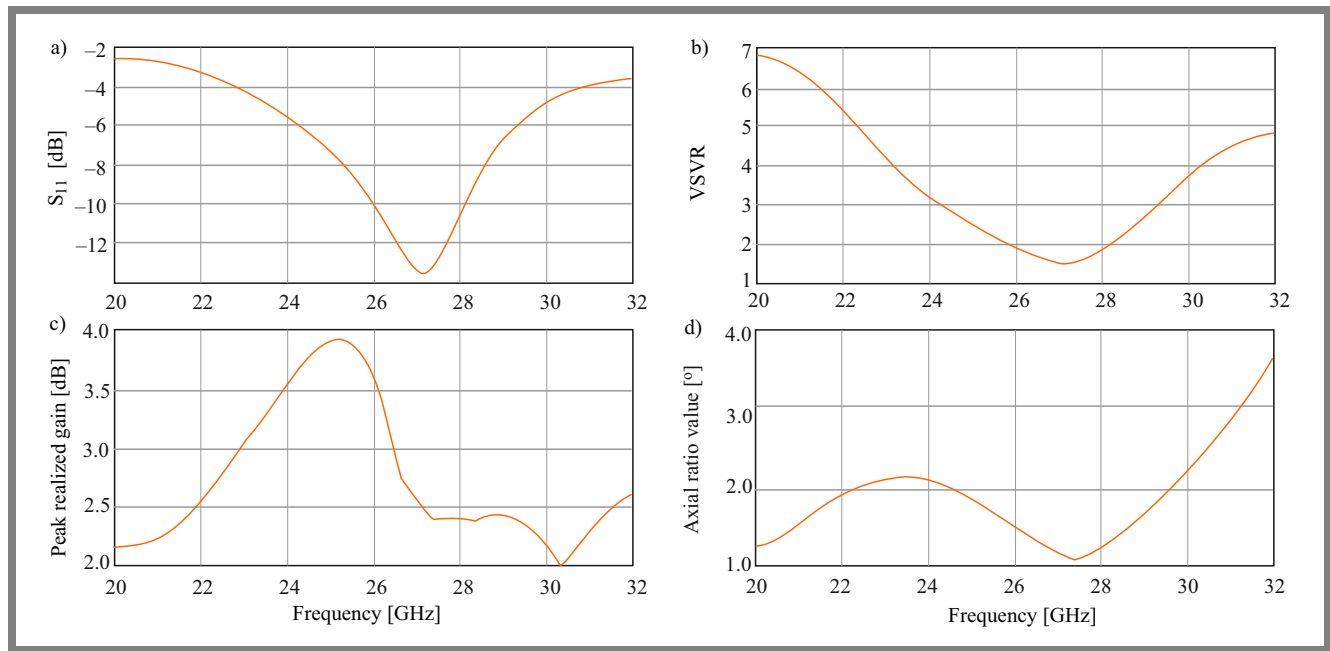


**Fig. 2.** Antenna A: a) return loss, b) VSWR, c) gain, and d) axial ratio.

be modified in several ways [31], e.g., by replacing the global swarm with several sub-swarms of particles. Such a sub-swarm design aims to mitigate the probability of the PSO algorithm being trapped in a suboptimal solution, which is a useful feature in the design of antenna. Here, PSO is utilized to acquire superior aerial characteristics and to find the best dimensions in order to reduce the return loss as:

$$\min f(x) = \text{ReturnLoss}(x) \text{ subject to } x_l \leq x \leq x_u, \quad (9)$$

**Tab. 1.** Dimensions of antennas A-D.

| Dimension        | Value [mm] | Dimension | Value [mm] |
|------------------|------------|-----------|------------|
| <b>Antenna A</b> |            |           |            |
| $W_S, W_G$       | 10.00      | $L$       | 2.00       |
| $L_S, L_G$       | 8.00       | R1        | 0.58       |
| $L_F$            | 3.00       | R2, R3    | 0.70, 1.12 |
| $W$              | 4.62       | $W_F$     | 0.5        |
| <b>Antenna B</b> |            |           |            |
| S1               | 0.08       | S8        | 0.44       |
| S2               | 0.30       | S9        | 0.18       |
| S3, S13          | 0.24       | S10       | 0.20       |
| S4               | 0.38       | S11       | 0.31       |
| S5               | 0.12       | S12       | 0.19       |
| S6, S7           | 0.16       | S14       | 0.24       |
| <b>Antenna C</b> |            |           |            |
| G1               | 2.00       | G2        | 2.00       |
| <b>Antenna D</b> |            |           |            |
| S15              | 0.13       | S16       | 0.14       |

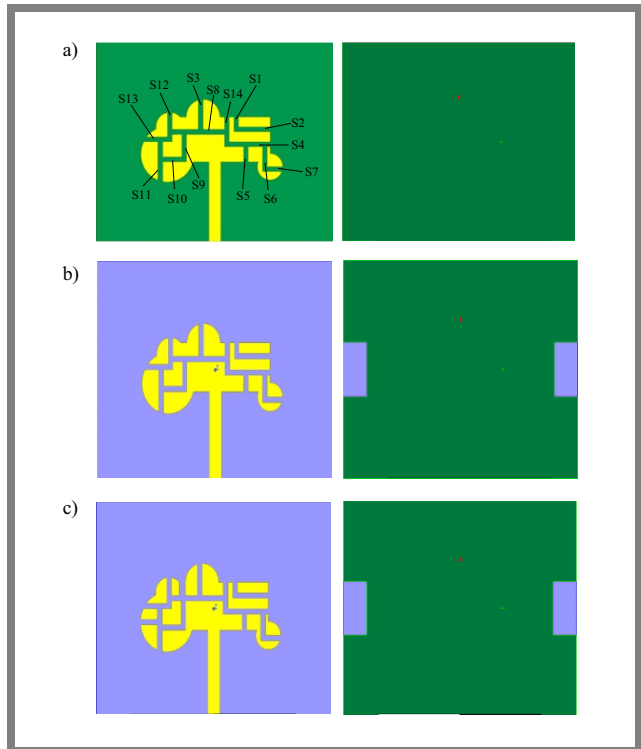


**Fig. 3.** Patch and ground plane for: a) antenna B, b) antenna C, and c) antenna D.

where  $x_l$  and  $x_u$  are the minimum and maximum values of MPA dimensions (length and width), respectively. For PSO optimization the following parameters were used:

- number of variables = 2 ( $L, W$ ),
- acceleration factors = 2, 2,
- number of iterations = 1000,
- population size = 100.

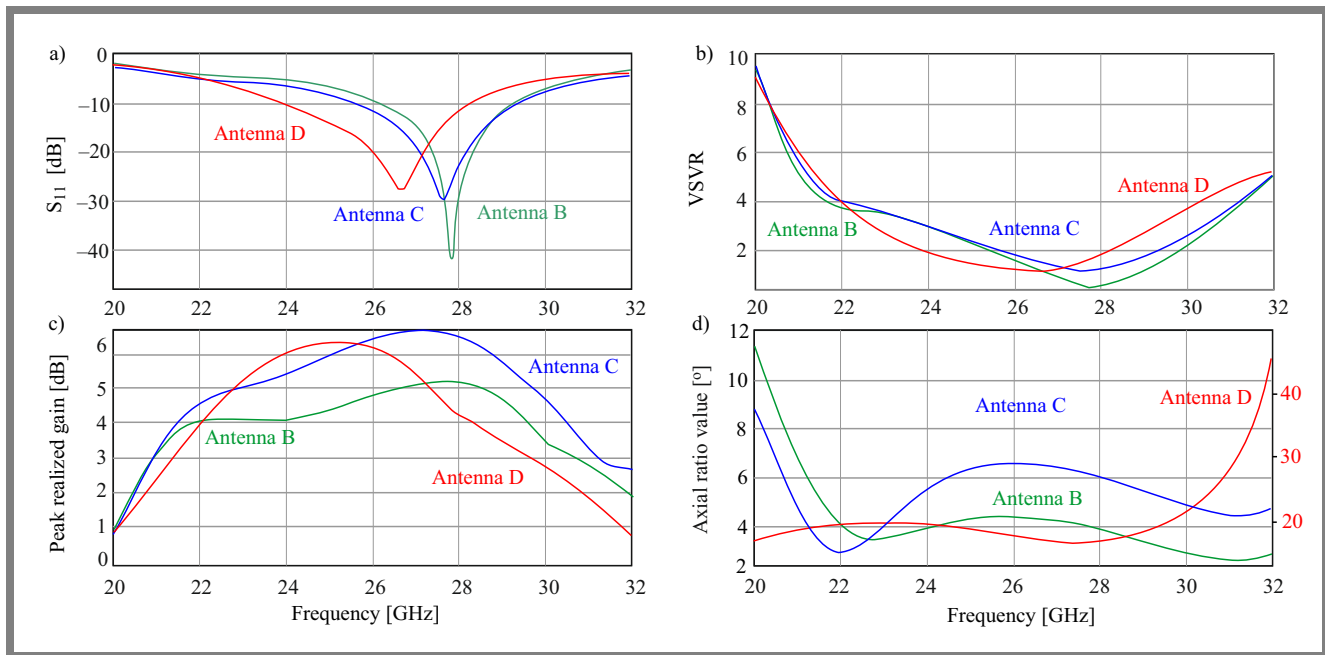


Fig. 4. Return loss a), VSWR b), gain c), and axial ratio d) for antennas B, C, D.

## 5. Antenna Simulation

The initial reference design is named as antenna A (see Fig. 1) and the dimensions of this design are presented in Tab. 1. The assumed initial dimensions offer a gain of 2.5 dB, return loss  $S_{11}$  of  $\geq 30$  dB at approx. 27 GHz, VSWR = 1.5 (Fig. 2a-c). However, the axial ratio does not meet the requirement of being lower than 3 dB (Fig. 2d). Therefore, further improvements are required.

In the next iteration (antenna B), slots have been introduced to the design and a new structure was created, as shown in Fig. 3a, Tab. 1. Simulation results presented in Fig. 4 (green curve) show that the slots have affected the axial ratio, the  $S_{11}$  has shifted towards 28 GHz with VSWR = 1. The frequency band with a high gain of approx. 4 dB is wider than for antenna A. Further modifications aim to create a version with an even higher gain (antenna C, Fig. 3b, and Tab. 1). Simulation results presented in Fig. 4 (blue curve) show that antenna C is characterized by a return loss of  $\geq 45$  dB, VSWR of 1, gain of 6 dBi and a bandwidth of 3.2 GHz. Due to the slots present

on the ground plane, a narrow band has been achieved with a minimum  $S_{11}$  value near 28 GHz.

To increase gain even more, another design was developed, referred as antenna D (Fig. 3c, Tab. 1). From the simulations we can conclude that for antenna D, the return loss equals  $\geq 40$ , with VSWR of 0.8 (Fig. 4, red curve) and the gain reached 6.6 dB. An acceptable change has been observed in the axial ratio and a wide bandwidth of 4 GHz. Table 2 shows the comparison of parameters of all simulated antennas (A to D) while Fig. 5 shows the evolution of the patch for A–D prototypes.

Tab. 2. Comparison of specific antenna designs.

| Antenna | Return loss $S_{11}$ | VSWR | Bandwidth | Gain   |
|---------|----------------------|------|-----------|--------|
| A       | $\geq 30$            | 1.5  | 2.1 GHz   | 2.5 dB |
| B       | $\geq 33$            | 1    | 3.0 GHz   | 4 dB   |
| C       | $\geq 45$            | 1    | 3.2 GHz   | 6 dB   |
| D       | $\geq 40$            | 0.8  | 4.0 GHz   | 6.6 dB |

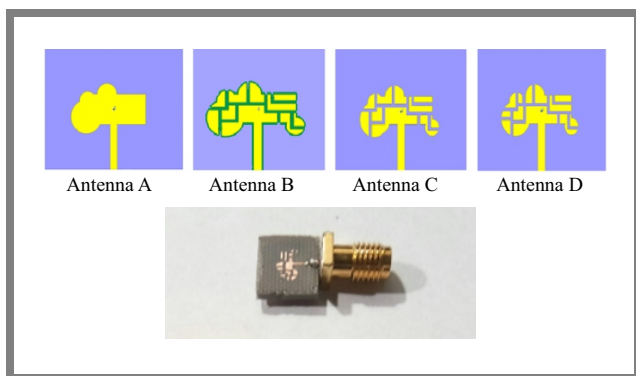


Fig. 5. Antenna design evolution (A to D) and fabricated prototype.

## 6. Antenna Fabrication and Measurements

To verify the simulation data, the designed antennas were fabricated and the return losses were measured using a vector network analyzer (VNA). Figure 6 shows the the value of the return loss parameter for all four prototypes.

## 7. Conclusions

In this work, the effects of changing the dimensions of ground slots, as well as the patch of a 5G microstrip antenna [31], [32],

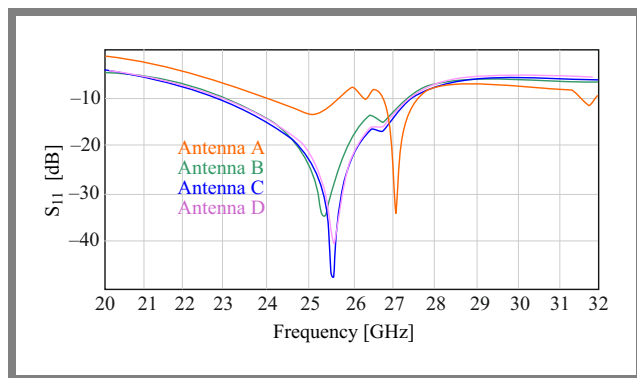


**Fig. 6.** Return loss of: a) antenna A, b) antenna B, c) antenna C, and d) antenna D.

impacting the different parameters of the antenna, have been analyzed. The proposed antenna is designed using the PSO approach and uses the low-cost and widely available FR4 substrate. It offers a high gain of 6.6 dB with a return loss  $|S_{11}|$  parameter of over 40 dB at 27 GHz. Post-optimization of the design using the PSO technique has resulted in a compact structure with minimal dimensions.

## References

- [1] J. Howell, "Microstrip Antennas", *IEEE Transactions on Antennas and Propagation*, vol. 23, no. 1, pp. 90–93, 1975 (<https://doi.org/10.1109/TAP.1975.1141009>).
- [2] C.A. Balanis, "Antenna Theory: a Review", *Proceedings of the IEEE*, vol. 80, no. 1, pp. 7–23, 1992 (<https://doi.org/10.1109/5.119564>).
- [3] C.A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Wiley, 848 p., 2005 (ISBN: 9780470576649).
- [4] G. Kumar and K.P. Ray, *Broadband Microstrip Antennas*, 1st ed. Artech, 451 p., 2002 (ISBN: 9781580537674).
- [5] P. Upender, P.A.H. Vardhini, and V. Prakasam, "Performance Analysis and Development of Printed Circuit Microstrip Patch Antenna with Proximity Coupled Feed at 4.3 GHz (C-band) with Linear Polarization for Altimeter Application", *International Journal of Computing and Digital Systems*, vol. 10, no. 1, 2020 (<https://dx.doi.org/10.12785/ijcds/1001116>).
- [6] O. Barrou, A. El Amri, and A. Reha, "Comparison of Feeding Modes for a Rectangular Microstrip Patch Antenna for 2.45 GHz Applications", *Lecture Notes in Electrical Engineering*, vol. 397, pp. 457–469, 2016 ([https://doi.org/10.1007/978-981-10-1627-1\\_36](https://doi.org/10.1007/978-981-10-1627-1_36)).
- [7] A.K. Khan, T.A. Shaem, and M.A. Alim, "Graphene Patch Antennas with Different Substrate Shapes and Materials", *Optik*, vol. 202, art. no. 163700, 2020 (<https://doi.org/10.1016/j.ijleo.2019.163700>).
- [8] G. Kaur *et al.*, "Performance analysis of conductive patch materials for the design and fabrication of microstrip patch antennas", in: *2017 Progress in Electromagnetics Research Symposium*, St. Petersburg, Russia, 2017 (<https://doi.org/10.1109/piers.2017.8261793>).
- [9] A. Khan and R. Nema, "Analysis of Five Different Dielectric Substrates on Microstrip Patch Antenna", *International Journal of Computer Applications*, vol. 55, no. 14, pp. 40–47, 2012 (<https://doi.org/10.5120/8826-2905>).
- [10] M. Sharma, "Design and Analysis of Multi-band Antenna for Wireless Communication", *Wireless Personal Communications*, vol. 114, no. 2, pp. 1389–1402, 2020 (<https://doi.org/10.1007/s11277-020-07425-9>).
- [11] M.J. Sathikbasha and V. Nagarajan, "Design of Multi-band Frequency Reconfigurable Antenna with Defected Ground Structure for Wireless Applications", *Wireless Personal Comm.*, vol. 113, no. 2, pp. 867–892, 2020 (<https://doi.org/10.1007/s11277-020-07256-8>).
- [12] S. Punith, S.K. Praveenkumar, A.A. Jugale, and M.R. Ahmed, "A Novel Multi-band Microstrip Patch Antenna for 5G Communications", *Procedia Computer Science*, vol. 171, pp. 2080–2086, 2020 (<https://doi.org/10.1016/j.procs.2020.04.224>).
- [13] Y. Luo *et al.*, "Graphene-based dual-band antenna in the millimeter-wave band", *Microwave and Optical Technology Letters*, vol. 60, no. 12, pp. 3014–3019, 2018 (<https://doi.org/10.1002/mop.31435>).
- [14] H.X. Araujo, A.E. Freitas, D.N. Prata, I.R.S. Casella, and C.E. Capovilla, "Multi-band Antenna Design Comprising the Future 5G Mobile Technology", *Electrotechnical Review*, vol. 95, no. 2, pp. 108–111, 2019 (<https://doi.org/10.15199/48.2019.02.25>).
- [15] A. Alieldin *et al.*, "A Triple-Band Dual-Polarized Indoor Base Station Antenna for 2G, 3G, 4G and Sub-6 GHz 5G Applications", *IEEE Access*, vol. 6, pp. 49209–49216, 2018 (<https://doi.org/10.1109/access.2018.2868414>).
- [16] D.K. Naji, "Design of Compact Dual-band and Tri-band Microstrip Patch Antennas", *International Journal of Electromagnetics and Applications*, vol. 8, no. 1, pp. 26–34, 2018 (<https://doi.org/10.5923/j.ijea.20180801.02>).
- [17] M. Ikram, N. Nguyen-Trong, and A. Abbosh, "Multi-band MIMO Microwave and Millimeter Antenna System Employing Dual-Function Tapered Slot Structure", *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 8, pp. 5705–5710, 2019 (<https://doi.org/10.1109/tap.2019.2922547>).
- [18] N.O. Parchin *et al.*, "Multi-Band MIMO Antenna Design with User-Impact Investigation for 4G and 5G Mobile Terminals", *Sensors*, vol. 19, no. 3, art. no. 456, 2019 (<https://doi.org/10.3390/s19030456>).
- [19] S. Asif *et al.*, "A Compact Multi-band Microstrip Patch Antenna with U-shaped Parasitic Elements", in: *2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, Vancouver, Canada, 2015 (<https://doi.org/10.1109/aps.2015.7304695>).
- [20] D.S. Yeole and U.P. Khot, "Reconfigurable Multi-band Microstrip Patch Antenna Design for Wireless Communication Applications", in: *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2016 (<https://doi.org/10.1109/rteict.2016.7807899>).
- [21] A. Bhattacharya, B. Roy, R.F.S. Caldeirinha, and A.K. Bhattacharjee, "Low-profile, Extremely Wideband, Dual-band-notched MIMO Antenna for UWB Applications", *International Journal of Microwave and Wireless Technologies*, vol. 11, no. 7, pp. 719–728, 2019 (<https://doi.org/10.1017/S1759078719000266>).
- [22] A. Bhattacharya, B. Roy, S.K. Chowdhury, and A.K. Bhattacharjee, "Computational and Experimental Analysis of a Low-profile, Isolation-enhanced, Band-notch UWB-MIMO Antenna", *Journal of Computational Electronics*, vol. 18, no. 2, pp. 680–688, 2019 (<https://doi.org/10.1007/s10825-019-01309-3>).
- [23] A. Bhattacharya, B. Roy, S.K. Chowdhury, and A.K. Bhattacharjee, "A Compact Fractal Monopole Antenna with Defected Ground Structure for Wideband Communication", *Applied Computational Electromagnetics Society Journal*, vol. 33, no. 3, pp. 347–350, 2018 (ISSN: 10544887).
- [24] A. Bhattacharya, B. Roy, S.K. Chowdhury, and A.K. Bhattacharjee, "Design and Analysis of a Koch Snowflake Fractal Monopole Antenna for Wideband Communication", *Applied Computational Electromagnetics Society Journal*, vol. 32, no. 6, 2017 (ISSN: 10544887).
- [25] N. Behnia, M. Zare, V. Moosavi, and S.J. Khajeddin, "An inter-comparison of Different PSO-Optimized Artificial Intelligence Algorithms for Thermal-based Soil Moisture Retrieval", *Earth Science Informatics*, vol. 15, no. 11, pp. 473–484, 2022 (<https://doi.org/10.1007/s12145-021-00747-7>).
- [26] V. Chourasia, S. Pandey, and S. Kumar, "Optimizing the Performance of Vehicular Delay Tolerant Networks Using Multi-objective PSO and Artificial Intelligence", *Computer Communications*, vol. 177, pp.

- 10–23, 2021 (<https://doi.org/10.1016/j.comcom.2021.06.006>).
- [27] S. Zhao, W. Xu, and L. Chen, “The Modeling and Products Prediction for Biomass Oxidative Pyrolysis Based on PSO–ANN Method: an Artificial Intelligence Algorithm Approach”, *Fuel*, vol. 312, art. no. 122966, 2022 (<https://doi.org/10.1016/j.fuel.2021.122966>).
- [28] A.G. Gad, “Particle Swarm Optimization Algorithm and its Applications: a Systematic Review”, *Archives of Computational Methods in Engineering*, vol. 29, no. 12, pp. 2531–2561, 2022 (<https://doi.org/10.1007/s11831-021-09694-4>).
- [29] J. Kennedy and R. Eberhart, “Particle Swarm Optimization”, in *Proc. of IEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995 (<https://doi.org/10.1109/ICNN.1995.488968>).
- [30] K. Sharma and G.P. Pandey, “Pattern Correction in Planar Antenna Array Using Multi-objective Particle Swarm Optimization and Continuous Excitation”, *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 45, no. 2, pp. 367–380, 2021 (<https://doi.org/10.1007/s40998-020-00378-6>).
- [31] K. Sharma, “Investigations of Microstrip Antenna Using Machine Learning”, *Ph.D. Thesis*, Pandit Deendayal Energy University, Gandhinagar, 2021 (<http://ir.pdpu.ac.in:8080/xmlui/handle/123456789/102>).
- [32] A. Bhattacharya and S. Pal, “An Extremely Compact and Low-Cost Antenna Sensor Designed for IoT-Integrated Biomedical Applications”, in *Internet of Things and Data Mining for Modern Engineering and Healthcare Applications*, pp. 261–268, Chapman and Hall/CRC, 2022 (<https://doi.org/10.1201/9781003217398>).

#### **Krishanu Kundu, Ph.D.**

Department of Electronics and Communication Engineering  
 <https://orcid.org/0000-0003-1057-4713>  
 E-mail: krishanukundu08@gmail.com  
 G.L. Bajaj Institute of Technology & Management, Greater Noida, India  
<https://www.glbitm.org>

#### **Ankan Bhattacharya, Ph.D.**

Department of Electronics and Communication Engineering  
 <https://orcid.org/0000-0003-2350-1687>  
 Hooghly Engineering & Technology College (HETC), Hooghly, West Bengal, India  
<https://www.hetc.ac.in>

#### **Firdous H. Mohammed, Ph.D. Scholar**

Computer Information System  
 University of the Cumberland’s Williamsburg, Kentucky, USA  
<https://www.ucumberlands.edu>

#### **Narendra Nath Pathak, Ph.D.**

Department of Electronics and Communication Engineering  
 <https://orcid.org/0000-0002-5670-3229>  
 Dr. B.C. Roy Engineering College Durgapur, India  
<https://bcrc.ac.in>

# Utilization of Reconfigurable Intelligent Surfaces with Context Information – Use Cases

Łukasz Kułacz

*Poznan University of Technology, Poznan, Poland*

<https://doi.org/10.26636/jtit.2023.3.1312>

**Abstract** — Utilization of reconfigurable intelligent surfaces is a topic that is very interesting in the context of complex radio environments, especially those used in dense urban areas. By relying on simple solutions to control the angle of reflection of the signal from the surface, different effects may be achieved in radio communication systems. Maximization or minimization of the received power at specific locations near the reflecting surface is the most important effect. This phenomenon allows to: receive a signal at a location at which it could not be received previously, detect spectrum occupancy at a place where the sensor could not perform a correct detection, or minimize interference affecting a specific receiver. In this paper, all three concepts are presented and a simple ray tracing simulation is used to show the potential profits attainable in each of the scenarios. Additionally, a scenario was analyzed in which several of the aforementioned situations are combined.

**Keywords** — context information, reconfigurable intelligent surfaces, spectrum occupancy detection

## 1. Introduction

The development of wireless communication has been based, recently, on an ever-greater level of intelligence of the networks. Dynamic spectrum access systems may serve as a good example here which often require access to a large amount of contextual information about the environment and network users. This information is necessary to make optimal decisions – based on many parameters (not necessarily directly related to the network itself) and their historical values. Data from the past allows, among other things, to capture trends in the behavior of network users and thus enables to predict upcoming events.

The urban environment has always been highly demanding for designers of wireless systems, where the transmitted signal itself is subjected to many propagation-related phenomena, i.e., reflections or dispersion. The huge number and the great variety of buildings and objects affecting the propagation of the signal collectively create a highly complex radio channel. Additionally, in an urban environment, a large number of users is present in a relatively small area. The combination of those two factors poses a significant challenge for the operation of wireless systems. It is worth noting that, in general, signal reflections are unavoidable. While they hinder the analysis

of signal propagation, they generally constitute a beneficial phenomenon, since the signal can reach a receiver which is not located in the transmitter's line of sight. The notion of reconfigurable intelligent surfaces (RIS) [1] is a promising approach that has recently gained much popularity. RIS allows to influence the process of radio wave reflection on the surface involved. This means that it is possible to modify the basic principle where the angle of incidence of the wave is equal to the angle of its reflection. This offers numerous opportunities for creating and controlling the propagation environment. On the one hand, it allows to direct the radio signal to previously unattainable places (due to the geometry of the objects), while on the other it enables to remove radio signal from specific locations (to minimize interference) by directing it in a completely different direction [2].

These two concepts alone seem to offer great potential for use in intelligent networks. They can be relied upon to detect the signal or improve the quality of the received signal by increasing or reducing the received power, in the case of desirable and unwanted signals, respectively. Signals may be detected even if the receiver is not located along the transmitter's line of sight if the power of the reflected signals is sufficient for the operation of signal detection algorithms. However, the solutions described require the development of new algorithms whose task is to control the described RIS [3].

## 2. Scenario Description

In this work, we consider a simple scenario of an elongated room with a wall between the signal transmitter and three receivers, obscuring direct visibility. At the far end of the room, however, there is a surface that will simulate an RIS. All devices are positioned at the same height, and the effects of the ceiling and floor have been ignored to simplify the scenario that is considered to involve a two-dimensional space only. The layout of the investigated area is presented in Fig. 1. The described room was chosen on purpose to demonstrate the potential of using RIS. This decision is dictated by the desire to highlight the concept of using RIS, and the ideas themselves can be used in more complex analyses where, however, it would be much more challenging to isolate the

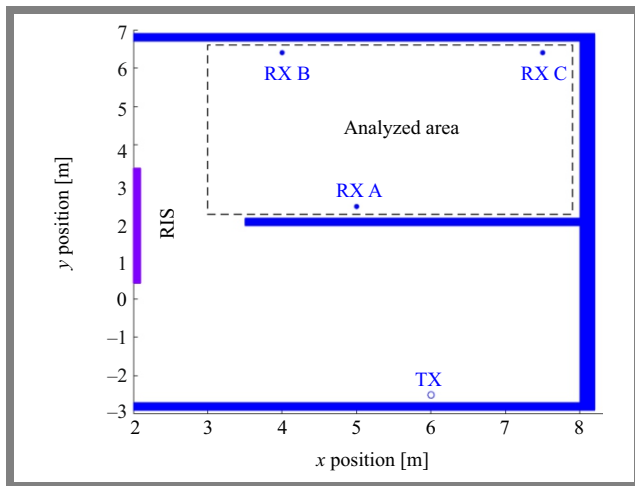


Fig. 1. Area under consideration.

impact of RIS. The work described in this paper compares three simulation scenarios:

- Scenario 1 assumes no use of RIS that has been replaced by a fragment of a regular wall – this is a benchmark for the remaining scenarios.
- Scenario 2 assumes that a simplified RIS is used that operates independently of external factors, i.e., periodically and alternately changes the angle of the reflected signal, simulating the use of a mirror rotated from left to right and back. This scenario has been proposed to demonstrate the potential profits from using even such a simple control mechanism.
- In Scenario 3, network users have access to information about the current RIS setting and indirectly influence those settings. In this example, context information (the current RIS setting) is used to achieve the current goal, i.e. to minimize interference.

Three use cases based on those three scenarios are analyzed in the paper: improved spectrum occupancy detection by the sensor, improvement in the received signal level for two receivers, and reduction of interference at the receiver. During the simulation, the operation of RIS was simulated as a rotational change in the position of the selected surface, with its angle varied from  $-20^\circ$  to  $20^\circ$ , with a step of  $5^\circ$ .

### 3. Simulation Results

Figure 2 presents the spatial distribution of the received power in the analyzed area for the given transmitting parameters and uses a propagation model based on ray tracing. Preliminary analysis of the indicated distribution shows that a certain part of the area under consideration is characterized by lower signal attenuation (warmer colors, receiver C). In comparison, the remaining part of the area has greater signal attenuation – darker colors, receivers A and B – remaining “in the shadow” of the signal. The received power in this situation results directly from the perpendicular arrangement of the wall marked as “RIS” in Fig. 1 which, however, acts as a regular wall here. It can be seen that receivers A and B would have a problem

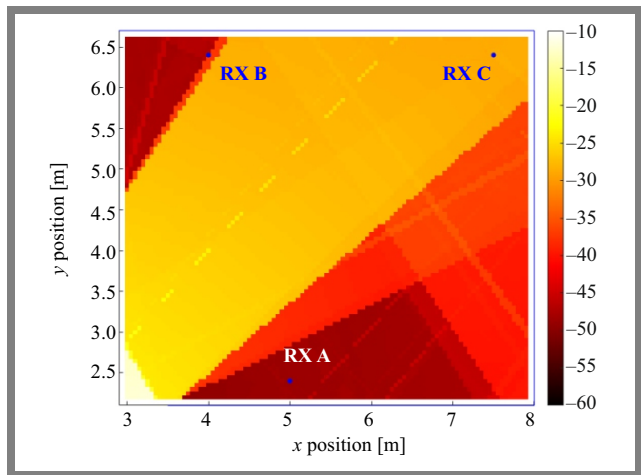


Fig. 2. Received power without influence of RIS.

with receiving and detecting the analyzed signal. However, they would be deprived of excessive interference. The situation is exactly the opposite for receiver C. Nevertheless, it should be remembered that in the opposite situation, when receivers A and B consider the analyzed signal undesirable and receiver C deems it a desirable signal, it would not be beneficial to influence the current propagation. Figure 3 shows the received power of all three analyzed receivers, depending on all RIS settings under consideration. It shows all received power values for all receivers and will be used depending on the current role of a particular receiver.

#### 3.1. Case 1 - Spectrum Occupancy Detection

For the first use case, i.e. the one in which the spectrum occupancy detection processes performed by receiver B are to be improved, preliminary simulations showed that in the first scenario, correct detection is challenging. In the second scenario, correct detection is possible for 31.25% of the time, and in the third scenario, for 79.17% of time there is a possibility of correct detection. The received power in the second scenario (depending on the RIS angle) is presented in Fig. 3 – see RX A. The last scenario depends, to a large

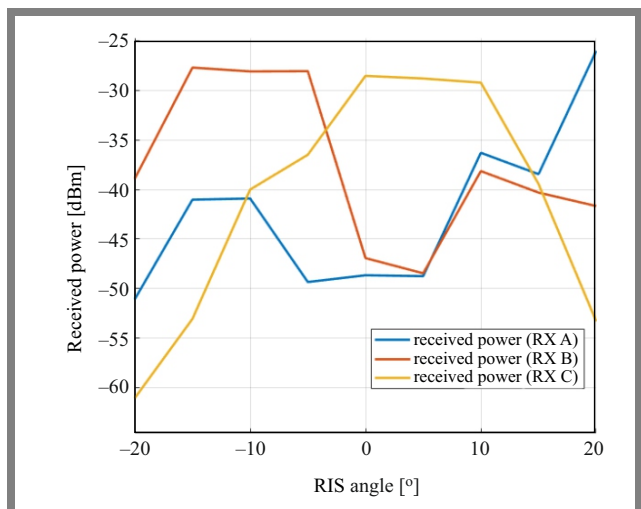


Fig. 3. Received power in receivers versus RIS setting.

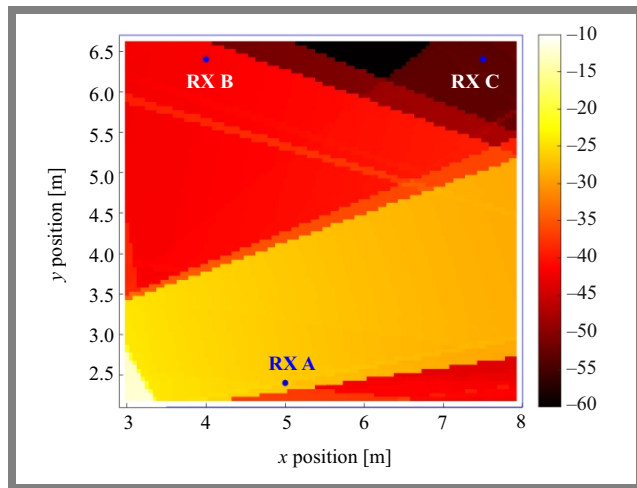


**Fig. 4.** Received power map for a specific RIS setting (angle  $20^\circ$ ).

extent, on the algorithm used and its purpose. But even if only information about the current RIS setting is available, this allows the sensor to assume that the occupancy decision is updated only at a particular time.

This approach is not ideal, and its effectiveness depends very much on the transmission characteristics (frequency, length of transmission, etc.). The solution, which assumes the ability to control the RIS even in a very simple scenario, enables, for example, periodic “combing” of all available settings and then maintaining one setting (or several alternating settings) for a certain period of time. In the considered system with one sensor, it is possible to correctly set the RIS for the entire observation period after a short time, achieving almost 100% of the detection time available. Figure 4 presents the received power map with the best RIS setting for receiver A which is considered to be a spectrum detection sensor in this particular use case.

### 3.2. Case 2 - Received Signal Level Improvement

For the second use case, i.e., improving the received signal level for receivers B and C, the preliminary results of the simulation show that in the first scenario, receiver B cannot receive the signal correctly, while receiver C can receive the signal all the time. In the second scenario, for 35.29% of the time, receiver B receives a signal with a sufficient power level, while receiver C receives it for 41.18% of the time. In this case, we notice the advantage stemming from the ability to service the receiver that was in the “shadow” of the signal and could not receive the signal without changing its position. In the third scenario, however, it was assumed that for an initial period of time, RIS checks all available settings and then, based on reports from the receivers, alternately uses the best-reported settings. As a result, receiver B could receive the signal for 45.1% of the time and receiver C for 47.06% of the time.

At this point, it is worth mentioning traffic steering algorithms used to adjust the duration of settings for individual receivers. For example, in Fig. 5, a received power map with the best RIS setting for receiver B is shown.

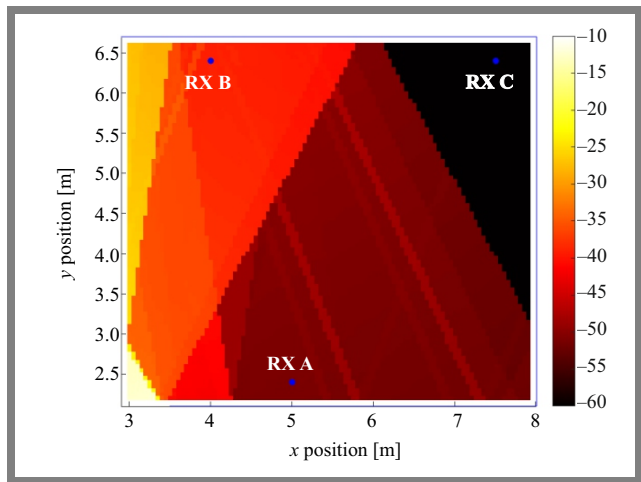


**Fig. 5.** Received power map for a specific RIS setting (angle  $-15^\circ$ ).

### 3.3. Case 3 - Interference Reduction

For the third use case, i.e., interference reduction in receiver C, preliminary simulation results show that in the first scenario, the receiver encounters significant interference for 100% of the time. In the second scenario, the interference exceeds the threshold 35.29% of the time. It is worth mentioning that the average observed interference level decreases by approx. 3.715 dB, the median by approx. 7.972 dB, and the 10th percentile by approx. 24.762 dB. The 90th percentile remains the same. The knowledge of the RIS setting could at least provide information about the occurrence of potential interference - which can, for example, be used in the transmission planning process. In the third scenario, we once again assume a specified period of checking all RIS settings and then, based on the report from receiver C, set up a configuration that causes minimum interference. Thanks to such an approach, it was possible to reduce the time with interference above the threshold value from 100% to approx. 19.61%.

This time, the outcomes are of course hugely dependent on a given configuration and the presence of other devices. Figure 6 presents the received power map with the best RIS setting for receiver C treating the analyzed signal as interference. It should be remembered that the simulations assumed a constant transmission throughout the entire experiment. Temporary signal transmissions for all three use cases would be an additional challenge for the algorithm that controls the operation of the RIS.

### 3.4. All Use Cases Combined

An additional experiment was carried out combining all three use cases analyzed, i.e. receiver A as a spectrum occupancy detection sensor, receiver B as a receiver of the analyzed signal (desired signal), and receiver C as not a receiver of the analyzed signal (unwanted signal). It needs to be borne in mind that in the benchmark scenario, the objectives of none of the receivers are met. Then, RIS operating according to the following scheme was used: initially, all possible RIS configurations are checked, and then all the best settings reported by the receivers are used alternately.

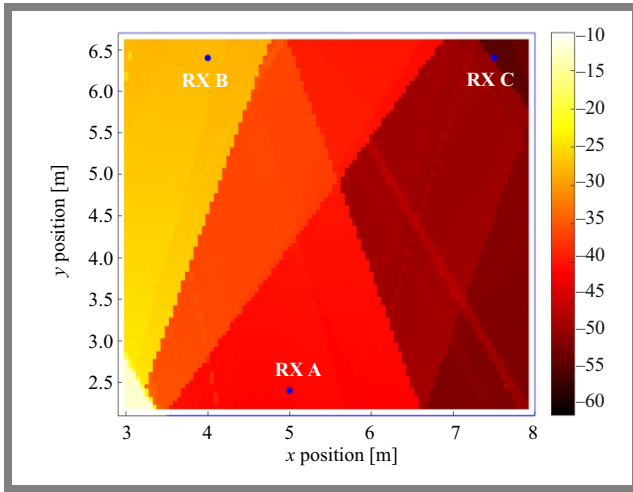


Fig. 6. Received power map for a specific RIS setting (angle  $-20^\circ$ ).

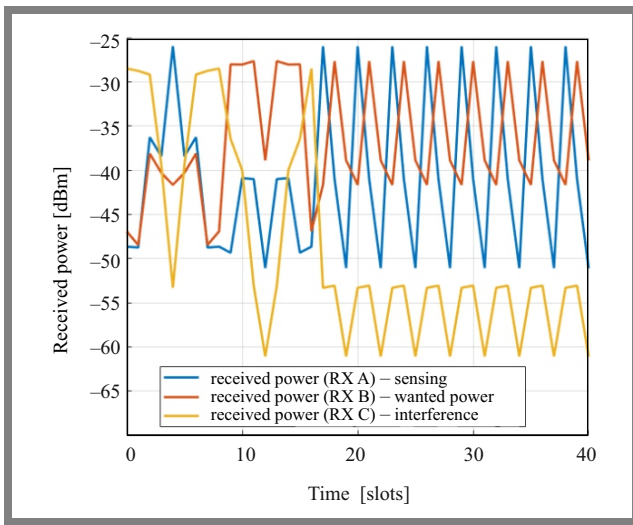


Fig. 7. Received power for all receivers in the time domain.

As a result of preliminary simulations, receiver A can successfully detect the signal for approx. 43.4% of the time; receiver B can successfully receive the desired signal for approx. 45.28% of the time, and receiver C is free of significant interference for approx. 79.25% of the time. In addition, the average interference power observed at receiver C decreases by approx. 8.58 dB. The received power in the time domain for each receiver is shown in Fig. 7.

## 4. Conclusions

The simulations presented in the paper show that using RIS, even with the most straightforward control mechanism, can

benefit each of the analyzed use cases. However, much better results can be obtained by adding a simple algorithm to the RIS control mechanism and providing the interested receivers with information about the current RIS setting. Such an approach offers a very flexible solution that should be carefully considered. In particular, the nature of the environment, the nature of the signal source or sources, potential signal receivers, their positions, and the intended use should be taken into account. On the one hand, the presented concepts allow to direct the signal to previously inaccessible locations. However, this situation may bring about both positive and negative outcomes, because there is a risk of introducing interference at a location where it has not been occurring so far. This problem is exacerbated in the case of devices that are unaware of the existence and presence of RIS. Nevertheless, the ideas presented in this paper seem promising in terms of potential gains for specific applications.

## Acknowledgments

The presented work has been funded by the National Science Centre in Poland under the project (no. 2021/43/B/ST7/01365) of the OPUS program.

## References

- [1] Y. Liu *et al.*, “Reconfigurable Intelligent Surfaces: Principles and Opportunities”, *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021 (<https://doi.org/10.1109/COMST.2021.3077737>).
- [2] M. Di Renzo, F.H. Danufane, and S. Tretyakov, “Communication Models for Reconfigurable Intelligent Surfaces: From Surface Electromagnetics to Wireless Networks Optimization”, *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1164–1209, 2022 (<https://doi.org/10.1109/JPROC.2022.3195536>).
- [3] C. Liaskos *et al.*, “A New Wireless Communication Paradigm through Software-Controlled Metasurfaces”, *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018 (<https://doi.org/10.1109/MCOM.2018.1700659>).
- [4] M. Hoyhtya *et al.*, “Spectrum Occupancy Measurements: A Survey and Use of Interference Maps”, *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386–2414, 2016 (<https://doi.org/10.1109/COMST.2016.2559525>).

**Łukasz Kulacz, Ph.D.**

Institute of Radiocommunications, Faculty of Computing and Telecommunications

 <https://orcid.org/0000-0002-3434-1917>

E-mail: lukasz.kulacz@put.poznan.pl

Poznan University of Technology, Poznan, Poland

<https://cat.put.poznan.pl>

# Minimized Group Delay FIR Low Pass Filter Design Using Modified Differential Search Algorithm

Sonelal Prajapati<sup>1</sup>, Sanjeev Rai<sup>1</sup>, Manish Tiwari<sup>1</sup>, and Atul Kumar Dwivedi<sup>2</sup>

<sup>1</sup>Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India,

<sup>2</sup>Bundelkhand Institute of Engineering and Technology, Jhansi, India

<https://doi.org/10.26636/jtit.2023.3.1313>

**Abstract** — Designing a finite impulse response (FIR) filter with minimal group delay has proven to be a difficult task. Many research studies have focused on reducing pass band and stop band ripples in FIR filter design, often overlooking the optimization of group delay. While some works have considered group delay reduction, their approaches were not optimal. Consequently, the achievement of an optimal design for a filter with a low group delay value still remains a challenge. In this work, a modified differential search optimization algorithm has been used for the purpose of designing a minimal group delay FIR filter. The results obtained have been compared with the classical techniques and they turned out to be promising.

**Keywords** — differential search optimization algorithm, FIR filter, optimization, small group delay

## 1. Introduction

Recently, digital finite impulse response (FIR) filters have caught the interest of researchers in a range of fields, including seismic and satellite signal processing, pattern recognition, biomedical signal processing and more. This is primarily because of their desirable characteristics, such as linear phase, inherent stability, suitability for multi-rate applications, and ease of implementation. FIR filters are known for their stable nature, low sensitivity, and ability to maintain a linear phase [1].

However, one common drawback of FIR filters is that they require higher filter orders to meet specific requirements, thus resulting in increased memory usage and processing delays. To ensure effective and reliable operation of FIR filters, a well-defined design procedure is necessary. Various approaches have been proposed to generate coefficient sets that aim to meet the desired specifications. Design engineers face such challenges as passband ripples (PBR), stopband ripples (SBR), and find it difficult to achieve sharper cutoffs with minimal filter orders. Apart from frequency domain specifications, a properly designed filter should also minimize group delays, as their high values lead to increased power consumption.

The motivation behind this study is to design filters that not only meet the desired frequency specifications, but also pro-

vide minimal delay once implemented in their hardware form. Group delay is impacted by the electrical length of the digital signal processing (DSP) system and can cause distortion of the output signal. Linear phase FIR filters offer constant group delay in the passband, if designed symmetrically. However, a major disadvantage is that they provide an overall group delay of  $N/2$ , where  $N$  represents the filter order. This large group delay reduces the speed of processing the signal by the filters, thus necessitating the mitigation of group delay in order to speed up signal processing.

A group delay reduction approach based on multirate techniques was proposed for active noise control systems in [2]. This approach utilizes polyphase decomposition and requires the design of multiple filters and multirate signal processing, consequently increasing design complexity and hardware requirements.

Other methods have explored small group delay for FIR filters using minimax constrained optimization, convex optimization based on semi-definite programming, quasi-convex optimization using second-order cone optimization, and minimax group delay reduction with the implementation in all-pass filters. However, the optimization approaches employed in these methods were not globally optimal, and the considered examples of filters were characterized by high orders, thus requiring significant power and chip area for implementation. Designing a minimum group delay filter using a global optimization approach still remains a rather challenging task.

Previously, evolutionary optimization algorithms, such as the genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), artificial bee colony (ABC), bacteria foraging algorithm (BFO), cat swarm optimization (CSO), differential evolution (DE), and orthogonal harmony search (OHS) have been utilized for solving complex, nonlinear, multidimensional FIR filter design problems. However, none of these approaches specifically considered the design of a low group delay FIR filter.

In this work, the optimization-based function for group delay reduction is formulated. The conventional differential search algorithm (DSA), bijective differential algorithm (BDSA), and the proposed differential search algorithm (mDSA) are employed to solve the problem of optimizing the design of a minimal group delay FIR filter. The effectiveness of the

proposed approach is evaluated by comparing the designed filters with existing methods.

The article is structured as follows. Section 2 presents the mathematical formulation of the problem, specifically focusing on the design of a minimum group delay FIR filter. This formulation enables the application of the proposed algorithm called mDSA. Section 3 contains a description of specific optimization algorithms, such as DSA, BDSA, and mDSA. In Section 4, the results obtained upon implementing DSA, BDSA, and mDSA are discussed. Finally, conclusions concerning the techniques proposed in the article are presented in Section 5.

## 2. Model of a Minimum Group Delay FIR Filter

The output  $y(t)$  of a signal processing system whose impulse response is  $h(t)$ , for an input  $x(t)$  is given by:

$$y(t) = (h * x)(t) = \int_{-\infty}^{\infty} x(d) h(t - \tau) d\tau, \quad (1)$$

where  $\tau$  is the unit sample delay in time domain. In the frequency analysis, the same can be represented as:

$$Y(s) = [H(s)] [X(s)], \quad (2)$$

where

$$X(s) = \mathcal{L}\{x(t)\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x(t) e^{-st} dt, \quad (3)$$

and

$$H(s) = \mathcal{L}\{h(t)\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} h(t) e^{-st} dt, \quad (4)$$

where  $h(t)$  is the time-domain impulse response of the LTI system,  $X(s)$ ,  $Y(s)$ ,  $H(s)$  are the Laplace transforms of the input  $x(t)$ , output  $y(t)$ , and impulse response  $h(t)$ , respectively.  $H(s)$  is called the transfer function of the LTI system and, like the impulse response  $h(t)$ , fully defines the input-output characteristics of the LTI system.

The main objective of this work is to formulate an objective function for optimizing the design of a low group delay FIR filter. The main requirement of an appropriately designed filter, as formulated in this work, is lower group delay with minimum pass band ripple, stop band ripples with smaller transition width. The objective function has been framed as deviation from the desired filter response to the designed filter response. The deviation of pass band ripple is:

$$f_1(b) = |(\max[e(\omega)] - \delta_{p_m}) - 1| \quad \text{for } \omega \leq \omega_p, \quad (5)$$

where  $\delta_{p_m}$  is the maximum pass band ripple,  $\omega_p$  is the pass-band edge frequency,  $e(\omega)$  is the error difference between the desired filter response  $B_d(\omega)$  and the actual filter response  $B(\omega)$  which is written as:

$$e(\omega) = [B_d(\omega) - B(\omega)]. \quad (6)$$

The second part of the objective function is the deviation from the desired filter response to the designed filter response of stop band ripple which can be written as:

$$f_2(b) = |(\max e(\omega) - \delta_{s_m})| \quad \text{for } \omega \geq \omega_s, \quad (7)$$

where  $\delta_{s_m}$  is the maximum stop-band ripple,  $\omega_s$  is the stop-band edge frequency. The frequency response  $B(\omega)$  of a linear phase designed FIR filter design is:

$$B(\omega) = \sum_{m=0}^{M-1} b(m) e^{-j(m\omega)}, \quad (8)$$

where  $b(m) = h(m)$  for  $m = 0, 1, \dots, M-1$ ,  $h(m)$  is the impulse response, and  $m$  is an integer. The third part of the objective function is formulated as group delay. It is calculated as a derivative of phase with respect to angular frequency  $\omega$ . The group delay response of a filter is a measure of the average delay of the filter as a function of frequency. It is the negative first derivative of the phase response of the filter. If the frequency response of a filter is  $H(e^{j\omega})$ , then the group delay is:

$$\tau_g(\omega) = -\frac{d\phi(\omega)}{d\omega}, \quad (9)$$

where  $\phi(\omega)$  is the phase, or argument, of  $H(e^{j\omega})$ . Group delay is considered to be the third part of the objective function for optimization:

$$f_3(b) = \tau_g(\omega). \quad (10)$$

The overall objective function optimization is:

$$f = f_1 + f_2 + f_3. \quad (11)$$

The optimization problem given by Eq. (11) has been solved using DSA, BDSA and the proposed mDSA optimization algorithm discussed in Section 3.

## 3. Differential Search Algorithm

DSA is a well-known population-based meta-heuristic optimization algorithm that is based on the migration behavior of superorganisms [3]. The availability of food resources varies throughout the year due to varying climate conditions. As a result, many live species move. Migration behavior permits living organisms to move from an inefficient, food-limited habitat to a productive, food-rich environment. Ants and birds, for example, follow a regular migration pattern. During migration, species develop superorganisms made up of several individuals. Superorganisms start relocating to more productive locations to suit their needs. They can linger at the new place for a while before moving on to more productive areas. In DSA, an artificial-superorganism is a population that consists of arbitrary solutions to the corresponding problem. During migration, an artificial superorganism decides where to stop. Individuals of the superorganisms that discovered it would immediately reside there and would continue their migration to the global minimum value.

In DSA, the artificial organism can be represented as  $X_i, i = 1, 2, \dots, N_{pop}$  and the superorganism made up of these artificial organisms is defined as  $superorganism_g = [X_i]$  where  $g = 1, 2, \dots, G$  contains the members according to the size of the problem such that  $x_{ij}, j = 1, 2, \dots, D$  where  $N_{pop}$  represents the number of individuals, and  $D$  signifies the size of the problem, while  $G$  is the maximum generation or iteration number. In such case, the operation of DSA can be described by the following steps:

**1. Initialization of artificial-organism.** The first step is to initialize each member of the artificial-organism to its random position such that:

$$x_{ij} = lb_j + \text{rand}[0, 1] \times (ub_j - lb_j), \quad (12)$$

where  $\text{rand}[0, 1]$  is the random value between 0 and 1 and  $ub_j, lb_j$  are the upper bounds and lower bounds of  $j$ -th dimension of the respective problem, respectively.

**2. Initialization of control parameters.** The DSA algorithm has only two control parameters  $p_1$  and  $p_2$  which are used to determine the number of disturbances corresponding to the position of each member of the artificial-organism. Scale is used to determine the magnitude of change in the position of an individual member of the artificial organism. The value of scale is generated by gamma random number generator  $\text{rand}_g$  so the values of  $p_1, p_2$ , and  $\text{scale}$  can be calculated as:

$$p_1 = 0.3 \times \text{rand}_1 \text{ and } p_2 = 0.3 \times \text{rand}_2, \quad (13)$$

$$\text{scale} = \text{rand}_g \times (2 \times \text{rand}_3) \times \text{rand}_4 - \text{rand}_5, \quad (14)$$

where  $\text{rand}_1, \text{rand}_2, \text{rand}_3, \text{rand}_4$ , and  $\text{rand}_5$  are the uniform random number generators. In this way, the artificial organism may radically change its direction within the habitat.

**3. Compute the stop-over site.** During the migration process, the stop-over site is a randomly shuffling intermediate position assumed by the artificial organism based on the Brownian-like random walk model. In order to calculate the stop-over site, the algorithm creates stop-over vector  $S_{i,G}$  which can be calculated as:

$$S_{i,G} = \text{superorganism}(X_{i,G}) + \text{scale} \times (\text{superorganism}(X_{i,G}) - \text{donor}), \quad (15)$$

where  $\text{donor}$  refers to the generation of the direction matrix based on the random shuffling of individual members of the artificial organism moving towards the target, such that:

$$\text{donor} = X_{\text{random\_shuffling}(i)}. \quad (16)$$

**4. Selection of next generation.** The selection operator  $G = G + 1$  is used to calculate the next generation between the artificial organism population and the stop-over vector population. This selection is based on the fitness value:

$$S_{i,G+1} = \begin{cases} S_{i,G}, & \text{if } f(S_{i,G}) \leq f(X_{i,G}) \\ X_{i,G}, & \text{if } f(S_{i,G}) > f(X_{i,G}) \end{cases}. \quad (17)$$

**5. Check the termination criterion.** If the stop-over site is more fruitful then the source owned by the artificial organism, then the artificial organism shifts to the newly discovered position. However, the superorganisms continue to migrate until the global optimum (best solution) is reached.

### 3.1. Bijective Differential Search Algorithm (BDSA)

The bijective differential search algorithm has been explored by Civicioglu [3] based on stochastic permutation of the original population of the algorithm. It is different from the original differential search algorithm in terms of scale factor  $R$  and the section method. In BDSA, the scale factor is

defined with the help of the stochastic walk of the population replacing the Brownian walk relied upon in the differential search algorithm.

### 3.2. Modified Differential Search Algorithm

The DSA and BDSA do not guarantee to find global optima in all types of situations, and the rate of exploitation at the local optima is relatively modest. Therefore, in this work, a modified differential algorithm has been proposed which utilizes a quantum-based selection operator. In the quantum-inspired differential search algorithm [4], the superorganism  $|\psi\rangle$  is represented by a pair of complex numbers, as a vector  $[\alpha\beta]^T$ :

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (18)$$

where  $\alpha$  and  $\beta$  are probabilities with the relation between them being  $\alpha^2 + \beta^2 = 1$ , while  $|0\rangle$  and  $|1\rangle$  are qbit (i.e. representation of a quantum bit). The superorganism of length  $n$  is represented by a qbit.

$$q = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} \begin{bmatrix} \dots \\ \dots \end{bmatrix} \begin{bmatrix} \alpha_n \\ \beta_n \end{bmatrix}. \quad (19)$$

One example of a superorganism of this type, having two qbits, can be represented as:

$$q = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}. \quad (20)$$

In quantum differential equations, there are NP individuals and each individual is composed of  $D$  qbits. Because the probability of states of every qbit satisfies the relationship of  $\alpha^2 + \beta^2 = 1$ , qbits can also be represented by  $\sin \theta$  and  $\cos \theta$ . After generation of the superorganism, other control parameters are initialized as real numbers, which are later on converted to qbits. Now, the stop over site is also calculated with the help of the quantum mechanism. In this operation, diversity of the population is maintained by the donor and it is increased because of the quantum-based representation of population. The Brownian motion of the stop over site is more diverse because of the quantum representation. The next generation is selected with the help of qbits represented by  $\sin \theta$  and  $\cos \theta$ . Therefore, Eq. (17) can be modified as:

$$\theta_{i,G+1} = \begin{cases} \theta_{i,G}, & \text{if } f(\theta_{i,G}) \leq f(\theta_{i,G}) \\ \theta_{i,G}, & \text{if } f(\theta_{i,G}) > f(\theta_{i,G}) \end{cases}. \quad (21)$$

These quantum operators increase the diversity of the population, thus expanding the algorithm's search space.

## 4. Results and Implications

In this work, the proposed technique has been tested and evaluated using Matlab. The performance of the proposed technique has been evaluated under three scenarios. In the first case, it has been tested in terms of standard benchmark functions. In the second case, it has been tested in connection with designing a low pass FIR filter. For each algorithm, the population size has been taken as 100. Each algorithm is executed over 500 iterations. The algorithm's parameters

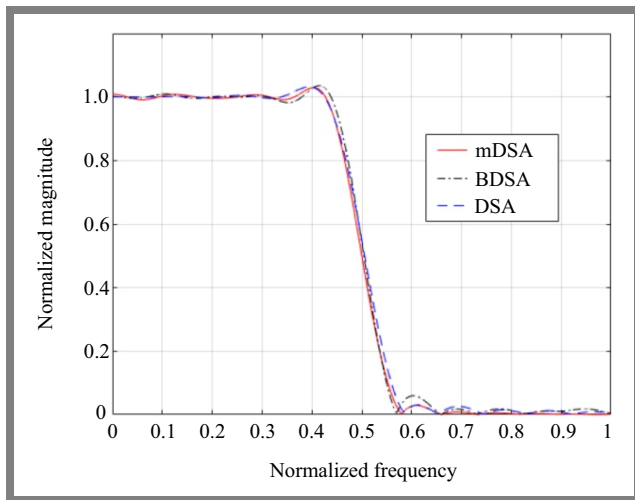


Fig. 1. Normalized magnitude response of the FIR filter.

have been set after number of trials. In order to take into account randomization of the optimization algorithm, 50 trials have been performed for each algorithm, with the minimum and maximum values of the parameters being calculated with a standard deviation. The statistical hypothesis test is performed using the  $t$  test to evaluate the significance of the proposed technique.

#### 4.1. Case 1 – Benchmark Function Evaluation

For evaluating the impact of the proposed technique on benchmark functions, five multimodal, non-linear benchmark functions are considered (Tab. 1). The benchmark functions have been tested at 30 dimensions. The DSA, BDSA and the proposed mDSA approaches have been applied to these selected benchmark equations.

The best and normalized values of the benchmark functions are presented in Tab. 2, where optimized results are shown, after normalization. The value in each row 50 is the representation of the minimum value and other values are scaled corresponding to this value. Thus, the proposed mDSA is capable of finding optimum results while performing standard benchmark functions. DSA performs better than BDSA for

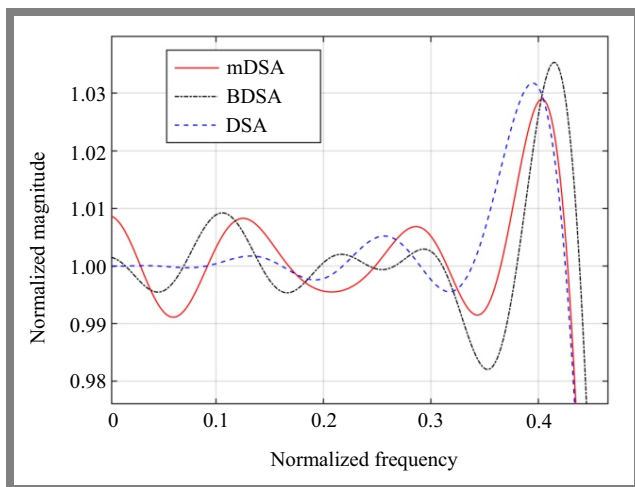


Fig. 2. Normalized PBR response of the FIR filter.

Tab. 1. Benchmark functions.

| No.   | Function        | Search range     | Optimized value (min. at) |
|-------|-----------------|------------------|---------------------------|
| $F_1$ | Griewank        | -512 ... 512     | $f(0, \dots, 0) = 0$      |
| $F_2$ | Ackley          | -32 ... 32       | $f(0, \dots, 0) = 0$      |
| $F_3$ | Rastrigin       | -5.12 ... 5.12   | $f(0, \dots, 0) = 0$      |
| $F_4$ | Schwefel's 2.21 | -100 ... 100     | $f(0, \dots, 0) = 0$      |
| $F_5$ | Whitley         | -10.24 ... 10.24 | $f(1, 1, \dots, 1) = 0$   |

Tab. 2. The best results of benchmark functions.

| Benchmark function | DSA | BDSA | Proposed mDSA |
|--------------------|-----|------|---------------|
| $F_1$ (Griewank)   | 129 | 76   | 50            |
| $F_2$ (Ackley)     | 156 | 121  | 50            |
| $F_3$ (Rastrigin)  | 142 | 124  | 50            |
| $F_4$ (Schwefel's) | 130 | 154  | 50            |
| $F_5$ (Whiteley)   | 80  | 164  | 50            |

Schwefel 2.21 and Whiteley functions. However, for Ackley, Griewank and Rastrigin functions, BDSA performs better than DSA. Therefore, it may be concluded that the proposed quantum-inspired differential search algorithm finds a better solution.

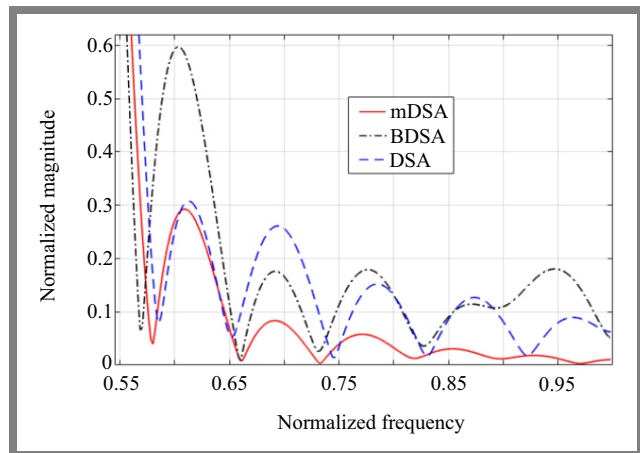


Fig. 3. Normalized SBR response of the FIR filter.

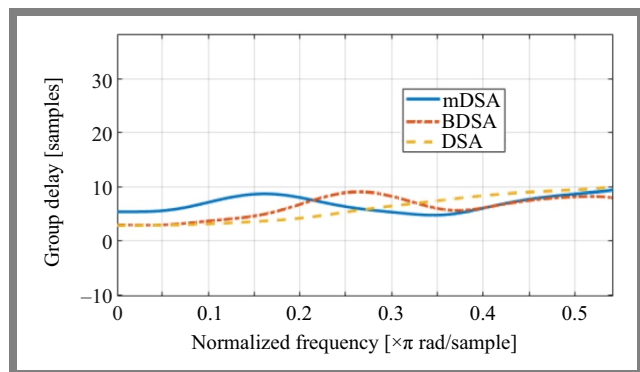


Fig. 4. Normalized group delay response of the FIR filter.

**Tab. 3.** Specifications for the desired FIR filter.

| Parameter                             | Value |
|---------------------------------------|-------|
| Order of designed filter              | 20.00 |
| Pass-band edge frequency (normalized) | 0.45  |
| Stop-band edge frequency (normalized) | 0.55  |
| Normalized transition width           | 0.10  |
| Normalized maximum pass band ripple   | 0.01  |
| Normalized maximum stop band ripple   | 0.01  |
| Maximum group delay                   | 0.00  |

**Tab. 4.** The normalized pass band ripple in FIR low pass filter.

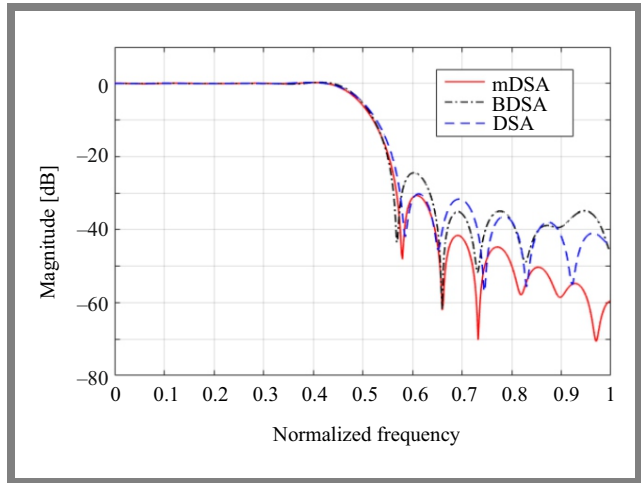
| Algorithm     | Max.    | Min.    | Avg.    | Std. dev. |
|---------------|---------|---------|---------|-----------|
| GA [7]        | 0.11926 | 0.01007 | 0.01932 | 0.00722   |
| PSO [8]       | 0.11740 | 0.08230 | 0.10071 | 0.00789   |
| ABC [8]       | 0.16620 | 0.09970 | 0.12800 | 0.02380   |
| MFO [8]       | 0.15170 | 0.10210 | 0.12580 | 0.01900   |
| BFO [9]       | 0.12900 | -       | -       | -         |
| CSO [6]       | 0.16400 | -       | -       | -         |
| DE [10]       | 0.04000 | -       | -       | -         |
| OHS [6]       | 0.14000 | -       | -       | -         |
| IOT [8]       | 0.11630 | 0.08890 | 0.10515 | 0.00535   |
| DSA           | 0.03613 | 0.03177 | 0.03289 | 0.00079   |
| BDSA          | 0.03536 | 0.03354 | 0.03412 | 0.00031   |
| Proposed mDSA | 0.02786 | 0.02634 | 0.02653 | 0.00021   |

**Tab. 5.** The normalized stop band ripple in FIR low pass filter.

| Algorithm     | Max.    | Min.    | Avg.     | Std. dev. |
|---------------|---------|---------|----------|-----------|
| GA [7]        | 0.04521 | 0.01250 | 0.02131  | 0.00681   |
| PSO [8]       | 0.0186  | 0.0122  | 0.014877 | 0.001104  |
| ABC [8]       | 0.1189  | 0.0825  | 0.09483  | 0.08711   |
| MFO [8]       | 0.02773 | -       | -        | -         |
| BFO [9]       | 0.02773 | -       | -        | -         |
| CSO [6]       | 0.01980 | -       | -        | -         |
| DE [10]       | 0.03900 | -       | -        | -         |
| OHS [6]       | 0.1746  | -       | -        | -         |
| IOT [8]       | 0.0175  | 0.0131  | 0.01417  | 0.00093   |
| DSA           | 0.02873 | 0.02631 | 0.02756  | 0.00092   |
| BDSA          | 0.05891 | 0.04376 | 0.05623  | 0.00090   |
| Proposed mDSA | 0.02781 | 0.02667 | 0.02730  | 0.00081   |

#### 4.2. Case 2 – Low Pass FIR Filter Design

In this work, in order to check the effectiveness of the proposed technique, a low group delay filter has been designed

**Fig. 5.** Magnitude response of the FIR filter.

by applying DSA, BDSA and the proposed optimization techniques. The upper limit of the solution is set as +1 and lower limits of filter coefficients have been set as -1. The design parameters are given in Tab. 3. These filter parameters related to pass band cut-off frequency, stop band cut-off frequency, transition width and stop band ripple are the same as shown in [5], [6]. In this work, pass band ripple of 0.1 is considered, being a better value than that achieved in past research.

Such specifications as pass band ripple, stop band ripple, group delay pass band cut-off frequency, stop band cut-off frequency and transition width of the low pass filter designed with the use of the proposed techniques have been compared with the results obtained using DSA, BDSA and other known techniques, i.e. GA [7], PSO [8], ABC [8], MFO [8], BFO [9], CSO [6], DE [10], OHS [6] and IOT [8].

The comparison concerning pass band ripple is shown in Tab. 4. It can be concluded that the proposed algorithm provides best results compared with other state-of-the-art techniques listed in the table, in terms of maximum, average and minimum PBR values (Figs. 1–5).

The normalized stop band ripple (SBR) obtained using the proposed technique is also compared with DSA, BDSA and other evolutionary optimization techniques listed earlier. It can be summarized from Tab. 5 that the proposed technique provides comparable results in terms of maximum, average and minimum stop band ripple (SBR) values.

The importance of the transition width is presented in Tab. 6. The proposed method offers a transition width that is very similar to the desired specifications (see Tab. 3). However, the transition width reported in many previous works that are based on state-of-the-art techniques is violating the assumed limits.

The main objective of this research was to design a low group delay filter and the group of the filters designed using the proposed techniques is shown in Tab. 8. The group delay presented in this table is the average value obtained from 50 iterations of the proposed algorithm.

It can be clearly observed from the Tab. 8 that the proposed mDSA approach outperforms all other optimization-based filter design techniques. In the majority of the previously

**Tab. 6.** Normalized transition width in an FIR low-pass filter.

| Algorithm     | Max.  | Min.  | Avg.  | Std. dev. |
|---------------|-------|-------|-------|-----------|
| GA            | 0.186 | 0.152 | 0.163 | 0.002     |
| PSO [11]      | 0.087 | -     | -     | -         |
| ABC [12]      | 0.102 | -     | -     | -         |
| BFO [9]       | 0.092 | -     | -     | -         |
| CSO [6]       | 0.095 | -     | -     | -         |
| DE [10]       | 0.180 | -     | -     | -         |
| OHS [6]       | 0.099 | -     | -     | -         |
| DSA           | 0.158 | 0.133 | 0.134 | 0.002     |
| BDSA          | 0.131 | 0.122 | 0.120 | 0.001     |
| Proposed mDSA | 0.119 | 0.098 | 0.102 | 0.001     |

reported FIR filter design approaches, group delay is not considered and symmetric coefficients are used which provide a constant group delay of 10 on a normalized scale. There are some approaches for filter design. These approaches do not provide better ripple specifications at lower orders in the range of 20.

## 5. Conclusions

In this work, low group delay FIR filters have been designed. The objective function has been formulated by considering the difference between the designed filter and the desired specification. Classical differential algorithms have been used to solve the filter design problem. Furthermore, the bijective differential search algorithm (BDSA) was also used during the design process. Inspired by modern quantum theory, a modified differential search algorithm (mDSA) has been developed. The improved mDSA algorithm enhanced accuracy and convergence parameters by exploring a larger search space. The results obtained have been compared with those achieved by deploying existing FIR filter design methods. The proposed technique has outperformed all other approaches in terms of group delay and other filter specifications.

## References

- [1] C. Wu, D. Gao, and K.L. Teo, "A Direct Optimization Method for Low Group Delay FIR Filter Design", *Signal Processing*, vol. 93, no. 7, pp. 1764–1772, 2013 (<https://doi.org/10.1016/j.sigpro.2013.01.015>).
- [2] R.M. Rizk-Allah, A.E. Hassanien, M. Elhoseny, and M. Gunasekaran, "A New Binary Salp Swarm Algorithm: Development and Application for Optimization Tasks", *Neural Computing and Applications*, vol. 31, no. 5, pp. 1641–1663, 2019 (<https://doi.org/10.1007/s00521-018-3613-z>).
- [3] P. Civicioglu, "Transforming Geocentric Cartesian Coordinates to Geodetic Coordinates by Using Differential Search Algorithm", *Computers & Geosciences*, vol. 46, pp. 229–247, 2012 (<https://doi.org/10.1016/j.cageo.2011.12.011>).

**Tab. 7.** Group delay in an FIR low-pass filter.

| Algorithm     | Order | Group delay |       |       |           |
|---------------|-------|-------------|-------|-------|-----------|
|               |       | Max.        | Min.  | Avg.  | Std. dev. |
| GA [7]        | 20    | 10          | 10    | 10    | 0.000     |
| PSO [8]       | 20    | 10          | 10    | 10    | 0.000     |
| ABC [8]       | 20    | 10          | 10    | 10    | 0.000     |
| MFO [8]       | 20    | 10          | 10    | 10    | 0.000     |
| IOT [8]       | 20    | 10          | 10    | 10    | 0.000     |
| CO [1]        | 32    | 6.943       | 4.25  | 6.532 | 1.234     |
| MCO [13]      | 24    | 7.1         | 5.25  | 6.173 | 1.411     |
| DSA           | 20    | 9.362       | 8.801 | 9.132 | 0.0029    |
| BDSA          | 20    | 9.044       | 8.743 | 8.892 | 0.0036    |
| Proposed mDSA | 20    | 8.704       | 8.602 | 8.704 | 0.00029   |

**Tab. 8.** Optimized filter coefficients.

| Filter coefficients | Algorithms |        |               |
|---------------------|------------|--------|---------------|
|                     | DSA        | BDSA   | Proposed mDSA |
| C (1.0)             | −0.042     | −0.072 | −0.032        |
| C (2.0)             | −0.101     | −0.150 | −0.095        |
| C (3.0)             | −0.140     | −0.127 | −0.079        |
| C (4.0)             | −0.217     | −0.063 | 0.060         |
| C (5.0)             | −0.344     | −0.165 | 0.119         |
| C (6.0)             | −0.355     | −0.374 | −0.098        |
| C (7.0)             | −0.101     | −0.369 | −0.416        |
| C (8.0)             | 0.216      | −0.045 | −0.459        |
| C (9.0)             | 0.239      | 0.257  | −0.188        |
| C (10)              | −0.021     | 0.229  | 0.049         |
| C (11)              | −0.174     | 0.009  | 0.046         |
| C (12)              | −0.052     | −0.094 | −0.025        |
| C (13)              | 0.092      | −0.028 | 0.020         |
| C (14)              | 0.052      | 0.007  | 0.092         |
| C (15)              | −0.046     | −0.031 | 0.054         |
| C (16)              | −0.035     | −0.028 | −0.022        |
| C (17)              | 0.026      | 0.032  | −0.024        |
| C (18)              | 0.021      | 0.043  | 0.013         |
| C (19)              | −0.017     | −0.004 | 0.008         |
| C (20)              | −0.011     | −0.023 | −0.017        |
| C (21)              | 0.009      | −0.004 | −0.013        |

- [4] W. Deng *et al.*, "An Improved Quantum-inspired Differential Evolution Algorithm for Deep Belief Network", *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319–7327, 2020 (<https://doi.org/10.1109/TIM.2020.2983233>).

- [5] A.K. Dwivedi, S. Ghosh, and N.D. Londhe, "Low Power FIR Filter Design Using Modified Multi-objective Artificial Bee Colony Algorithm", *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 58–69, 2016 (<https://doi.org/10.1016/j.engappai.2016.06.006>).
- [6] S.K. Saha *et al.*, "Efficient and Accurate Optimal Linear Phase FIR Filter Design Using Opposition-based Harmony Search Algorithm", *The Scientific World Journal*, art. no. 320489, 2013 (<https://doi.org/10.1155/2013/320489>).
- [7] A.K. Dwivedi, S. Ghosh, and N.D. Londhe, "Review and Analysis of Evolutionary Optimization-based Techniques for FIR Filter Design", *Circuits, Systems, and Signal Processing*, vol. 37, no. 10, pp. 4409–4430, 2018 (<https://doi.org/10.1007/s00034-018-0772-1>).
- [8] T. Mittal, "Design of Optimal FIR Filters Using Integrated Optimization Technique", *Circuits, Systems, and Signal Processing*, vol. 40, no. 6, pp. 2895–2925, 2021 (<https://doi.org/10.1007/s00034-020-01602-8>).
- [9] S.K. Saha, R. Kar, D. Mandal, and S.P. Ghoshal, "Bacteria Foraging Optimization Algorithm for Optimal FIR Filter Design", *International Journal of Bio-Inspired Computation*, vol. 5, no. 1, pp. 52–66, 2013 (<https://doi.org/10.1504/IJBIC.2013.053039>).
- [10] K.S. Reddy and S.K. Sahoo, "An Approach for FIR Filter Coefficient Optimization Using Differential Evolution Algorithm", *AEU - International Journal of Electronics and Communications*, vol. 69, no. 1, pp. 101–108, 2015 (<https://doi.org/10.1016/j.aeue.2014.07.019>).
- [11] K. Boudjelaba, F. Ros, and D. Chikouche, "Potential of Particle Swarm Optimization and Genetic Algorithms for FIR Filter Design", *Circuits, Systems, and Signal Processing*, vol. 33, no. 10, pp. 3195–3222, 2014 (<https://doi.org/10.1007/s00034-014-9800-y>).
- [12] A. K. Dwivedi, S. Ghosh, and N. D. Londhe, "Modified Artificial Bee Colony Optimization-based FIR Filter Design with Experimental Validation Using Field-programmable Gate Array", *IET Signal Processing*, vol. 10, no. 8, pp. 955–964, 2016 (<https://doi.org/10.1049/iet-spr.2015.0214>).
- [13] J. Konopacki, "Design of Sparse FIR Filters with Low Group Delay", *International Journal of Electronics and Telecommunications*, vol. 67, no.1, pp. 121–126, 2021 (<https://doi.org/10.24425/ijet.2021.135953>).

---

**Sonelal Prajapati, Ph.D.**

 <https://orcid.org/0000-0001-5891-9528>

E-mail: sonelal59@gmail.com

Motilal Nehru National Institute of Technology Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

**Sanjeev Rai, Ph.D.**

E-mail: srai@mnnit.ac.in

Motilal Nehru National Institute of Technology Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

**Manish Tiwari, Ph.D.**

 <https://orcid.org/0000-0002-0361-9373>

E-mail: mtiwari@mnnit.ac.in

Motilal Nehru National Institute of Technology Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

**Atul Kumar Dwivedi, Ph.D.**

 <https://orcid.org/0000-0003-4221-806X>

E-mail: akd@bietjhs.ac.in

Bundelkhand Institute of Engineering and Technology, Jhansi,  
India

<http://bietjhs.ac.in>

# Performance Evaluation of a Zone-based Three-level Heterogeneous Clustering Protocol for WSNs

Sanapala Shanmukha Rao, Shilpi, and Arvind Kumar

Motilal Nehru National Institute of Technology, Allahabad, Prayagraj, India

<https://doi.org/10.26636/jtit.2023.3.1381>

**Abstract** — This paper proposes a zone-based three-level heterogeneous clustering protocol (ZB-TLHCP) for heterogeneous WSNs. In ZB-TLHCP, the sensor field/region is divided into zones where super, advance, and normal nodes are deployed uniformly and randomly. The performance of the proposed ZB-TLHCP system is compared with that of zonal-stable election protocol (Z-SEP), distributed energy efficient clustering (DEEC), and threshold-based DEEC (TDEEC) protocol by varying the number of super and advance nodes, their energy levels for the fixed sensor field, and the total number of nodes. Matlab simulation results revealed that the proposed ZB-TLHCP solution performed better than Z-SEP, DEEC, and TDEEC protocols, as it increased the instability period, prolonged the network's lifetime, and achieved higher throughput values.

**Keywords** — clustering, heterogeneous, network lifetime, stability period, zone-based WSN

## 1. Introduction

Wireless sensor networks (WSNs) comprise numerous self-governing, low-energy, lightweight, densely deployed sensor nodes communicating via the wireless medium [1]. WSNs are widely used in the military, various industries, in environment protection schemes, in multimedia, as well as in underwater, home control/automation, forest fire detection, transportation, logistics, monitoring, parking slot management, biomedical, health and other settings [2], [3]. Due to the dense deployment of sensor nodes, WSNs provide more information and collect data to the mutual collaboration of nodes, thus ensuring more accurate and comprehensive results. The field deployment of sensor nodes defines the communication architecture of sensor networks [4]. These nodes collect information from the physical medium and transmit it to the base station (BS) in

multi-hop or single-hop transmission processes. Sensor node architecture [4] consists of four modules: a power unit, a processing unit, a transceiver unit, and a sensing unit, as shown in Fig. 1. The sensing unit comprises an analog-to-digital converter (ADC) and a MEMS sensor. Sensors sense data in the analog form, and ADC converts it into digital data for further processing. The processing unit consists of a storage block and a computational unit, processing the digital data received from the sensor and store it based on the requirements of a specific application. The transceiver unit either transmits or receives the data through the communication medium. The power unit supplies energy to all the modules. Along with these four modules, there is a GPS-based location-finding system and a mobilizer for establishing mobile communications.

Sensor nodes within the network are deployed in harsh environments, where battery replacement is expensive and human entry is discouraged [4]. This makes the sensor nodes power-constrained, emphasizing the need for efficient energy utilization. The BS is assumed to have a sufficient power supply in any wireless sensor network communication architectures. The clustering technique is an effective solution to reduce the sensor node's energy consumption in WSNs [5] (Fig. 2). The nodes are organized into clusters and each cluster chooses one node to serve as its cluster head (CH). Within the specified TDMA schedule, CH gathers data from cluster nodes. After performing data aggregation and integration, the information is transferred, in a single-hop or with the use of multi-hop communication, to the BS. At the BS, this data is uploaded to the Internet, making it accessible to end-user [5]. The remaining sections are structured as follows. Section 2 discusses related research work concerned with the proposed protocol. Section 3 discusses the implementation steps. Simu-

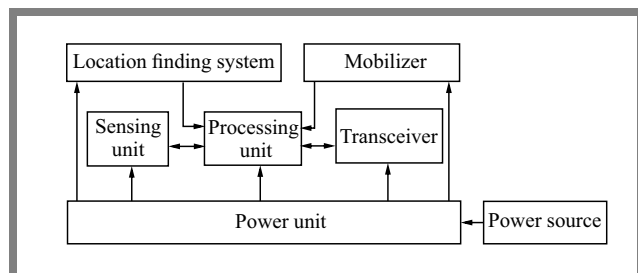


Fig. 1. Sensor node architecture.

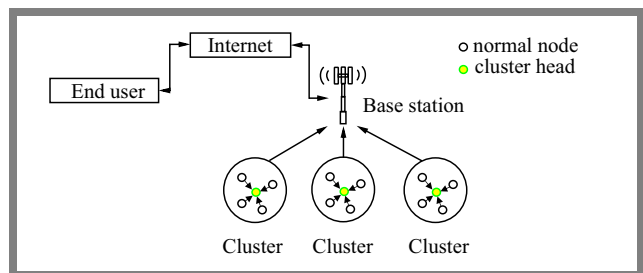


Fig. 2. Clustering architecture in WSN.

lation results and related discussions are presented in Section 4. In Section 5, a conclusion based on the findings and outcomes of the proposed protocol is presented, along with the suggested direction of future research.

## 2. Related Work

Homogeneous and heterogeneous clustering networks are the two main types of WSNs that need to be taken into account when considering the amount of energy used. All sensor nodes deployed within a homogeneous clustering network have the same level of energy, while nodes in a heterogeneous clustering network are deployed with different energy levels.

### 2.1. Homogeneous Clustering Protocols

A low energy adaptive clustering hierarchy (LEACH) protocol was developed by Heinzelman *et al.* [6] for homogeneous WSNs. LEACH is considered a benchmark for the various clustering algorithms in WSNs. In this protocol, each node behaves, at least once, as a CH. Data transfer from nodes to the BS is divided into two phases: the setup phase and the steady phase. During the setup phase, sensor nodes are grouped into a cluster by choosing one node to serve as their CH. Each node randomly selects a number between 0 and 1 and compares it with the threshold value. If the chosen number is below the threshold, the node acts as the CH. In the steady state, the CH gathers information from cluster nodes, performs data aggregation and fusion, and transmits it to the BS. In terms of functionality, LEACH increases stability period and network lifetime by distributing energy load among all sensor nodes, unlike in conventional protocols.

Farman *et al.* in [7] presented an improved grid-based hybrid network deployment (IGHND) approach for grid-based WSNs. In IGHND, data transfer from nodes to the BS is achieved by selecting a zone head (ZH) and by developing topology phases. During the topology development phase, if the number of nodes in a zone falls below the lower threshold, then the nodes in that zone will be combined with those in an adjacent zone, based on density and distance. A zone will be divided into sub-zones if the node count rises above the upper threshold. In the ZH selection phase, distance from the zone's center point, distance from neighboring nodes inside the zone, residual energy of a node, number of times a node has undergone ZH, and whether a node is merged parameters, are considered for ZH selection. Using this scheme, the IGHND prolongs network lifespan and increases network stability in IoT-based WSNs.

Singh *et al.* [8] proposed the LEACH – uniform size clusters (LEACH-USC) protocol, being an improved version of the balanced cluster size Formation (BCF) protocol. The BCF protocol achieves balanced cluster size formation in two phases: the initial cluster formation phase and the rescue phase. In this method, the CH selects and forms balanced clusters in the initial cluster formation phase. In the rescue phase, nodes not clustered join nearby clusters as normal nodes. Instead of a rescue phase, the LEACH-USC protocol uses a cluster

refurbishment phase as the second phase, in which excess nodes from large clusters are joined with other clusters as the second-best choice of the CH.

Kumar *et al.* [9] introduced a divide and rule sectorization (DRS) scheme for homogeneous WSNs to reduce the energy-hole creation problem. The DRS scheme divides the sensor field into sectors and inner circle regions. These sectors, in turn, are divided into segments and nodes are deployed in these segments uniformly. The three-tier communication model transfers data from the node to the BS by dynamically selecting forward nodes. The DRS scheme increases the network's lifetime by reducing the amount of energy consumed by nodes and provides better stability than existing protocols.

### 2.2. Heterogeneous Clustering Protocols

In paper [10], Verma *et al.* presented a new cost and sub-epoch-based stable energy-efficient clustering (CSSEEC) protocol for heterogeneous WSNs. The CSSEEC protocol uses a cost function to reduce the energy consumption of sensor nodes and a modified sub-epoch to balance energy among the nodes. The CSSEEC protocol ensures better network lifetime, network throughput, and stability period than existing protocols. Rawat *et al.* [11] proposed an energy-efficient heterogeneous clustering protocol, a two-level network model consisting of normal and power nodes. To select the appropriate CH, the residual energy of the node, total network energy and initial node energy are introduced into the threshold equation. The model reduces the amount of energy consumed by the nodes, prolongs network lifetime, and increases network throughput in comparison with existing protocols.

In [12], Faisal *et al.* described the Z-SEP protocol, a variation of the SEP protocol based on zones. SEP is a two-level heterogeneous clustering protocol comprising advance and normal nodes. Advance nodes are equipped with extra energy levels compared with normal nodes. In Z-SEP, the sensor field is divided into zone 0 and two head zones. Normal nodes are deployed only at zone 0 and send the sensed data directly to the BS, while advance nodes are deployed in both head zone 1 and head zone 2, uniformly and randomly. These data is sent to the BS using the CH. Z-SEP offers a longer network life and a longer stability period compared with existing protocols. Khan *et al.* [13] proposed the advanced Z-SEP (AZ-SEP) protocol, an enhancement of Z-SEP. In the AZ-SEP protocol, zone 0 nodes send data to the BS in a single-hop, while nodes in head zone 1 and head zone 2 transfer data to the BS in a multi-hop transmission. CHs in head zone 1 or head zone 2 transfer data to the nearby CH in head zone 1 or head zone 2, respectively. The AZ-SEP protocol optimizes average energy consumption and increases the stability period and network throughput compared to existing protocols.

Rowayda *et al.* [14] introduced the hybrid IoT (Hy-IoT) protocol for IoT applications, where the sensor field is divided into two regions: superior and regular. The superior region consists of higher energy nodes, while the regular region consists mainly of normal nodes and includes a minority of higher energy nodes. The superior region uses the LEACH protocol for CH selection, while the regular region uses the SEP pro-

toloc. The Hy-IoT protocol prolongs the network lifetime and increases its throughput and instability period compared with existing protocols.

A two-level heterogeneous clustering protocol in which the sensor field is divided into two regions is proposed by Abidi *et al.* in [15]. Normal nodes are deployed in region 1, near the BS, and advance nodes are deployed in region 2, far from the BS. Nodes in region 1 do not participate in the CH selection process, and directly send information directly to the BS. Nodes in region 2 participate in CH selection and send information to the BS through selected CHs. The proposed protocol gives a better network lifetime, stability period, and network throughput compared with the Z-SEP protocol.

Chithra *et al.* [16] researched an energy-efficient concentric circular clustering protocol (EECCCP). Three concentric circular zones make up the sensor field in EECCCP. The nodes are deployed in three zones: normal, advance and super. Nodes in normal and super zones submit data directly to the BS, while nodes in advance zone sent data to the BS using the clustering technique. The proposed protocol increases network lifetime and throughput compared with existing protocols. Tyagi *et al.* [17] introduced a lifetime extended multi-level heterogeneous routing (LE-MHR) protocol with k-levels horizontal heterogeneity. In LE-MHR, data transmission is divided into two stages: cluster formation and data transmission stage. These are similar to the setup and steady phases of the LEACH protocol. The authors presented four theorems for implementing the LE-MHR protocol. Several advanced nodes were implemented at each level of the network's hierarchy. The proposed protocol was simulated by varying energy heterogeneity parameters and initial energy values. The results reveal that LE-MHR enhances stability period, network lifetime, and ensures better distribution of cluster heads compared with existing protocols.

Alsafi *et al.* [18] and Qureshi *et al.* [19] presented a comparative performance analysis of distributed energy efficient clustering (DEEC), Developed DEEC (DDEEC), extended DEEC (EDEEC), and threshold-based DEEC (TDEEC) protocols. All these protocols follow a similar pattern in transferring data. If the nodes are close to the BS, relative to the nearby CH, the data is transferred directly to the BS. Otherwise, nodes clustered near the CH transfer data to their respective CH. The

DEEC and DDEEC protocols are designed for multi-level networks. In the DEEC protocol, advance nodes continue to punish energy consumption even after their energy level becomes low, compared with the normal node. EDEEC and TDEEC are designed for three-level networks comprising normal, advance, and super nodes. By incorporating residual and average energy levels in the current round into the threshold value equation, the performance of the EDEEC protocol was improved. TDEEC outperforms all other protocols in terms of network throughput, lifetime, and stability period.

### 3. Proposed Heterogeneous Clustering Protocol

TDEEC is a three-level heterogeneous clustering protocol consisting of normal, advance, and super nodes. These nodes are placed randomly in the sensor field. In the Z-SEP protocol, the sensor field is split into three regions, zone 0, head zone 1, and head zone 2, as depicted in Fig. 3. Only zone 0 has normal nodes deployed which communicate directly with the BS. Advance nodes are randomly and uniformly distributed in head zones 1 and 2, and these transfer data to the BS through the elected CH in either head zone 1 or head zone 2. From Fig. 3, the following conclusions can be drawn:

- Normal nodes in zone 0, close to the CH in head zone 1 or head zone 2, but far from the BS, only send data to the BS. As a result, normal nodes far from the BS consume more energy and perish quickly;
- Advance nodes are deployed only in the lower and upper parts of the sensor field, so normal nodes in the left- and right-hand parts dissipate more energy than those in the lower and upper parts of the sensor field.

Considering the problem formulated above, the following approach has been implemented to increase the network's lifespan and reduce node energy consumption:

- Instead of two-level networks, three-level networks are used for a longer network lifetime;
- Making normal nodes in zone 0 participate in cluster formation by considering nodes that are closer to the CH than the BS, only clustered into respective CH cluster formation;
- Making zones on each side of the sensor field and deploying high-energy nodes in the corners of the sensor field instead of the sides of the sensor field helps balance energy consumption.

By putting all these solutions together, we proposed a new zone-based three level heterogeneous clustering protocol (ZB-TLHCP) for energy optimization in WSNs.

#### 3.1. Network Model

As shown in Fig. 4, the sensor field is split into nine zones, based on the distance from the BS. We assume that  $n\_zone$  will end at  $d_0$  distance from the BS. All normal nodes are deployed in  $n\_zone$  only, all advance nodes are deployed uniformly in  $a\_zone$  1 ... 4 and super nodes are deployed

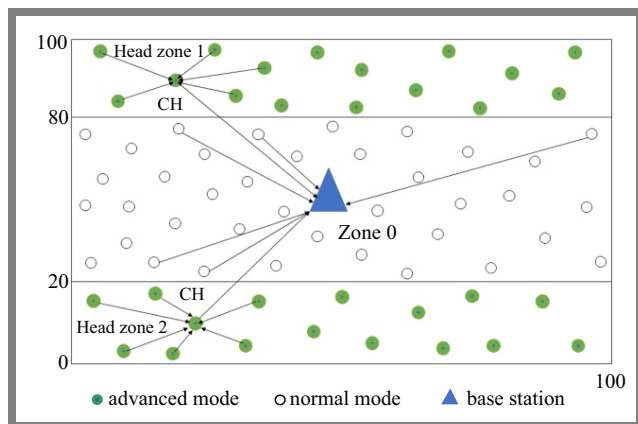


Fig. 3. Z-SEP network communication model.

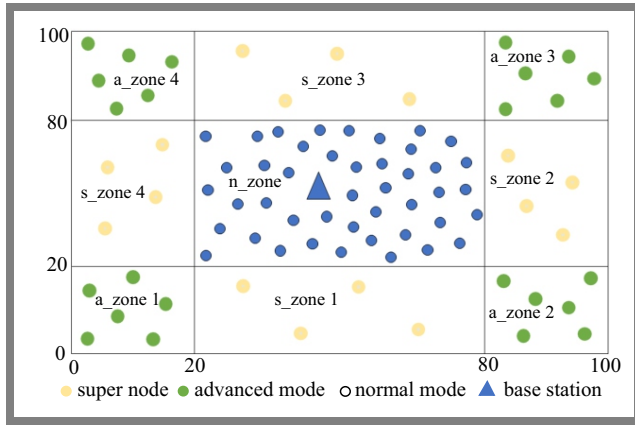


Fig. 4. ZB-TLHCP network model.

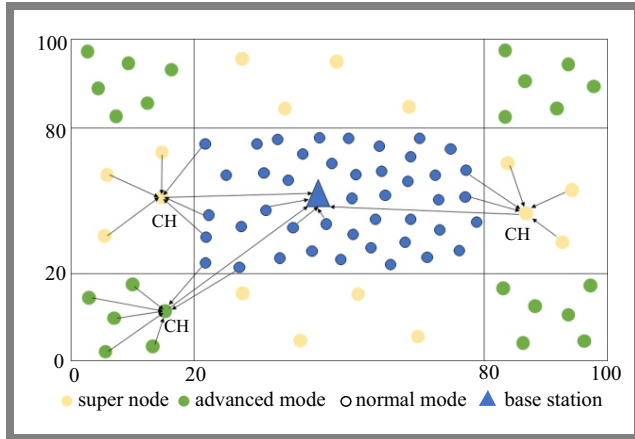


Fig. 5. ZB-TLHCP communication network model.

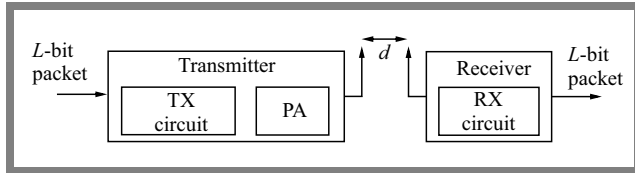


Fig. 6. The radio energy model for WSN.

uniformly in  $s\_zone\ 1 \dots 4$ . Nodes from  $n\_zone$  will not participate in the CH selection phase. In the cluster formation phase, all nodes in  $n\_zone$  calculate the distance from CHs and the BS. If the node is close to the BS, data packets are sent directly to the BS. Otherwise, the nodes join the nearby cluster's CH. Nodes not located in  $n\_zone$  and not selected as CH join the nearby CH, as shown in Fig. 5.

### 3.2. Radio Energy Model

During data transmission, the sensor node consumes energy [19]–[21], as shown in Fig. 6. The energy required by a sensor node to send an  $L$ -bit message over distance  $d$  is:

$$E_{TX} = \begin{cases} L \times E_{elec} + L \times \varepsilon_{mp} \times d^4, & d \geq d_0 \\ L \times E_{elec} + L \times \varepsilon_{fs} \times d^2, & d < d_0 \end{cases}, \quad (1)$$

where  $E_{TX}$  is the energy consumed by the transmitter,  $E_{elec}$  is energy consumed during the transmission by the companion device,  $\varepsilon_{mp}$ ,  $\varepsilon_{fs}$  are multipath and free space amplification

factors, respectively, and  $d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$  is the threshold distance.

The energy needed by a sensor node to receive an  $L$ -bit message is:

$$E_{RX} = L \times E_{elec}, \quad (2)$$

where  $E_{elec}$  is receiver circuit energy consumption.

The amount of energy left after either transmission or reception of the  $L$ -bit message is called residual energy  $E_R$  and is given by:

$$E_R = E_{TOTAL} - (E_{TX} + E_{RX}). \quad (3)$$

In the proposed protocol, regular nodes have an energy level of  $E_0$ , while advance nodes have an additional energy level of  $a$ . Super nodes are equipped with an extra energy level of  $b$ , such that  $a < b$ . The total initial energy of the network is:

$$E_{TOTAL} = N(1 - m)E_0 + N m(1 - m_0) \times (1 + a)E_0 + N m m_0(1 + b)E_0, \quad (4)$$

where  $N$  is the total number of nodes,  $m_0$  is the share of super nodes and  $m$  is the share of advance nodes. After CH selection and cluster formation, all nodes send data packets are sent to the BS as one round.

The energy consumption in one round is:

$$E_{Round} = 2L N E_{elec} + N E_{DA} + k \varepsilon_{mp} \times d_{toBS}^4 + N \varepsilon_{fs} d_{toCH}^2, \quad (5)$$

where  $E_{DA}$  denotes the data aggregation energy,  $k$  is the number of clusters in the current round,  $d_{toCH}$  is the distance between cluster member nodes and the CH, and  $d_{toBS}$  is the distance between the nodes and the BS, such that:

$$d_{toCH} = 0.765 \frac{M}{2},$$

$$d_{toBS} = \frac{M}{\sqrt{2\pi k}},$$

$$K_{opt} = \sqrt{\frac{N}{2\pi}} \cdot \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \cdot \frac{M}{d_{toBS}^2}. \quad (6)$$

In Eq. (6),  $M$  denotes the sensor field size,  $K_{opt}$  is an optimum number of clusters for efficient energy consumption ( $k = K_{opt}$ ). The total maximum rounds for WSNs can be estimated using Eqs. (4) and (5):

$$R = \frac{E_{TOTAL}}{E_{Round}}. \quad (7)$$

In order to identify which CH is best in a given round of iteration, the probability of threshold  $T(S_i)$  is introduced. If the probability threshold  $T(S_i)$  is exceeded, the node is chosen as the CH for that round.

$$T(S_i) = \begin{cases} \frac{P_i}{1 - P_i(r\% \frac{1}{P_i})} \cdot \frac{E_i(r)}{E(r)}, & S_i \in G \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

where  $r$  is the current round,  $P_i$  is the probability used for CH selection and  $G$  is the collection of nodes that have not been chosen as CHs in previous  $1/P_i$  rounds.

$$P_i = \begin{cases} \frac{P_{opt} \cdot E_i(r)}{1 + m(a + m_0 b) \cdot E_r}, & S_i \in \text{normal node} \\ \frac{P_{opt} \cdot (1 + a) E_i(r)}{1 + m(a + m_0 b) \cdot E_r}, & S_i \in \text{advance node} \\ \frac{P_{opt} \cdot (1 + b) E_i(r)}{1 + m(a + m_0 b) \cdot E_r}, & S_i \in \text{super node} \end{cases}. \quad (9)$$

In Eq. (9),  $E_{i(r)}$  is residual energy,  $P_{opt}$  is the reference value of the average probability of  $P_i$ ,  $\bar{E}(r)$  stands for the estimated energy that acts as standard reference energy for each node and is equal to:

$$\bar{E}(r) = \frac{1}{N} E_{TOTAL} \left( 1 - \frac{r}{R} \right). \quad (10)$$

The algorithm for the proposed ZB-TLHCP protocol is given as Algorithm 1.

**Algorithm 1.** Pseudo code of the proposed ZB-TLHCP protocol

```

1: Initialize sensor network parameters
2: Initialize population and energy for each sensor node
3: Sensor nodes deployment
4: for  $r = 0$  to maximum rounds do
5:   Check nodes' eligibility to become CH using
      $\text{mod}(r, \text{round}(\frac{1}{P})) == 0$ 
6:   Calculate average energy  $\bar{E}(r)$  of network in
     current round using Eq. (10)
7:   Check status of each node alive or dead.
8:   for  $i = 1$  to  $N$  do
9:     if  $\bar{E}(r)$  then
10:      Allot  $P_i$  respectively using Eq. (9)
11:     if  $E_R > 0$  then
12:        $temp_{rand} \rightarrow$  node assume some random
         value between 0 to 1
13:       if node not in  $n\_zone$  &  $temp_{rand} \leq T(S_i)$  then
14:          $N$  node considered as CH
15:       end if
16:     end if
17:   end if
18: end for
19: for  $i = 1$  to  $N$  do
20:   if  $E_R > 0$  then
21:     if node in  $n\_zone$  then
22:       Node check distance from BS and CH
23:       if node near to BS then
24:         Data directly sent to BS
25:       else node is clustered to near CH and sends data to
         that respective CH
26:     end if
27:     if node not in  $n\_zone$  & not selected as CH then
28:       Node is clustered to near CH and sends data to
         that respective CH
29:     end if
30:   end if
31: end if
32: end for
33: end for

```

## 4. Results and Discussion

We simulated our model by considering the parameters given in Tab. 1. Four performance parameters are considered:

**Tab. 1.** Parameters used for simulation.

| Parameters                 | Notation        | Value                        |
|----------------------------|-----------------|------------------------------|
| Number of nodes            | $N$             | 100                          |
| Transmit or receive energy | $E_{elec}$      | 50 nJ/bit                    |
| Initial energy             | $E_0$           | 0.5 J                        |
| Data aggregation energy    | $E_{DA}$        | 5 nJ/bit/message             |
| Multipath loss constant    | $\epsilon_{mp}$ | 0.0013 pJ/bit/m <sup>4</sup> |
| Free space loss constant   | $\epsilon_{fs}$ | 10 pJ/bit/m <sup>2</sup>     |
| Data packet size           | $L$             | 4000 bits                    |
| Network size               | $M$             | 100 × 100 m <sup>2</sup>     |
| Probability                | $P_{opt}$       | 0.1                          |
| Maximum rounds             | $R$             | 100,000                      |

- **Stability period** – Time from when networks become active to when the first node runs out of energy,
- **Instability period** – Time from the first node dying to the last node running out of energy, or the entire network collapsing,
- **Network lifetime** – Time from when a network becomes active until the final node runs out of energy or the entire network collapses,
- **Throughput** – The overall network data transmission rate. The rate of data transmission from nodes to the BS as well as from CHs to the BS.

By altering the number of advance nodes and super nodes, as well as their energy levels for given  $N$  and  $m$  value, we could examine overall performance.

The proposed model is simulated for the following four cases:

- Case 1:  $m = 0.3, m_0 = 0.2, a = 1.2, b = 2$ ,
- Case 2:  $m = 0.4, m_0 = 0.3, a = 1.3, b = 2.5$ ,
- Case 3:  $m = 0.5, m_0 = 0.4, a = 1.5, b = 3$ ,
- Case 4:  $m = 0.6, m_0 = 0.5, a = 1.6, b = 3.2$ .

We simulated the model ten times, and then averaged the results.

### 4.1. Matlab Simulation Results

**In case 1**, DEEC, Z-SEP, TDEEC [18], and ZB-TLHCP comprise 70 normal nodes, each with an energy level of  $E_0$ . In the DEEC protocol, 30 advanced nodes are considered, each with the energy level of  $(1 + rand \times 1.2) \times E_0$ . In Z-SEP, 30 advance nodes have an energy level of  $2.2E_0$ . In TDEEC and ZB-TLHCP, 24 advance nodes are simulated, each with the energy level of  $2.2E_0$  and six super nodes, each with the energy level of  $3E_0$ .

According to Fig. 7a, the longest stability period of the network (in descending order) is achieved by: Z-SEP, ZB-TLHCP, DEEC, and TDEEC. The order for the network lifetime is: ZB-TLHCP, TDEEC, Z-SEP, and DEEC (Fig. 7b). From Fig.

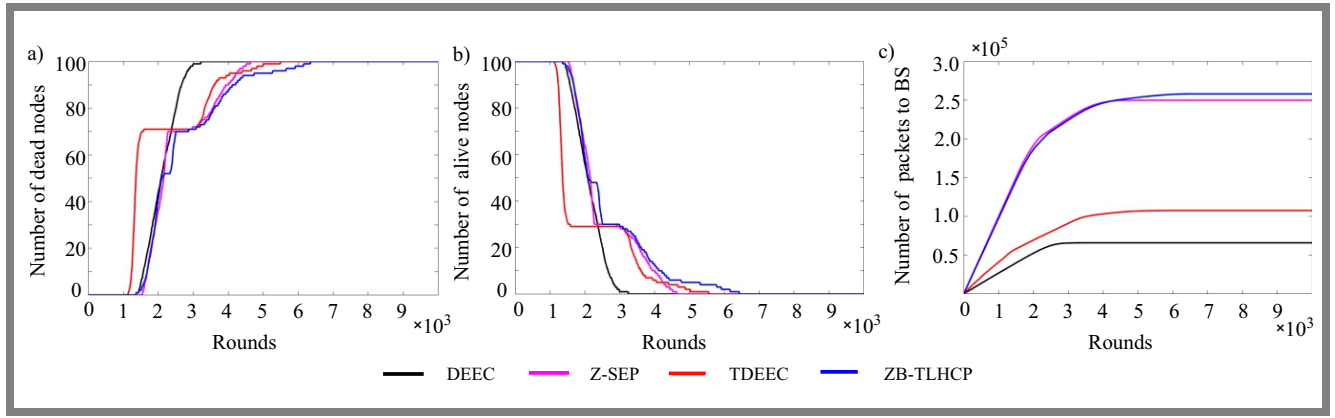


Fig. 7. Simulation results for case 1: a) stability period, b) network lifetime, and c) throughput.

7a and Fig. 7b, we can say that the results concerning network instability period are: ZB-TLHCP, TDEEC, Z-SEP, and DEEC. The throughput of the network is, in descending order: ZB-TLHCP, Z-SEP, TDEEC, and DEEC (Fig. 7c).

**In case 2**, it is assumed that DEEC, Z-SEP, TDEEC, and ZB-TLHCP have 60 normal nodes, each with an energy level of  $E_0$ . In the DEEC protocol, 40 advanced nodes, each with the energy level of  $(1 + rand \times 1.3) \times E_0$  are simulated. In Z-SEP, 40 advance nodes have an energy level of  $2.3E_0$ , while in TDEEC and ZB-TLHCP, 28 advance nodes, each with the energy level of  $2.3E_0$  and 12 super nodes, each with the energy level of  $3.5E_0$  are considered.

According to Fig. 8a, the longest stability period of the network (in descending order) is: Z-SEP, ZB-TLHCP, DEEC, and TDEEC. From Fig. 8b, the order for network lifetime is ZB-TLHCP, TDEEC, Z-SEP, and DEEC, while results concerning instability period of the network are, in descending order: ZB-TLHCP, TDEEC, Z-SEP, DEEC (Fig. 8a-b). The throughput of the network is, in descending order: ZB-TLHCP, Z-SEP, TDEEC, DEEC (Fig. 8c).

**The results for case 3** are as follows. DEEC, Z-SEP, TDEEC [18], and ZB-TLHCP have 50 normal nodes, each with an energy level of  $E_0$ . In the DEEC protocol, there are 50 advanced nodes, each with the energy level of  $(1 + rand \times 1.5) \times E_0$ . In Z-SEP, 50 advance nodes have an energy level of  $2.5E_0$ . In TDEEC & ZB-TLHCP, 30 advance nodes are simulated each with the energy level of  $2.5E_0$  and 20 super nodes, each with the energy level of  $4E_0$ .

According to Fig. 9, the longest stability period of the network is achieved by Z-SEP, ZB-TLHCP, DEEC, and TDEEC. The results for the network lifetime are: ZB-TLHCP, TDEEC, Z-SEP, and DEEC, while those concerning the instability period of the network are: ZB-TLHCP, TDEEC, Z-SEP, DEEC. From Fig. 9c the one can see that the order related to the highest throughput of the network is: ZB-TLHCP, Z-SEP, TDEEC, DEEC.

**In the last simulated case 4**, DEEC, Z-SEP, TDEEC, and ZB-TLHCP have 40 normal nodes, each with an energy level of  $E_0$ . In the DEEC protocol, 60 advanced nodes are analyzed, each with the energy level of  $(1 + rand \times 1.6) \times E_0$ , while in Z-SEP, 60 advance nodes have an energy level of  $2.6E_0$ . In TDEEC and ZB-TLHCP protocols, 30 advance nodes, each

with the energy level of  $2.6E_0$  and 30 super nodes, each with the energy level of  $4.2E_0$ .

From Fig. 10 we can conclude that the longest stability period of the network is achieved by Z-SEP, ZB-TLHCP, DEEC, and TDEEC. The sequence for network lifetime is: ZB-TLHCP, TDEEC, Z-SEP, and DEEC and the results related to instability period of the network are: ZB-TLHCP, TDEEC, Z-SEP, and DEEC. The network throughput is, in descending order: ZB-TLHCP, Z-SEP, TDEEC, and DEEC.

#### 4.2. Overall Analysis

Here, by increasing the number of advance and super nodes and their energy levels, we examine the performance of the proposed protocol by combining all results in one figure. The results shown in Fig. 11a indicate that the stability period remain almost at the same level as the number of advance nodes and super nodes, as well as their energy levels, increase. In case 1, the proposed ZB-TLHCP protocol offers an improvement in stability period of 2.25% and 22.81% over DEEC and TDEEC protocols, respectively, and an improvement in instability period of 61.63% over the Z-SEP protocol.

In case 2, the proposed ZB-TLHCP provides an improvement in stability period of 22.82% over the TDEEC protocol and an improvement in instability period of 222.84% and 90.15% over DEEC and Z-SEP protocols, respectively. In case 3, the proposed ZB-TLHCP solution offers an improvement in stability period of 5.97% and 26.49% over DEEC, and TDEEC protocols, respectively, as well as an improvement in instability period of 99.15% over the Z-SEP protocol.

In case 4, the proposed ZB-TLHCP offers an improvement in stability period of 1.36 % and 37.75% over DEEC and TDEEC protocols, respectively and an improvement in instability period of 106.30% over the Z-SEP protocol. In all four cases, ZB-TLHCP ensures higher better when compared with TDEEC, while lacking in terms of stability when compared with Z-SEP. Among these, Z-SEP performs better than ZB-TLHCP, DEEC, and TDEEC protocols. Due to the high network lifetime, the instability period of all protocols is increasing.

According to the results shown in Fig. 11b, the network lifetime of all protocols increases with a higher number of ad-

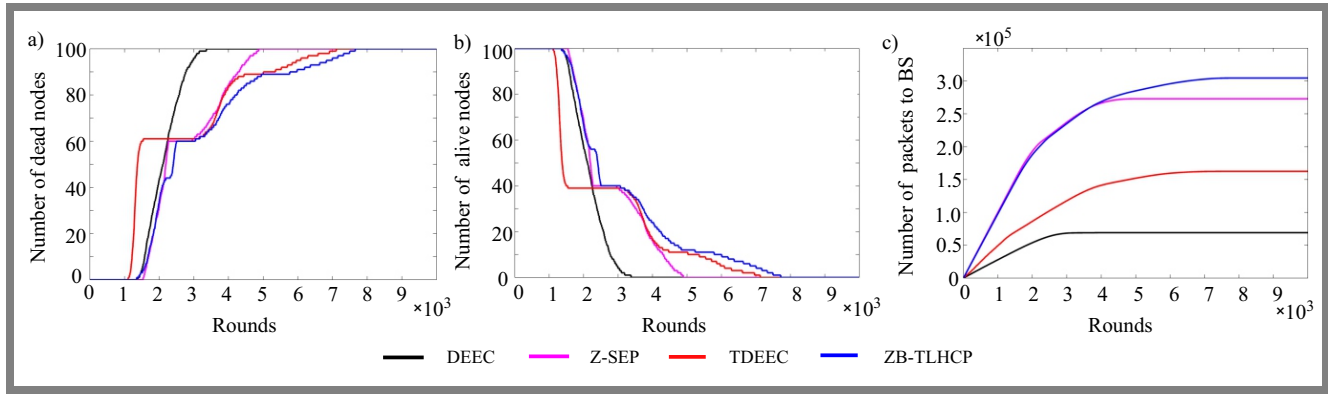


Fig. 8. Simulation results for case 2: a) stability period, b) network lifetime, and c) throughput.

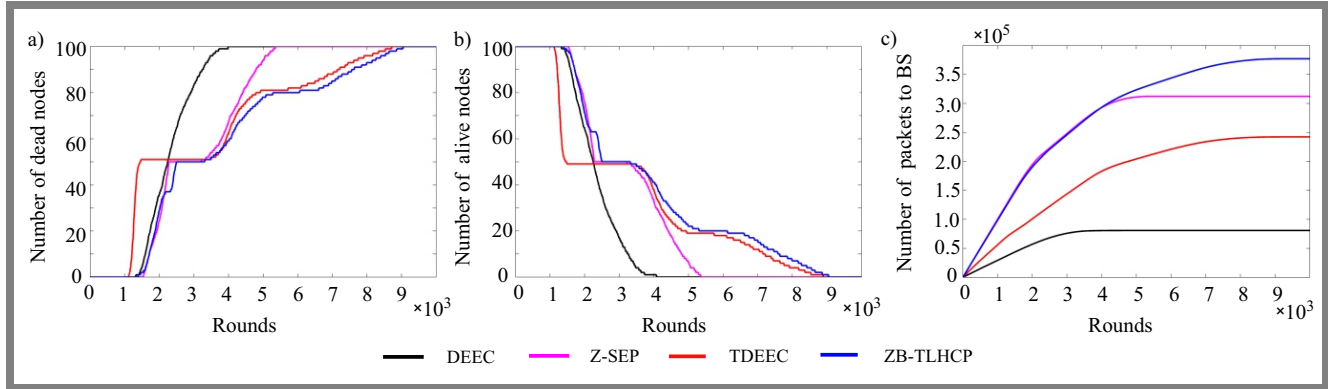


Fig. 9. Simulation results for case 3: a) stability period, b) network lifetime, and c) throughput.

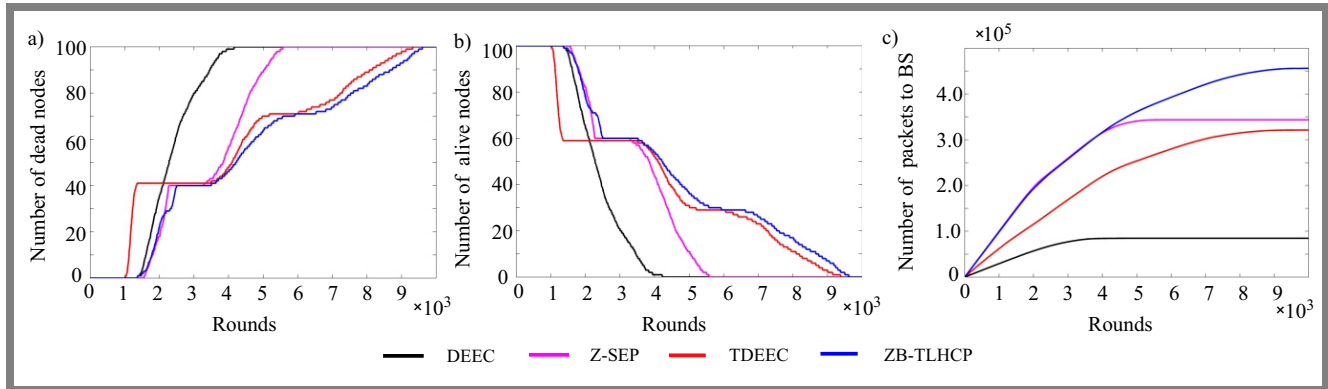


Fig. 10. Simulation results for case 4: a) stability period, b) network lifetime, and c) throughput.

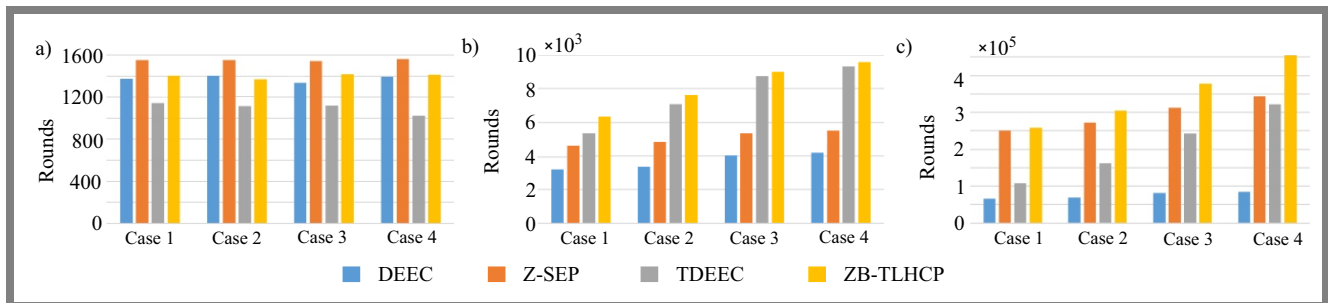


Fig. 11. Overall analysis of DEEC, Z-SEP, TDEEC, and ZB-TLHCP protocols for: a) stability period, b) network lifetime, and c) throughput.

vance nodes and super nodes and their energy levels. In case 1, ZB-TLHCP provides an improvement of 98.21%, 37.62%, and 18.56% over DEEC, Z-SEP, and TDEEC protocols, re-

spectively. In case 2, ZB-TLHCP ensures an improvement of 128.29%, 57.49%, and 8.01% over DEEC, Z-SEP, and TDEEC protocols, respectively. In case 3, ZB-TLHCP of-

fers an improvement of 123.59 %, 68.26%, and 2.97% over DEEC, Z-SEP, and TDEEC, respectively. In case 4, ZB-TLHCP achieves an improvement of 129.21 %, 73.60%, and 2.85% over DEEC, Z-SEP, and TDEEC, respectively. Hence, ZB-TLHCP achieves the longest network lifetime compared to DEEC, Z-SEP, and TDEEC, in all four scenarios.

Figure 11c, shows that the network throughput increases along with the number of advance nodes, super nodes, and their energy levels. In case 1, ZB-TLHCP provides an improvement of 292.64%, 3.30%, and 140.26% in throughput compared to DEEC, Z-SEP, and TDEEC, respectively. In case 2, ZB-TLHCP guarantees and improvement of 341.49%, 11.62%, and 87.56% in throughput compared to DEEC, Z-SEP, and TDEEC, respectively. In case 3, ZB-TLHCP ensures an improvement of 366.27%, 20.78%, and 55.66% in throughput compared to DEEC, Z-SEP, and TDEEC, respectively.

In case 4, ZB-TLHCP offers an improvement of 438.92%, 32.76%, 41.98% in throughput compared to DEEC, Z-SEP, and TDEEC, respectively. Hence, ZB-TLHCP has the highest throughput compared to DEEC, Z-SEP, and TDEEC in all four cases.

## 5. Conclusions and Future Work

This paper proposes a heterogeneous clustering protocol for WSNs, known as ZB-TLHCP. It divides the sensor region into zones, based on distance from the BS. Sensor nodes are deployed uniformly and randomly within the zone. The proposed model is simulated by varying the number of super and advance nodes and their energy levels for four different cases. The results reveal that the proposed protocol prolongs the network lifetime and increases the instability period and network throughput. In the future, this protocol can be integrated with nature-inspired algorithms [21], [22] to further enhance the effectiveness of the proposed protocol.

## References

- [1] A. Verma, S. Kumar, P.R. Gautam, T. Rashid, and A. Kumar, "Broadcast and Reliable Coverage Based Efficient Recursive Routing in Large-scale WSNs", *Telecommunication Systems*, vol. 75, no. 1, pp. 63–78, 2020 (<https://doi.org/10.1007/s11235-020-00679-5>).
- [2] S.R.J. Ramson and D.J. Moni, "Applications of Wireless Sensor Networks – A Survey", *IEEE International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT)*, Coimbatore, India, pp. 325–329, 2017 (<https://doi.org/10.1109/ICEEIMT.2017.8116858>).
- [3] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy Conservation in Wireless Sensor Networks: A Survey", *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009 (<https://doi.org/10.1016/j.adhoc.2008.06.003>).
- [4] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: a Survey", *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002 ([https://doi.org/10.1016/S1389-1286\(01\)00302-4](https://doi.org/10.1016/S1389-1286(01)00302-4)).
- [5] J. Amutha, S. Sharma, and S.K. Sharma, "Strategies Based on Various Aspects of Clustering in Wireless Sensor Networks Using Classical, Optimization and Machine Learning Techniques: Review, Taxonomy, Research Findings, Challenges and Future Directions", *Computer Science Review*, vol. 40, art. no. 100376, 2021 (<https://doi.org/10.1016/j.cosrev.2021.100376>).
- [6] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", in: *IEEE Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, USA, 2000 (<https://doi.org/10.1109/HICSS.2000.926982>).
- [7] H. Farman *et al.*, "Multi-criteria Based Zone Head Selection in Internet of Things Based Wireless Sensor Networks", *Future Generation Computer Systems*, vol. 87, pp. 364–371, 2018 (<https://doi.org/10.1016/j.future.2018.04.091>).
- [8] J. Singh *et al.*, "A Node Overhaul Scheme for Energy Efficient Clustering in Wireless Sensor Networks", *IEEE Sensors Letters*, vol. 5, no. 4, pp. 1–4, 2021 (<https://doi.org/10.1109/LSENS.2021.3068184>).
- [9] S. Kumar, P.R. Gautam, T. Rashid, A. Verma, and A. Kumar, "Division Algorithm Based Energy-efficient Routing in Wireless Sensor Networks", *Wireless Personal Communications*, vol. 122, no. 3, pp. 2335–2354, 2022 (<https://doi.org/10.1007/s11277-021-08996-x>).
- [10] A. Verma, T. Rashid, P.R. Gautam, S. Kumar, and A. Kumar, "Cost and Sub-epoch Based Stable Energy-efficient Clustering Algorithm for Heterogeneous Wireless Sensor Networks", *Wireless Personal Communications*, vol. 107, no. 2, pp. 1865–1879, 2019 (<https://doi.org/10.1007/s11277-019-06362-6>).
- [11] P. Rawat, S. Chauhan, and R. Priyadarshi, "A Novel Heterogeneous Clustering Protocol for Lifetime Maximization of Wireless Sensor Network", *Wireless Personal Communications*, vol. 117, no. 2, pp. 825–841, 2021 (<https://doi.org/10.1007/s11277-020-07898-8>).
- [12] S. Faisal *et al.*, "Z-SEP: Zonal-Stable Election Protocol for Wireless Sensor Networks", *Journal of Basic and Applied Scientific Research*, vol. 3, no. 5, pp. 132–139, 2013 (<https://doi.org/10.48550/arXiv.1303.5364>).
- [13] F.A. Khan, M. Khan, M. Asif, A. Khalid, and I.U. Haq, "Hybrid and Multi-hop Advanced Zonal-stable Election Protocol for Wireless Sensor Networks", *IEEE Access*, vol. 7, pp. 25334–25346, 2019 (<https://doi.org/10.1109/ACCESS.2019.2899752>).
- [14] R.A. Sadek, "Hybrid Energy Aware Clustered Protocol for IoT Heterogeneous Network", *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 166–177, 2018 (<https://doi.org/10.1016/j.fcij.2018.02.003>).
- [15] W. Abidi and T. Ezzedine, "Effective Clustering Protocol Based on Network Division for Heterogeneous Wireless Sensor Networks", *Computing*, vol. 102, no. 2, pp. 413–425, 2020 (<https://doi.org/10.1007/s00607-019-00757-w>).
- [16] A. Chithra and R.S.S. Kumari, "A Novel 3-level Energy Heterogeneity Clustering Protocol with Hybrid Routing for a Concentric Circular Wireless Sensor Network", *Cluster Computing*, vol. 22, no. 6, pp. 11101–11108, 2019 (<https://doi.org/10.1007/s10586-017-1310-9>).
- [17] S. Tyagi, S. Tanwar, S.K. Gupta, N. Kumar, and J. Rodrigues, "A Lifetime Extended Multi-levels Heterogeneous Routing Protocol for Wireless Sensor Networks", *Telecommunication Systems*, vol. 59, no. 1, pp. 43–62, 2015 (<https://doi.org/10.1007/s11235-014-9884-5>).
- [18] S. Alsafi and S.A. Talab, "Implementation of DEEC, DDEEC, EDEEC and TDEEC Protocols using MATLAB in Wireless Sensor Network", *International Journal of Advanced Networking and Applications*, vol. 12, no. 3, pp. 4596–4600, 2020 (<https://doi.org/10.35444/IJANA.2020.12304>).
- [19] T.N. Qureshi, N. Javaid, M. Malik, U. Qasim, and Z.A. Khan, "On Performance Evaluation of Variants of DEEC in WSNs", in: *IEEE Seventh International Conference on Broadband, Wireless Computing, Communication and Applications*, Victoria, Canada, pp. 162–169, 2012 (<https://doi.org/10.1109/BWCCA.2012.35>).
- [20] S. Kaur *et al.*, "Heterogeneous Load Balancing Clustering Protocol for Wireless Sensor Networks", *Cognitive Systems Research*, vol. 70, pp. 10–17, 2021 (<https://doi.org/10.1016/j.cogsys.2021.07.001>).

- [21] I.E. Agbehadji *et al.*, “Clustering Algorithm Based on Nature-inspired Approach for Energy Optimization in Heterogeneous Wireless Sensor Network”, *Applied Soft Computing*, vol. 104, art. no. 107171, 2021 (<https://doi.org/10.1016/j.asoc.2021.107171>).
- [22] A. Slowik and H. Kwasnicka, “Nature Inspired Methods and Their Industry Applications: Swarm Intelligence Algorithms”, *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1004–1015, 2017 (<https://doi.org/10.1109/TII.2017.2786782>).

---

**Sanapala Shanmukha Rao, M.Tech. Student**

 <https://orcid.org/0000-0001-5293-5690>

E-mail: shanmukhasanapala19@gmail.com

Motilal Nehru National Institute of Technology, Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

**Shilpi, Ph.D. Student**

 <https://orcid.org/0000-0001-7527-5090>

E-mail: shilpi@mnnit.ac.in

Motilal Nehru National Institute of Technology, Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

**Arvind Kumar, Ph.D.**

 <https://orcid.org/0000-0003-3283-5849>

E-mail: arvindk@mnnit.ac.in

Motilal Nehru National Institute of Technology, Allahabad,  
Prayagraj, India

<http://www.mnnit.ac.in>

# Information for Authors

**Journal of Telecommunications and Information Technology (JTIT)** is published quarterly since 1999. 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/IS-SPIT.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 acceptance. 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.

## Design and Analysis of a Low-profile Microstrip Antenna for 5G Applications Using AI-based PSO Approach

*K. Kundu, A. Bhattacharya, F.H. Mohammed, and N.N. Pathak*

*Paper*

68

## Utilization of Reconfigurable Intelligent Surfaces with Context Information: Use Cases

*Ł. Kułacz*

*Paper*

74

## Minimized Group Delay FIR Low Pass Filter Design Using Modified Differential Search Algorithm

*S. Prajapati, S. Rai, M. Tiwari and A.K. Dwivedi*

*Paper*

78

## Performance Evaluation of a Zone-based Three-level Heterogeneous Clustering Protocol for WSNs

*S.S. Rao, Shilpi, and A. Kumar*

*Paper*

85



National Institute  
of Telecommunications

### Editorial Office

National Institute  
of Telecommunications  
Szachowa st 1  
04-894 Warsaw, Poland

phone +48 22 512 81 83  
fax +48 22 512 84 00  
e-mail: [journal@jt.it.pl](mailto:journal@jt.it.pl)  
[editorial.office@il-pib.pl](mailto:editorial.office@il-pib.pl)  
<https://www.jt.it.pl>