

Wireless Communications and Mobile Computing

# Multidimensional Modulation for 6G Communications

Lead Guest Editor: Jun Li

Guest Editors: Shuping Dang and Yingyang Chen





---

# **Multidimensional Modulation for 6G Communications**

Wireless Communications and Mobile Computing

---

# **Multidimensional Modulation for 6G Communications**

Lead Guest Editor: Jun Li

Guest Editors: Shuping Dang and Yingyang Chen



---



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Chief Editor

































Zhipeng Cai , USA

## Associate Editors

Ke Guan , China  
Jaime Lloret , Spain  
Maode Ma , Singapore

## Academic Editors

Muhammad Inam Abbasi, Malaysia  
Ghufran Ahmed , Pakistan  
Hamza Mohammed Ridha Al-Khafaji ,  
Iraq  
Abdullah Alamoodi , Malaysia  
Marica Amadeo, Italy  
Sandhya Aneja, USA  
Mohd Dilshad Ansari, India  
Eva Antonino-Daviu , Spain  
Mehmet Emin Aydin, United Kingdom  
Parameshchhari B. D. , India  
Kalapaveen Bagadi , India  
Ashish Bagwari , India  
Dr. Abdul Basit , Pakistan  
Alessandro Bazzi , Italy  
Zdenek Becvar , Czech Republic  
Nabil Benamar , Morocco  
Olivier Berder, France  
Petros S. Bithas, Greece  
Dario Bruneo , Italy  
Jun Cai, Canada  
Xuesong Cai, Denmark  
Gerardo Canfora , Italy  
Rolando Carrasco, United Kingdom  
Vicente Casares-Giner , Spain  
Brijesh Chaurasia, India  
Lin Chen , France  
Xianfu Chen , Finland  
Hui Cheng , United Kingdom  
Hsin-Hung Cho, Taiwan  
Ernestina Cianca , Italy  
Marta Cimitile , Italy  
Riccardo Colella , Italy  
Mario Collotta , Italy  
Massimo Condoluci , Sweden  
Antonino Crivello , Italy  
Antonio De Domenico , France  
Floriano De Rango , Italy

Antonio De la Oliva , Spain  
Margot Deruyck, Belgium  
Liang Dong , USA  
Praveen Kumar Donta, Austria  
Zhuojun Duan, USA  
Mohammed El-Hajjar , United Kingdom  
Oscar Esparza , Spain  
Maria Fazio , Italy  
Mauro Femminella , Italy  
Manuel Fernandez-Veiga , Spain  
Gianluigi Ferrari , Italy  
Luca Foschini , Italy  
Alexandros G. Fragkiadakis , Greece  
Ivan Ganchev , Bulgaria  
Óscar García, Spain  
Manuel García Sánchez , Spain  
L. J. García Villalba , Spain  
Miguel Garcia-Pineda , Spain  
Piedad Garrido , Spain  
Michele Girolami, Italy  
Mariusz Glabowski , Poland  
Carles Gomez , Spain  
Antonio Guerrieri , Italy  
Barbara Guidi , Italy  
Rami Hamdi, Qatar  
Tao Han, USA  
Sherief Hashima , Egypt  
Mahmoud Hassaballah , Egypt  
Yejun He , China  
Yixin He, China  
Andrej Hrovat , Slovenia  
Chunqiang Hu , China  
Xuexian Hu , China  
Zhenghua Huang , China  
Xiaohong Jiang , Japan  
Vicente Julian , Spain  
Rajesh Kaluri , India  
Dimitrios Katsaros, Greece  
Muhammad Asghar Khan, Pakistan  
Rahim Khan , Pakistan  
Ahmed Khattab, Egypt  
Hasan Ali Khattak, Pakistan  
Mario Kolberg , United Kingdom  
Meet Kumari, India  
Wen-Cheng Lai , Taiwan

Jose M. Lanza-Gutierrez, Spain  
Pavlos I. Lazaridis , United Kingdom  
Kim-Hung Le , Vietnam  
Tuan Anh Le , United Kingdom  
Xianfu Lei, China  
Jianfeng Li , China  
Xiangxue Li , China  
Yaguang Lin , China  
Zhi Lin , China  
Liu Liu , China  
Mingqian Liu , China  
Zhi Liu, Japan  
Miguel López-Benítez , United Kingdom  
Chuanwen Luo , China  
Lu Lv, China  
Basem M. ElHalawany , Egypt  
Imadeldin Mahgoub , USA  
Rajesh Manoharan , India  
Davide Mattera , Italy  
Michael McGuire , Canada  
Weizhi Meng , Denmark  
Klaus Moessner , United Kingdom  
Simone Morosi , Italy  
Amrit Mukherjee, Czech Republic  
Shahid Mumtaz , Portugal  
Giovanni Nardini , Italy  
Tuan M. Nguyen , Vietnam  
Petros Nicolitidis , Greece  
Rajendran Parthiban , Malaysia  
Giovanni Pau , Italy  
Matteo Petracca , Italy  
Marco Picone , Italy  
Daniele Pinchera , Italy  
Giuseppe Piro , Italy  
Javier Prieto , Spain  
Umair Rafique, Finland  
Maheswar Rajagopal , India  
Sujan Rajbhandari , United Kingdom  
Rajib Rana, Australia  
Luca Reggiani , Italy  
Daniel G. Reina , Spain  
Bo Rong , Canada  
Mangal Sain , Republic of Korea  
Praneet Saurabh , India

Hans Schotten, Germany  
Patrick Seeling , USA  
Muhammad Shafiq , China  
Zaffar Ahmed Shaikh , Pakistan  
Vishal Sharma , United Kingdom  
Kaize Shi , Australia  
Chakchai So-In, Thailand  
Enrique Stevens-Navarro , Mexico  
Sangeetha Subbaraj , India  
Tien-Wen Sung, Taiwan  
Suhua Tang , Japan  
Pan Tang , China  
Pierre-Martin Tardif , Canada  
Sreenath Reddy Thummaluru, India  
Tran Trung Duy , Vietnam  
Fan-Hsun Tseng, Taiwan  
S Velliangiri , India  
Quoc-Tuan Vien , United Kingdom  
Enrico M. Vitucci , Italy  
Shaohua Wan , China  
Dawei Wang, China  
Huaqun Wang , China  
Pengfei Wang , China  
Dapeng Wu , China  
Huaming Wu , China  
Ding Xu , China  
YAN YAO , China  
Jie Yang, USA  
Long Yang , China  
Qiang Ye , Canada  
Changyan Yi , China  
Ya-Ju Yu , Taiwan  
Marat V. Yuldashev , Finland  
Sherali Zeadally, USA  
Hong-Hai Zhang, USA  
Jiliang Zhang, China  
Lei Zhang, Spain  
Wence Zhang , China  
Yushu Zhang, China  
Kechen Zheng, China  
Fuhui Zhou , USA  
Meiling Zhu, United Kingdom  
Zhengyu Zhu , China

## Contents



---

### **Fairness-Oriented Transmission Schemes for Multiuser MISO Broadcast Channels**

Jianjian Wu , Xinyu Liu , Chanzi Liu , Chi-Tsun Cheng , and Qingfeng Zhou 

Research Article (10 pages), Article ID 8172300, Volume 2022 (2022)

### **Optimal Element Allocation for RIS-Aided Physical Layer Security**

Ying Zhang , Guoan Zhang , Siyu Chen, Jaeho Choi, and Pin-Han Ho


Research Article (7 pages), Article ID 4617366, Volume 2022 (2022)

### **Reconfigurable Intelligent Surface Assisted Non-Terrestrial NOMA Networks**

Xuanhao Lian , Xinwei Yue , Xuehua Li , Xiang Yun, Tian Li, and Dehan Wan 



Research Article (13 pages), Article ID 8494630, Volume 2022 (2022)

### **Multiple Dimensional Encoding/Modulation Shift-and-Addition Design for Distributed Systems**

Mingjun Dai, Chanting Zhang, Zhaoyan Hong , Paweł Wawrzyński, Tomasz Trzciński, and Xiaohui Lin





Research Article (10 pages), Article ID 5615041, Volume 2022 (2022)

### **Multiband RoF System with Good Transmission Performance Based on Additional SCS**

Anliang Liu  and Hongzhi Li 


Research Article (7 pages), Article ID 5747959, Volume 2022 (2022)

### **Highly Efficient MIMO-OFDM with Index Modulation**

Zeng Hu , Dehuan Wan , Yankun Chen, Jinpei Yan, Yun Liu , Pengfei Guo , Liming Chen, and Wen Zhou

Research Article (14 pages), Article ID 7399457, Volume 2022 (2022)

### **Chebyshev-Gauss Approximation Analysis for Mobile Edge Computing-Aided IoT Networks**

Fusheng Zhu, Liming Chen, Wen Zhou, Dan Deng, Yajuan Tang, Jun Liu , Yuwei Zhang, Tao Cui, Lin Zhang, Jing Wang, and Sun Li

Research Article (6 pages), Article ID 3099914, Volume 2022 (2022)

## Research Article

# Fairness-Oriented Transmission Schemes for Multiuser MISO Broadcast Channels

Jianjian Wu <sup>1</sup>, Xinyu Liu <sup>2</sup>, Chanzi Liu <sup>2</sup>, Chi-Tsun Cheng <sup>3</sup> and Qingfeng Zhou <sup>1,2</sup>

<sup>1</sup>School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China

<sup>2</sup>School of Electric Engineering and Intelligentization, Dongguan University of Technology, Dongguan, Guangdong 523808, China

<sup>3</sup>Manufacturing, Materials and Mechatronics, School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

Correspondence should be addressed to Qingfeng Zhou; [enqfzhou@ieee.org](mailto:enqfzhou@ieee.org)

Received 17 August 2022; Accepted 3 October 2022; Published 18 November 2022

Academic Editor: Jun Li

Copyright © 2022 Jianjian Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasing number of Internet of Things (IoT), Industry 4.0 (I4.0), and mobile devices, it can be expected that base stations will have to serve more and more clients with a limited number of antennas. For their broadcast channels, nonorthogonal multiple access (NOMA) and blind interference alignment (BIA) are two efficient and commonly adopted transmission schemes. This paper conducts a comparison study on these techniques on a 3-user  $2 \times 1$  multiple-input single-output (MISO) broadcast channel with a limited number of transmit antennas. Specifically, space-time block coding based NOMA (STBC-NOMA) and NOMA-assisted beamforming (NOMA-BF) are compared with BIA. Both perfect and imperfect successive interference cancellation (SIC) have been considered for NOMA-based schemes, and the theoretical achievable rates of all schemes have been derived. Furthermore, with a given fairness constraint among end users, the power allocation (PA) problems have been solved for cases when accurate channel state information is available at the transmitter (CSIT) as well as when only path loss information is available. Numerical results show the following: (1) none of the schemes under this study can always outperform the others under different SNR regions. (2) With imperfect SIC, NOMA-BF, and STBC-NOMA both suffer from a significant performance loss under a high SNR condition. (3) Fairness PA with only path loss information provides similar performance as that with perfect CSIT, thus partial CSIT is adequate for system or scheme designs in practice.

## 1. Introduction

In modern wireless communications networks, the ever-increasing demand for device mobility and connectivity makes system design challenging [1–3]. Over the years, various advanced communication solutions and technologies have been emerged to address the issues [4–6]. Furthermore, conventional wireless systems have been coupled with Artificial Intelligence (AI) techniques in response to the desire for a higher level of intelligence in future networks [7–10]. Among these techniques, multiuser access becomes challenging when the number of end users is much larger than the antennas of the base station. Several technologies have been developed to solve the problem, including beamforming (BF), interference alignment (IA), and nonorthogonal multiple access (NOMA).

It is worth noting that channel state information (CSI) is a critical input parameter for enhancing performance in most wireless technologies [11, 12].

BF has been studied to improve MIMO systems' capacity and spectrum efficiency [13–15] for decades. For scenarios with base station's antenna number larger than its end user's number, BF-based methods like matched-filter-based BF, zero-forcing-based BF (ZF-BF), regularized ZF-BF [16], hybrid BF [17], and deep-learning-based BF algorithm [18] work well. However, when the transmit antenna number is less than the number of end users, the aforementioned BF-based methods start to fail or drop their performances. To tackle this problem, remedies like beamforming based on the signal-to-leakage-noise ratio (SLNR) have been proposed [19]. The common drawback is that the corresponding



interference leakage causes significant performance reductions, especially when trying to achieve a fair transmission power allocation (PA) among end users.

Unlike BF, which is mainly designed for serving a limited number of end users simultaneously, interference alignment (IA) is designed for scenarios where the end users' number is much larger than the base station's antenna number [20]. The main idea of IA is to adopt symbol extension to yield a big precoding space, then align interference signals into a small-enough subspace and use the remaining space for transmitting desired symbols. IA can be applied to manage interference under different scenarios, such as an IA algorithm for managing multitier interference in a two-tier HetNet [21], and an IA application for Coordinated Multi-Point (CoMP) systems [22]. Even with its remarkable potential, the practical implementation of IA is challenging. One of its limitations comes from the requirement of having a perfect CSIT. Fortunately, this requirement could be relaxed for broadcast channels by a so-called blind interference alignment (BIA) scheme which has been proposed in [23]. BIA is achieved by elaborately designing channel patterns over several symbol extensions, often known as "supersymbols". The channel pattern of BIA could be implemented naturally or artificially. Over the years, different BIAs have been proposed, such as a naturally implemented BIA in a homogeneous block fading system [24], an artificial BIA implementation with reconfigurable antennas in  $K$ -user SISO interference channels [25], a hybrid BIA and NOMA approach for inter-cluster interference canceling [26], a reconfigurable-antenna-based BIA with minimized mode-switching overhead [27], and a BIA implementation to maintain mode-switching fairness among end users [28].

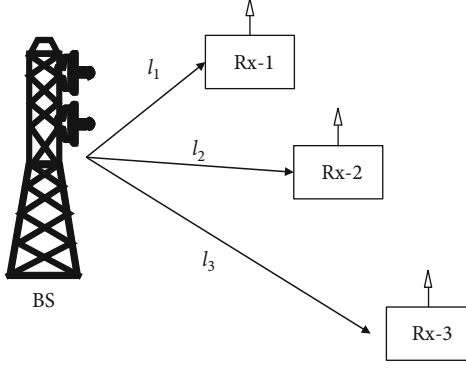
Rather than divisions of time, frequency, or code, NOMA considers the power domain and uses the power division to distinguish end users, also known as power-domain NOMA (PD-NOMA). So, unlike orthogonal multiple access (OMA), NOMA uses the same time, frequency, and code resource to serve multiple end users and can achieve higher spectrum efficiency, user fairness, and lower system delay simultaneously. Since its appearance, it has been considered as a promising technique for resolving the multiple access issues mentioned before. Shortly, a general framework combining MIMO and NOMA has been constructed [29]. As the power domain is unrelated to the domains of time, frequency, and code, NOMA is compatible with most conventional methods to deliver multiple benefits simultaneously, like the NOMA-based power transfer and backscatter communication in [30], the NOMA-based mobile edge computing (MEC) in [31], and the NOMA-assisted UAV system in [32]. In PD-NOMA, successive interference cancellation (SIC) is the key component for decoding. In SIC, sequential decoding is applied, where signals with higher power than the desired signal are decoded first and then subtracted from received signals. Subsequently, the desired signal is decoded by regarding other signals as noise. Due to this sequential property in SIC, the decoding order influences performances in PD-NOMA and different attempts have been made to resolve the issues. In [33], it is shown that the decoding order has no effect on

the achievable sum-rate in uplink PD-NOMA but may lead to unfairness among users. Considering users with a minimum rate requirement, they studied a joint user grouping, decoding order, and power control problem. In [34], a joint position, decoding order, and power allocation problem is investigated in an Unmanned Aerial Vehicle- (UAV-) based downlink NOMA. Besides, imperfect SIC has also been studied intensively. Since the recovery of each symbol with SIC depends on previous decodings, imperfect SIC causes severe residual error propagation, which makes NOMA difficult to adopt in practice. In [35], a power consumption problem in a multicell NOMA system has been studied, which evaluated the influence of imperfect SIC. In [36], the impact of imperfect SIC has been investigated in MIMO-NOMA systems when user fairness is considered. Results show that employing NOMA is not always beneficial when SIC imperfection is significant when compared with the MIMO-OMA scheme.

With the increasing popularity of IoT and I4.0 devices, the ratio between the number of end users and the number of antennas at the corresponding base station is expected to be more and more extreme, which can hardly be handled by conventional MIMO BF approaches. In contrast, IA and NOMA approaches appear to be more promising. However, to our best knowledge, their practicalities have not been extensively studied. In this paper, we study and compare three potential schemes, i.e., BIA, STBC-based NOMA, and NOMA-assisted BF, when power allocation with fairness is considered. Achievable sum-rates are derived for all schemes with the consideration of both perfect and imperfect SIC. PA problems for all schemes are solved by assuming accurate CSIT is available. PA problems for BIA and STBC-NOMA have been further studied by assuming only path loss information is available. The numerical simulation results show that the fair PA with partial channel information can achieve similar performance than that with perfect CSIT, and it also has a lower computation complexity. Among the three candidate schemes under this study, considering both the imperfect SIC and the overhead of perfect CSIT, BIA appears to be a promising candidate in practical scenarios.

## 2. System Model

Consider a  $K$ -user  $M \times 1$  broadcast channel with  $K = 3$  as shown in Figure 1, where the base station (BS) is equipped with  $M = 2$  antennas, and the  $k$ -th user (Rx- $k$ ) is equipped with  $N_k = 1$  antenna. Let  $d_k$  be the achievable degree-of-freedom (DoF) of the  $k$ -th user, then  $d_k \leq \min(M, N_k)$ . Moreover, the maximum achievable sum DoF is upper bounded by  $\min(M, \sum_{k=1}^K N_k)$  [20]. This model is a simple example of a multiuser channel in which the end user's number is larger than the number of BS's antennas, i.e.,  $M < K$ . In this channel, we apply a composite link model with both quasistatic Rayleigh fading and large-scale path loss considered. Then the channel matrix from BS to Rx- $k$  is modeled as  $\mathbf{h}_k = \mathbf{g}_k / \sqrt{L(l_k)}$ , where  $\mathbf{h}_k, \mathbf{g}_k \in \mathbb{C}^{N_k \times M}$  and  $\mathbf{g}_k$  being a Rayleigh fading channel matrix, and  $l_k$  is the distance between BS and Rx- $k$ . The path loss function  $L(l_k)$  is given by

FIGURE 1: System model of 3-user  $2 \times 1$  broadcast channel.

$$L(l_k) = \begin{cases} l_k^\alpha, & \text{if } l_k > r_0 \\ r_0^\alpha, & \text{otherwise} \end{cases}, \quad (1)$$

where  $\alpha = 3$  denotes the path loss exponent [37–39] and the parameter  $r_0$  avoids the singularity problem when  $l_k$  is very small. Without loss of generality, we assume  $l_1 < l_2 < l_3$ , so user-1 has the least path loss. Moreover, we suppose all channels undergo additive white Gaussian noise (AWGN) with zero-mean and variance  $\sigma^2$ , where the effect of noise on the communication systems can be found in [40, 41].

Denote  $s_k^j \in \mathbb{C}$  as the  $j$ -th symbol transmitted to the  $k$ -th user, we assume  $\mathbb{E}[|s_k^j|^2] = \rho_k P$  where  $\rho_k$  is the corresponding power allocation coefficient for user- $k$  and  $P$  is the average transmit energy per time slot from BS.

To sever this multiuser channel, simply apply orthogonal multiaccess (OMA) schemes. However, OMA schemes, in general, have lower QoS flexibility than NOMA and IA techniques.

### 3. Schemes with Full CSIT

**3.1. Blind Interference Alignment.** The key component of BIA is the particular feasible channel patterns, also known as supersymbols. Table 1 shows the feasible channel patterns conducted by [23, 24], respectively. Without loss of generality, in this paper, we consider the reconfigurable-antenna-based BIA in [23]. In this scheme, each receiver node is assumed to be equipped with a reconfigurable antenna, which can switch its operating mode across time slots. Table 1(a) shows how the receive antennas switch their modes over a 4-slots supersymbol. In the table, we denote  $\mathbf{h}_k(m)$  as the channel vector when the Rx- $k$ 's antenna is on the  $m$ -th mode. Here, we use three precoding matrices  $[\mathbf{I} \mathbf{0} \mathbf{0}]^T$ ,  $[\mathbf{0} \mathbf{I} \mathbf{0}]^T$  and  $[\mathbf{0} \mathbf{0} \mathbf{I}]^T$ , for user-1, user-2, and user-3, respectively.

Consider Rx-1, the received signal through 4 slots is

$$\mathbf{y}_1 = \begin{bmatrix} \mathbf{h}_1(1) \\ \mathbf{h}_1(2) \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} s_1^1 \\ s_1^2 \end{bmatrix} + \begin{bmatrix} \mathbf{h}_1(1) \\ \mathbf{0} \\ \mathbf{h}_1(1) \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} s_2^1 \\ s_2^2 \end{bmatrix} + \begin{bmatrix} \mathbf{h}_1(1) \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{h}_1(1) \end{bmatrix} \begin{bmatrix} s_3^1 \\ s_3^2 \end{bmatrix} + \mathbf{z}_1, \quad (2)$$

where  $s_k^j$  is the  $j$ -th symbol for the  $k$ -th user and  $\mathbf{z}_1$  is the AWGN vector with covariance matrix  $\sigma^2 \mathbf{I}$ . Then, a normalized decoding matrix  $\mathbf{U}_1 = \begin{bmatrix} 1/\sqrt{3} & 0 & -(1/\sqrt{3}) & -(1/\sqrt{3}) \\ 0 & 1 & 0 & 0 \end{bmatrix}$  is used for user-1 and thus we have

$$\mathbf{U}_1 \mathbf{y}_1 = \begin{bmatrix} (1/\sqrt{3}) \mathbf{h}_1(1) \\ \mathbf{h}_1(2) \end{bmatrix} \begin{bmatrix} s_1^1 \\ s_1^2 \end{bmatrix} + \mathbf{u}_1 \mathbf{z}_1, \quad (3)$$

where  $\mathbf{U}_1 \mathbf{z}_1$  denotes the AWGN after the postprocessing and its elements are iid noise with zero-mean and variance  $\sigma^2$ .

Let  $\hat{\mathbf{H}}_k = [(1/\sqrt{3}) \mathbf{h}_k^T(1) \quad \mathbf{h}_k^T(2)]^T$ , then the achievable rate for user- $k$  is

$$R_k^{\text{BIA}} = \frac{1}{4} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \gamma \rho_k \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right) \right], \quad (4)$$

where  $\rho_k$  is the power allocation coefficient for user- $k$ ,  $\gamma = P/\sigma^2$  is the transmit SNR from BS, and  $\hat{\mathbf{H}}_k^\dagger$  is the conjugate transpose of  $\hat{\mathbf{H}}_k$ . Here, the design of one supersymbol occupying 4 time slots gives the coefficient 1/4.

Since each user's symbols are transmitted twice in a BIA supersymbol, and each supersymbol has four time slots, we have the power constraint  $2(\sum_{k=1}^K \sum_{j=1}^M |s_k^j|^2) = 2MP(\sum_{k=1}^K \rho_k) \leq 4P$ , hence  $\sum_{k=1}^K \rho_k \leq 1$ . Further, we consider a power allocation with max-min fairness

$$\begin{aligned} (P1) \quad & \max_{\rho_k (k \in [1, K])} \min R_k^{\text{BIA}} \\ \text{s.t.} \quad & C_1 : \sum_{k=1}^K \rho_k \leq 1, \\ & C_2 : \rho_k \geq 0. \end{aligned} \quad (5)$$

Note that BIA usually does not consider the power allocation problem since its implementation needs no CSIT or any feedback. The power allocation problem for BIA provides an upper bound and is used in the following performance comparison.

To deal with the complicated objective function in equation (5), we introduce an auxiliary variable  $r \geq 0$ , and equation (5) can be transformed as

$$\begin{aligned} (P2) \quad & \max_{\rho_k (k \in [1, K]), r} r \\ \text{s.t.} \quad & C_1 : R_k^{\text{BIA}} \geq r, \forall k \in [1, K]. \\ & C_2 : \sum_{k=1}^K \rho_k \leq 1, \\ & C_3 : \rho_k \geq 0, r \geq 0. \end{aligned} \quad (6)$$

Thus, problem (6) is a convex programming problem and can be solved by convex optimization tools.

**3.2. STBC-Based NOMA.** In downlink power-domain NOMA, signals for end users are superposed with different power allocated, then each receiver deploys successive

TABLE 1: Supersymbols of the 3-user  $2 \times 1$  BIA

(a) 4-slots supersymbol in [23]				
	Slot 1	Slot 2	Slot 3	Slot 4
User-1	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(1)$
User-2	$\mathbf{h}_2(1)$	$\mathbf{h}_2(1)$	$\mathbf{h}_2(2)$	$\mathbf{h}_2(1)$
User-3	$\mathbf{h}_3(1)$	$\mathbf{h}_3(1)$	$\mathbf{h}_3(1)$	$\mathbf{h}_3(2)$

(b) 4-slots supersymbol in [24]				
	Slot 1	Slot 2	Slot 3	Slot 4
User-1	$\mathbf{h}_1(1)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(2)$
User-2	$\mathbf{h}_2(1)$	$\mathbf{h}_2(2)$	$\mathbf{h}_2(2)$	$\mathbf{h}_2(2)$
User-3	$\mathbf{h}_3(1)$	$\mathbf{h}_3(1)$	$\mathbf{h}_3(1)$	$\mathbf{h}_3(2)$

interference cancellation (SIC) to decode its desired signal. Unlike in BF, in NOMA, a BS can serve multiple end users simultaneously through the power-domain scheme, even though the number of transmit antennas is less than the number of end users. Further, since two transmit antennas are available, to achieve the full transmit diversity gain, we utilize the space-time block code (STBC). Specifically, we consider the Alamouti code.

The STBC-based and superposed signals at BS is given by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^K s_k^1 \\ \sum_{k=1}^K s_k^2 \end{bmatrix}, \quad (7)$$

where  $s_k^j$  is the  $j$ -th symbol for the  $k$ -th user, with transmit power allocated by  $\rho_k P$ .

On the receiver side, for instance, the received signal at user- $k$  is

$$\begin{bmatrix} y_k(1) \\ y_k^*(2) \end{bmatrix} = \begin{bmatrix} h_k^1 & h_k^2 \\ h_k^{2*} & -h_k^{1*} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_k(1) \\ z_k^*(2) \end{bmatrix}, \quad (8)$$

where  $y_k(t)$  denotes the received signal of the  $k$ -th user at the  $t$ -th time/symbol slot,  $y^*$  denotes the conjugate of a complex number  $y$ , and  $z_k(t)$  denotes the corresponding AWGN with zero-mean and the variance of  $\sigma^2$ . Note that  $h_k^j$  is the channel coefficient from the  $j$ -th transmit antenna to the  $k$ -th user. Since  $l_1 < l_2 < l_3$  is assumed in our model and the decoding order of SIC depends on the channel quality, then the decoding order of three end users' signals is user-3, user-2, and user-1. For instance, user-1 needs to decode the signals for user-3 and user-2 successively and then subtract them from the superposed signal, before decoding its desired symbol.

In this paper we consider imperfect SIC, thus we denote a parameter  $\mu$  as the level of residual interference because of

SIC imperfection. Particularly,  $\mu = 0$  implies perfect SIC, and  $\mu = 1$  implies no SIC. The value of  $\mu$  is influenced by the type of receivers, channel characteristics, and hardware sensibility. In practice,  $\mu$  can be easily calculated at the receivers [36].

Assume  $\hat{\mathbf{H}}_k = \begin{bmatrix} h_k^1 & h_k^2 \\ h_k^{2*} & -h_k^{1*} \end{bmatrix}$ , we have  $\hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger = \hat{\mathbf{H}}_k^\dagger \hat{\mathbf{H}}_k = (|h_k^1|^2 + |h_k^2|^2) \mathbf{I}$ . Note that  $\gamma = P/\sigma^2$  is the transmit SNR, thus the covariance matrix of the desired signal vector can be derived as

$$\mathbf{V}^k = \mathbb{E} \left[ \gamma \rho_k \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right]. \quad (9)$$

Further, we can derive the covariance matrices of interference from users with decoding order larger than user- $k$  and from residual interference due to SIC imperfection, which can be formulated, respectively, as

$$\mathbf{V}_{SIC}^{(k)} = \mathbb{E} \left[ \gamma \left( \sum_{i=1}^{k-1} \rho_i \right) \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right] = \mathbb{E} \left[ \gamma \rho_k^{SIC} \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right], \quad (10)$$

$$\mathbf{V}_{imp}^{(k)} = \mathbb{E} \left[ \gamma \left( \mu \sum_{i=k+1}^K \rho_i \right) \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right] = \mathbb{E} \left[ \gamma \rho_k^{imp} \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger \right]. \quad (11)$$

Note that  $\rho_k^{SIC} = 0$  when  $k = 1$  and  $\rho_k^{imp} = 0$  when  $k = K$ .

Recall the Alamouti scheme requires 2 time slots, the achievable rate per time slot of user- $k$  is derived as

$$\begin{aligned} R_k^{S\text{-NOMA}} &= \frac{1}{2} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \frac{\gamma \rho_k \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger}{\mathbf{I} + \gamma \rho_k^{SIC} \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger + \gamma \rho_k^{imp} \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger} \right) \right] \\ &= \frac{1}{2} \mathbb{E} \left[ \log \det \left( \frac{\mathbf{I} + \gamma (\rho_k + \rho_k^{SIC} + \rho_k^{imp}) \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger}{\mathbf{I} + \gamma (\rho_k^{SIC} + \rho_k^{imp}) \hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^\dagger} \right) \right] \\ &= \frac{1}{2} \mathbb{E} \left[ \log \left( \frac{1 + \gamma (\rho_k + \rho_k^{SIC} + \rho_k^{imp}) (|h_k^1|^2 + |h_k^2|^2)}{1 + \gamma (\rho_k^{SIC} + \rho_k^{imp}) (|h_k^1|^2 + |h_k^2|^2)} \right)^2 \right] \\ &= \mathbb{E} \left[ \log \left( 1 + \frac{\gamma \rho_k (|h_k^1|^2 + |h_k^2|^2)}{1 + \gamma (\rho_k^{SIC} + \rho_k^{imp}) (|h_k^1|^2 + |h_k^2|^2)} \right) \right] \\ &= \mathbb{E} \left[ \log \left( 1 + \frac{\mathcal{F}_k}{\mathcal{G}_k} \right) \right], \end{aligned} \quad (12)$$

By applying similar transformation in equation (6), the power allocation problem for STBC-based NOMA is formulated as

$$\begin{aligned}
(P3) \quad & \max_{\rho_k(k \in [1, K]), r} \quad r \\
s.t. \quad & C_1 : \frac{\mathcal{F}_k}{\mathcal{G}_k} \geq r, \forall k \in [1, K], \\
& C_2 : \sum_{k=1}^K \rho_k \leq 1, \\
& C_3 : \rho_k \geq 0, r \geq 0.
\end{aligned} \tag{13}$$

Problem (13) is intractable due to the nonconvexity of  $C_1$ , but can be transformed into geometric programming (GP) [42]. Here, we introduce a variable  $t = 1/r$ , thus problem (13) can be transformed as

$$\begin{aligned}
(P4) \quad & \min_{\rho_k(k \in [1, K]), t} \quad t \\
s.t. \quad & C_1 : \frac{\mathcal{G}_k}{\mathcal{F}_k} \leq t, \forall k \in [1, K], \\
& C_2 : \sum_{k=1}^K \rho_k \leq 1, \\
& C_3 : \rho_k \geq 0, t \geq 0.
\end{aligned} \tag{14}$$

Since  $\mathcal{G}_k$  is a posynomial and  $\mathcal{F}_k$  is a monomial,  $\mathcal{G}_k/\mathcal{F}_k$  should be a posynomial. Then problem (14) is GP and can be solved by convex programming [42].

**3.2.1. NOMA-Based Beamforming.** In this scheme, we consider full CSIT and employ the conventional BF method. Note that the maximum achievable DoF in this system is 2, which is smaller than the number of end users. In this model, traditional zero-forcing-based beamforming (ZF-BF) cannot be well applied, since the whole signal space cannot provide interference-free transmission for all end users simultaneously. Other BF schemes, such as BF based on the signal-to-leakage-noise ratio (SLNR) [19], are declared to be feasible when  $M < K$ . However, in this model, an overloaded user causes much interference leakage, hence significant performance declines especially when considering a max-min fair power allocation. So, inspired by [43], we use a singular value decomposition- (SVD-) based BF scheme to serve some end users and use power-domain NOMA simultaneously to serve the remaining end users.

In this joint scheme, we suppose a simple NOMA pairing scheme has been adopted for user clustering. Assume that user-1 experiences the best channel and user-3 experiences the worst, then we collect user-1 and user-3 to form a NOMA pair. By the joint scheme, we use BF to cancel the intercluster interference and suppress the intracluster interference by using NOMA.

Let  $s_k \in \mathbb{C}$  be the scheduled symbol for the  $k$ -th user, the precoded signal at the BS is

$$\mathbf{x} = \mathbf{v}_1(s_1 + s_3) + \mathbf{v}_2 s_2, \tag{15}$$

where  $\mathbf{v}_k \in \mathbb{C}^{M \times 1}$  denotes the transmit BF vector. By applying SVD for  $\mathbf{h}_2$ , we have  $\mathbf{h}_2 = \mathbf{S}_2 \mathbf{A}_2 \mathbf{D}_2^H$ , where  $\mathbf{A}_2$  contains the singular values of  $\mathbf{h}_2$ . Besides, the columns of  $\mathbf{S}_2$  and  $\mathbf{D}_2$  are the left-singular vectors and right-singular vectors.

We then choose a right-singular vector  $\mathbf{v}_1$  in  $\mathbf{D}_2$  corresponding to the zero singular value in  $\mathbf{A}_2$ . Thus  $\mathbf{h}_2 \mathbf{v}_1 = 0$  is achieved to eliminate interuser interference. Similarly, we can choose  $\mathbf{v}_2$  by applying SVD of  $\mathbf{h}_1$  and thus  $\mathbf{h}_1 \mathbf{v}_2 = 0$ . Note that in the NOMA-based BF scheme, the channel information of user-1 is chosen for deriving  $\mathbf{v}_2$ , so there is no interference from user-2 to user-1, but the interference leakage from user-2 to user-3 is unavoidable.

In this section, we assume  $\mathbb{E}[|s_k|^2] = \rho_k P$ . For user-2, interference signals due to the NOMA pair are canceled by BF, hence the SINR is

$$\text{SINR}_2^{\text{NOMA-BF}} = \gamma \rho_2 |\mathbf{h}_2 \mathbf{v}_2|^2. \tag{16}$$

For user-1, BF is used to suppress the interference from user-2. User-1 first decodes the symbols for user-3, then subtracts them to get its desired signal  $s_1$ . Considering imperfect SIC, the SINR is

$$\text{SINR}_1^{\text{NOMA-BF}} = \frac{\gamma \rho_1 |\mathbf{h}_1 \mathbf{v}_1|^2}{1 + \mu \gamma \rho_3 |\mathbf{h}_1 \mathbf{v}_1|^2}. \tag{17}$$

For user-3, after considering the interference from both  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , the SINR is given by

$$\text{SINR}_3^{\text{NOMA-BF}} = \frac{\rho_3 |\mathbf{h}_3 \mathbf{v}_1|^2}{\rho_2 |\mathbf{h}_3 \mathbf{v}_2|^2 + \rho_1 |\mathbf{h}_3 \mathbf{v}_1|^2 + (1/\gamma)}. \tag{18}$$

Since no symbol extensions are required in NOMA-based beamforming, the average achievable rate of user- $k$  is denoted as

$$R_k^{\text{NOMA-BF}} = \mathbb{E}[\log(1 + \text{SINR}_k)]. \tag{19}$$

Similarly in (6), an auxiliary variable  $t$  is introduced to formulate a max-min fairness problem as

$$\begin{aligned}
(P5) \quad & \min_{\rho_k(k \in [1, K]), t} \quad t \\
s.t. \quad & C_1 : \frac{1}{\text{SINR}_k^{\text{NOMA-BF}}} \leq t, \forall k \in [1, K], \\
& C_2 : \sum_{k=1}^3 \rho_k \leq 1, \\
& C_3 : \rho_k, t \geq 0,
\end{aligned} \tag{20}$$

where  $C_2$  is provided to meet the transmit power constraint. Since the precoders  $\mathbf{v}_k$  have been normalized, therefore  $\mathbb{E}[\mathbf{x}] = \sum_k \rho_k P \leq P$ , and hence  $C_2$  is derived.

Note that  $1/\text{SINR}_k^{\text{NOMA-BF}}$  are posynomials, thus equation (20) is a GP problem.

#### 4. Schemes with Partial CSIT

As shown in Sect.(III-A) and Sect.(III-B), the formulated power allocation problems require CSIT to solve. However, in practice, perfect CSIT is hard to acquire. A more general

situation is utilizing partial CSIT for power allocation, e.g., path loss information or statistical information of channels. Here, assume an LoS path loss channel model in power allocation. That is,

$$\mathbf{h}_k = \frac{\mathbf{1}_{N_k \times M}}{\sqrt{L(l_k)}}, \quad (21)$$

where  $\mathbf{1}_{N_k \times M}$  is a  $N_k \times M$  matrix with each element be 1.

Although we neglect the small-scale fading, the corresponding optimized power allocation is still useful for the optimization problem under the Rayleigh fading channel.

**4.1. Power Allocation for BIA.** By substituting the LoS model (21) into (4), we have the achievable rate of the  $k$ -th user in BIA expressed by

$$\begin{aligned} \dot{R}_k^{\text{BIA}} &= \frac{1}{4} \log \det \left( \mathbf{I} + \rho_k \gamma l_k^{-\alpha} \begin{bmatrix} \frac{1}{\sqrt{3}} \mathbf{1}_{1,2} \\ \mathbf{1}_{1,2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} \mathbf{1}_{2,1} & \mathbf{1}_{2,1} \end{bmatrix} \right) \\ &= \frac{1}{4} \log \left( 1 + \frac{8}{3} \rho_k \gamma l_k^{-\alpha} \right). \end{aligned} \quad (22)$$

Assume  $\text{sinr}_k^B = (8/3)\rho_k \gamma l_k^{-\alpha}$ , the fairness power allocation problem is

$$\begin{aligned} (P6) \quad & \max_{\rho_k(k \in [1, K]), r} \quad r \\ \text{s.t.} \quad & C_1 : \text{sinr}_k^B \geq r, \forall k \in [1, K], \\ & C_2 : \sum_{k=1}^K \rho_k \leq 1, \\ & C_3 : 0 \leq \rho_k. \end{aligned} \quad (23)$$

Note that (23) is a linear programming problem and can be easily solved by the Lagrange dual theory. By applying the Karush-Kuhn-Tucker (KKT) condition, we can get the optimal power allocation by

$$\rho_k = \frac{l_k^\alpha}{\sum_{i=1}^K l_i^\alpha}, \forall k \in [1, K] \quad (24)$$

**4.2. Power Allocation for STBC-NOMA.** Assuming the LoS model in (21), the covariance matrix of the desired signal vector for the  $k$ -th user in STBC-NOMA is

$$\mathbf{V}^k = 2l_k^{-\alpha} \rho_k \gamma \mathbf{I}, \quad (25)$$

where

$$\mathbf{V}_{\text{SIC}}^{(k)} = 2l_k^{-\alpha} \left( \sum_{i=1}^{k-1} \rho_i \right) \gamma \mathbf{I} = a^{(k)} \mathbf{I}, \quad (26)$$

$$\mathbf{V}_{\text{imp}}^{(k)} = 2l_k^{-\alpha} \mu \left( \sum_{i=k+1}^K \rho_i \right) \gamma \mathbf{I} = b^{(k)} \mathbf{I}. \quad (27)$$

Thus the achievable rate is

$$R_k^{\text{S-NOMA}} = \log \left( 1 + \frac{2l_k^{-\alpha} \rho_k \gamma}{1 + a^{(k)} + b^{(k)}} \right) = \log \left( 1 + \text{sinr}_k^{\text{S-NOMA}} \right) \quad (28)$$

Similarly to the problem in (14), the power allocation problem can be formulated as

$$\begin{aligned} (P7) \quad & \min_{\rho_k(k \in [1, K]), t} \quad t \\ \text{s.t.} \quad & C_1 : \frac{1}{\text{sinr}_k^B} \leq t, \forall k \in [1, K], \\ & C_2 : \sum_{k=1}^K \rho_k \leq \frac{1}{2}, \\ & C_3 : \rho_k \geq 0, t \geq 0, \end{aligned} \quad (29)$$

which is a GP problem.

## 5. Simulation Results and Discussion

In this section, all the schemes mentioned previously are compared via numerical simulation. Suppose  $l_1 = 4$ ,  $l_2 = 5$ , and  $l_3 = 6$ . Since in this paper, we consider end users with different distances to the BS, and the channel quality is related directly to the distance, we adopt the NOMA clustering scheme according to the distance. For example, since  $l_1 < l_2 < l_3$ , the decoding order in SIC of STBC-NOMA will be user-3, user-2, and then user-1, while in NOMA-BF, user-3 will be paired with user-1.

**5.1. Schemes with Perfect CSIT.** Figure 2 shows the ergodic achievable sum-rate for the three schemes with perfect SIC in linear scale and logarithmic scale, respectively. Note that in this section, perfect CSIT is considered for power allocation. The figures show that, under the low and medium SNR scenarios, the STBC-NOMA scheme outperforms other schemes, while BIA is better than NOMA-BF. The reasons are as follows: (1) the rate enhancement yielded by the diversity gain of STBC-NOMA is larger than the contributions of multiplexing gain that the other two schemes obtain, (2) BIA sacrifices some DoF for its fairness among three end users, (3) In NOMA-BF, because of the fairness-oriented max-min rate objective, the user with the worst channel state is allocated with much more power, which reduces the sum-rate at the low-SNR scenario.

Under the high SNR scenario, STBC-NOMA performs the worst because it offers no multiplexing gain. By contrast, NOMA-BF outperforms the other two schemes. This can be explained by the DoF/multiplexing-gain difference. In the system model considered, the sum DoF achieved by NOMA-BF is 2, the achievable sum DoF of BIA is 3/2, and STBC-NOMA only has a sum DoF 1. Interestingly, applying

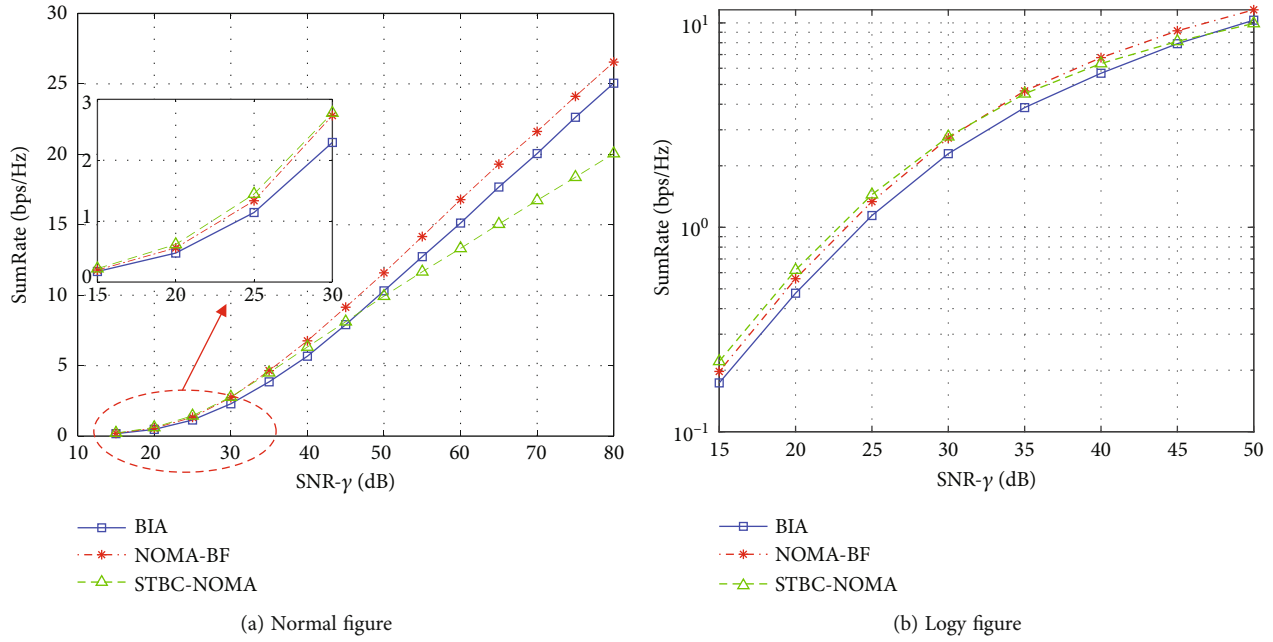


FIGURE 2: Sum-rate comparison with perfect SIC (normal graph).

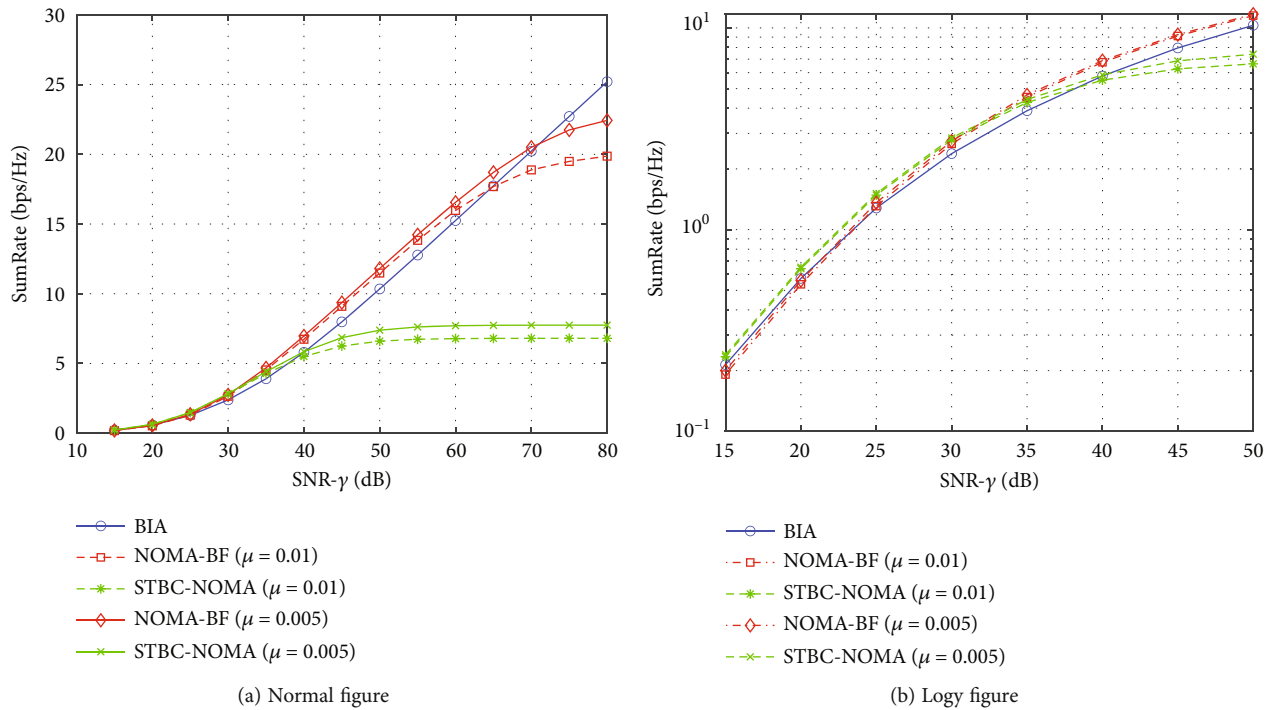


FIGURE 3: Sum-rate comparison with imperfect SIC.

PA with perfect CSIT, BIA can deliver decent performances under both low and high SNR scenarios.

Figure 3 shows the ergodic achievable sum-rate consider imperfect SIC in NOMA-based schemes. From the figure we observe the following: (1) compared with systems with perfect SIC, there is performance loss on NOMA-BF and STBC-NOMA with imperfect SIC under the high SNR scenarios. The reason is that residual interference

on SIC is proportional to SNR, thus leading to a significant performance loss on the sum-rate under high SNR. (2) The imperfection of SIC causes much performance loss on STBC-NOMA than on NOMA-BF. This can be explained by the different levels of residual interference in the two schemes. In NOMA-BF, only two users are clustered as a NOMA pair and apply SIC for intercluster interference cancellation. While in STBC-NOMA, all users are clustered and SIC is required for

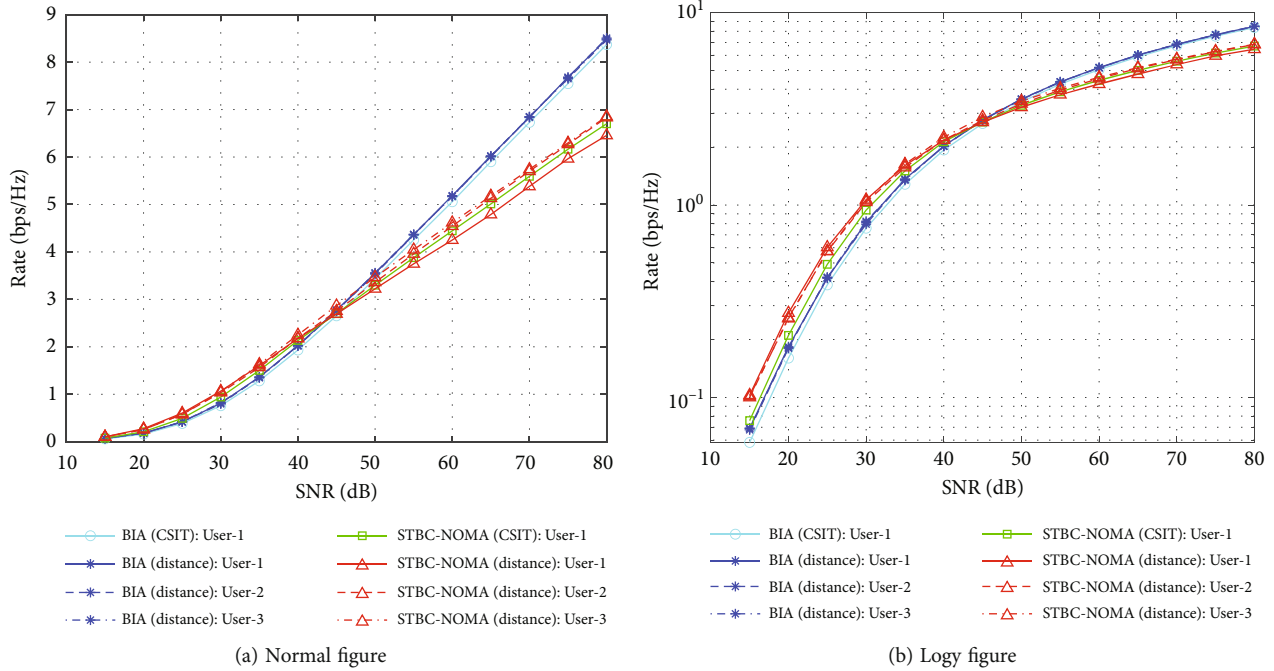


FIGURE 4: PA with CSIT vs. PA with distance (perfect SIC).

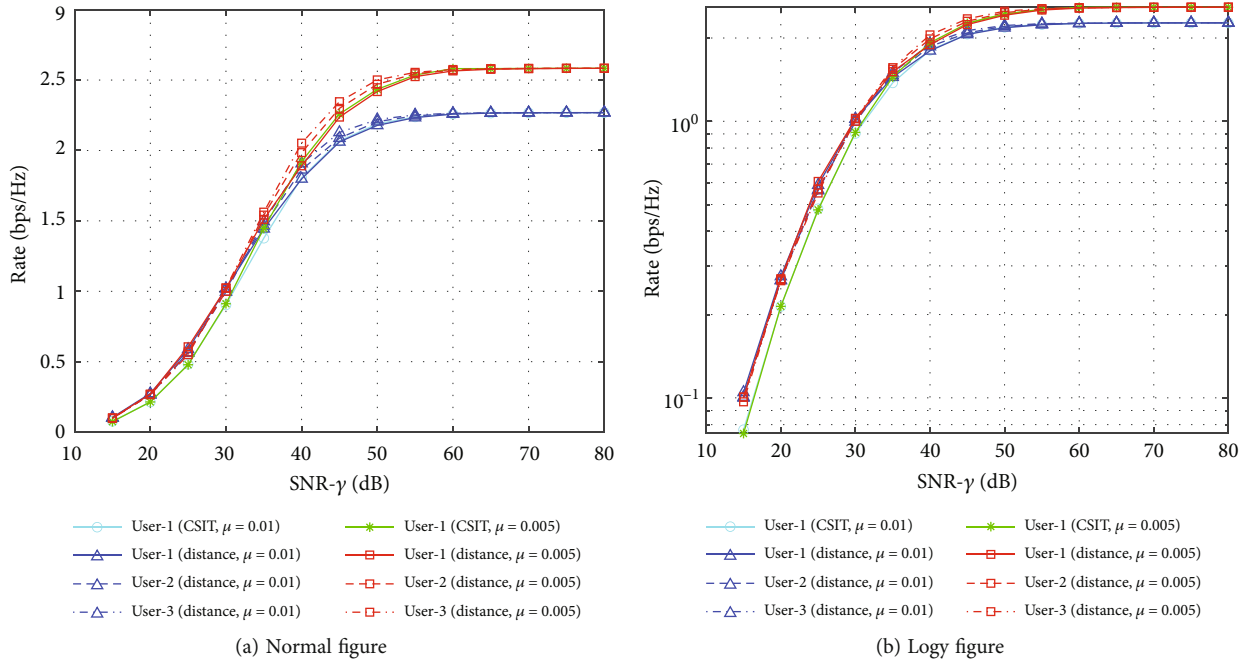


FIGURE 5: PA with CSIT vs. PA with distance (in STBC-NOMA with imperfect SIC).

canceling interference among all users. (3) Under low SNR scenarios, STBC-NOMA achieves the highest sum-rate as indicated in Figure 2.

**5.2. Schemes with Partial CSIT.** In this section, PA considering only path loss information is compared with PA considering perfect CSIT in BIA and STBC-NOMA. Note that NOMA-BF is not considered in this section, because with only path loss information available, the precoder in NOMA-BF cannot be

fully functional. Since PA with perfect CSIT can yield fair rates among users, for simplicity, we only chose user-1 considering perfect CSIT for comparison in this section.

In Figure 4, PA with perfect CSIT and with only distance information available are compared, with perfect SIC considered in NOMA. From the figure, we observe that the proposed PA with only distance information available has unequal ergodic rates among users, especially in the high SNR region. However, its performance is acceptable since

the absence of accurate CSIT and low computation complexity. The performance of BIA with perfect CSIT is very close to that of BIA with path loss information only.

Figure 5 compares the PA with perfect CSIT and with only distance information available in STBC-NOMA. In the medium SNR region, PA with only distance available achieves unequal rates among users. However, as SNR increases, equal rates among users are achieved by considering only distance information, which is the same as performance when considering accurate CSIT. The reason is that imperfect SIC provides the same rate constraint in high SNR region for each user, thus users achieve equal rates as SNR increases. Results presented in Figure 5 show that in the high SNR region, it is possible to achieve max-min fairness optimization with only distance information available when imperfect SIC is considered.

## 6. Conclusion

In this paper, we consider scenarios when the number of BS transmit antennas is less than the number of end users. Three feasible schemes have been investigated, i.e., BIA, STBC-NOMA, and NOMA-BF. Under PA of a max-min fairness objective, the sum-rates of the three schemes have been derived with the consideration of both perfect and imperfect SIC. Then, the formulated PA problems are solved assuming perfect CSIT is available as well as assuming only path loss information is available. Numerical simulations show that partial channel information is adequate, which can be used as a criterion in system or scheme designs. Outcomes of this study serve as guidelines for system integrators to pick the best approach for their systems. The formulations also serve as benchmarks and an evaluation framework for up-and-coming designs. For future research, different constraints/objectives can be considered in the PA. Nevertheless, the benefits of integrating NOMA and BIA should be further investigated.

## Data Availability

No underlying data was collected or produced in this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by NSFC under Grant 61971138, in part by the Science and Technology Planning Project of Guangdong Province under Grant 2016B010108002, in part by Guangdong Higher Education Innovation Project under Grant 2020ZDZX3047, and in part by the Applied Basic Research Project of Guangdong Province under Grant No. 2019A151511149.

## References

- [1] X. Lai, Y. Deng, G. K. Karagiannidis, and A. Nallanathan, "Secure mobile edge computing networks in the presence of multiple eavesdroppers," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 500–513, 2022.
- [2] X. Lai, J. Xia, L. Fan, T. Q. Duong, and A. Nallanathan, "Outdated access point selection for mobile edge computing with cochannel interference," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7445–7455, 2022.
- [3] L. He, K. He, L. Fan, X. Lei, A. Nallanathan, and G. K. Karagiannidis, "Toward optimally efficient search with deep learning for large-scale MIMO systems," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3157–3168, 2022.
- [4] J. Lu and M. Tang, "Performance analysis for IRS-assisted MEC networks with unit selection," *Physical Communication*, vol. 52, article 101869, 2022.
- [5] R. Zhao and M. Tang, "Impact of direct links on intelligent reflect surface-aided MEC networks," *Physical Communication*, vol. 50, article 101905, 2022.
- [6] L. Chen, "Physical-layer security on mobile edge computing for emerging cyber physical systems," *Computer Communications*, vol. 194, pp. 180–188, 2022.
- [7] Y. Guo, R. Zhao, S. Lai, L. Fan, X. Lei, and G. K. Karagiannidis, "Distributed machine learning for multiuser mobile edge computing systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 3, pp. 460–473, 2022.
- [8] L. Zhang and C. Gao, "Deep reinforcement learning based IRS-assisted mobile edge computing under physical-layer security," *Physical Communication*, vol. 51, no. 1, article 101896, 2022.
- [9] L. Zhang, W. Zhou, J. Xia et al., "DQN-based mobile edge computing for smart internet of vehicle," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, 45 pages, 2022.
- [10] S. Tang, L. Chen, K. He, J. Xia, L. Fan, and A. Nallanathan, "Computational intelligence and deep learning for next-generation edge-enabled industrial IoT," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1–13, 2022.
- [11] S. Tang, "Dilated convolution based CSI feedback compression for massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 211–216, 2022.
- [12] Y. Wu and C. Gao, "Task offloading for vehicular edge computing with imperfect CSI: a deep reinforcement approach," *Physical Communication*, vol. 55, no. 1, article 101867, 2022.
- [13] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.
- [14] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multi-user transmit beamforming: a difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, 2014.
- [15] S. Ali, M. Sohail, S. B. H. Shah et al., "New trends and advancement in next generation mobile wireless communication (6G): a survey," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9614520, 14 pages, 2021.
- [16] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: a survey," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3920–3931, 2017.
- [17] I. Ahmed, H. Khammari, A. Shahid et al., "A survey on hybrid beamforming techniques in 5G: architecture and system model perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3060–3097, 2018.



- [18] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1065–1069, 2019.
- [19] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, 2007.
- [20] C. S. Vaze and M. K. Varanasi, "The degrees of freedom region of the two-user MIMO broadcast channel with delayed CSIT," in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 199–203, St. Petersburg, Russia, 2011.
- [21] C. Wang, D. Deng, L. Xu, W. Wang, and F. Gao, "Joint interference alignment and power control for dense networks via deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 966–970, 2021.
- [22] Z. Li, J. Chen, L. Zhen, S. Cui, K. G. Shin, and J. Liu, "Coordinated multi-point transmissions based on interference alignment and neutralization," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3347–3365, 2019.
- [23] T. Gou, C. Wang, and S. A. Jafar, "Aiming perfectly in the dark-blind interference alignment through staggered antenna switching," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2734–2744, 2011.
- [24] Q. F. Zhou, Q. Zhang, and F. C. Lau, "Diophantine approach to blind interference alignment of homogeneous K-user 2x1 MISO broadcast channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2141–2153, 2013.
- [25] M. Johnny and M. R. Aref, "BIA for the k-user interference channel using reconfigurable antenna at receivers," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2184–2197, 2019.
- [26] M. Morales-Céspedes, O. A. Dobre, and A. Garcia-Armada, "Semi-blind interference aligned NOMA for downlink MU-MISO systems," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1852–1865, 2019.
- [27] Q. F. Zhou, A. Huang, M. Peng, F. Qu, and L. Fan, "On the mode switching of reconfigurable-antenna-based blind interference alignment," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 6958–6968, 2017.
- [28] J. Wu, X. Liu, C. Qu, C.-T. Cheng, and Q. Zhou, "Balanced-switching-oriented blind interference-alignment scheme for 2-user MISO interference channel," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2324–2328, 2020.
- [29] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4438–4454, 2016.
- [30] Z. Ding, "Harvesting devices' heterogeneous energy profiles and QoS requirements in IoT: WPT-NOMA vs BAC-NOMA," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 2837–2850, 2021.
- [31] P. Lai, Q. He, G. Cui et al., "Cost-effective user allocation in 5G NOMA-based mobile edge computing systems," *IEEE Transactions on Mobile Computing*, p. 1, 2021.
- [32] Y. Dai, Z. Liang, L. Lyu, and B. Lin, "Deep reinforcement learning-based UAV data collection and offloading in NOMA-enabled marine IoT systems," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 8805416, 13 pages, 2022.
- [33] J. Zhang, L. Zhu, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "Optimal and sub-optimal uplink NOMA: joint user grouping, decoding order, and power control," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 254–257, 2020.
- [34] D. Hu, Q. Zhang, Q. Li, and J. Qin, "Joint position, decoding order, and power allocation optimization in UAV-based NOMA downlink communications," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2949–2960, 2020.
- [35] M. Zeng, W. Hao, O. A. Dobre, Z. Ding, and H. V. Poor, "Power minimization for multi-cell uplink NOMA with imperfect SIC," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2030–2034, 2020.
- [36] A. S. de Sena, F. R. M. Lima, D. B. da Costa et al., "Massive MIMO-NOMA networks with imperfect SIC: design and fairness enhancement," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6100–6115, 2020.
- [37] J. Lu, L. Chen, J. Xia et al., "Analytical offloading design for mobile edge computing based smart internet of vehicle," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, no. 1, p. 10, 2022.
- [38] R. Zhao and M. Tang, "Profit maximization in cache-aided intelligent computing networks," *Physical Communication*, vol. 99, pp. 1–10, 2022.
- [39] S. Tang and X. Lei, "Collaborative cache-aided relaying networks: performance evaluation and system optimization," *IEEE Journal on Selected Areas in Communications*, vol. 99, pp. 1–12, 2022.
- [40] L. Chen, R. Zhao, K. He, Z. Zhao, and L. Fan, "Intelligent ubiquitous computing for future UAV-enabled MEC network systems," *Cluster Computing*, vol. 25, pp. 1–11, 2021.
- [41] K. He, L. He, L. Fan, X. Lei, Y. Deng, and G. K. Karagiannis, "Efficient memory-bounded optimal detection for GSM-MIMO systems," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4359–4372, 2022.
- [42] M. Chiang, "Geometric programming for communication systems," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 1–2, pp. 1–154, 2005.
- [43] B. Kimy, S. Lim, H. Kim et al., "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *MILCOM 2013-2013 IEEE Military Communications Conference*, pp. 1278–1283, San Diego, CA, USA, 2013.

## Research Article

# Optimal Element Allocation for RIS-Aided Physical Layer Security

Ying Zhang <sup>1</sup>, Guoan Zhang <sup>1</sup>, Siyu Chen,<sup>1</sup> Jaeho Choi,<sup>2</sup> and Pin-Han Ho<sup>3</sup>

<sup>1</sup>School of Information Science and Technology, Nantong University, Nantong 226019, China

<sup>2</sup>Department of Electronic Engineering, Chonbuk National University, Jeonju 54896, Republic of Korea

<sup>3</sup>Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Correspondence should be addressed to Guoan Zhang; [g Zhang@ntu.edu.cn](mailto:g Zhang@ntu.edu.cn)

Received 26 July 2022; Accepted 8 September 2022; Published 21 September 2022

Academic Editor: Yingyang Chen

Copyright © 2022 Ying Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we investigate an innovative physical layer security (PLS) scheme for an uncertain reconfigurable intelligent surface (RIS-) assisted communication system under eavesdropping attack. Specifically, in our proposed system, we consider that the uncertain RIS is known to the legitimate user while not to the eavesdropper (Eve). In this manner, the reflective elements of the uncertain RIS are adjusted to strengthen the keen signals for the legitimate user while suppressing the signals for Eve via jamming. We analyze the system by assuming legitimate and wiretap channels, respectively, where the secrecy capacity maximization problem is formulated and its exact closed-form expression is derived. Simulation results verifying the accuracy of our analysis demonstrate the validity and superiority of the uncertain RIS-assisted communication system against its counterparts.

## 1. Introduction

Due to the fast development of the mobile communication network, the continuous emergence of new services, and the fulminic growth of information interaction, the fifth-generation (5G) wireless communications are anticipated to provide massive data transmissions and wide-area coverage as a new type of communication network [1]. For the upcoming applications of 5G, future wireless communication systems are anticipated to support more efficient, lower latency, more reliable, and secure designs. With the rapid evolution of science, lots of technologies have been implemented in 5G communication network to satisfy its diverse application scenarios, e.g., multiple access technology [2], deep learning (DL) [3], and beamforming [4]. Meanwhile, due to the explosive growth of information dissemination and massive information interaction, public communication networks and private communication networks are often mixed, which makes information security issues more important [5].

Physical layer security (PLS) enhancement has aroused extensive attentions in both industry and academia thanks to its superb capability of light-weight authentication on received signals against eavesdropping attacks [6]. Since

the PLS takes advantage of the natural peculiarities and physical properties of propagation environment to promote the security performance of 5G communication network and secure the data transmission, it has attracted extensive attention in the academia and industry circles [7]. Several dimensions of designing PLS schemes have been identified in the literature, via nonorthogonal multiple access (NOMA) [8], artificial noise (AN), and cooperative interference [9]. Although effective, they may consume extra power and/or incur additional hardware cost, thus forming significant barriers in the real implementations.

Fortunately, with the rapid evolution of metamaterial technology, a low cost and energy efficient device named reconfigurable intelligent surface (RIS) offers an available way to strengthen the PLS [10]. The RIS is an artificial electromagnetic surface consisting a great many cheap passive reflecting elements governed by a preprogrammed controller, which can skillfully change the amplitude and/or phase shift of arriving galvanomagnetic signals to make the direction and strength of the signal greatly controllable at the destination. Thus, the RIS cannot only strengthen the required signals and restrain the undesirable signals by befittingly tuning the phase shifts of all reflecting components but also

effectively improves the radio transmission environment leading to the random channel state information (CSI) controllable [11, 12]. Therefore, it can be anticipated that the RIS will occupy an important position in the 5G wireless communication network [13].

With such characteristics, RIS has recently been considered to enable the PLS of wireless communication system to promote the secrecy capacity [14–18]. In [14], the minimum secrecy rate maximization problem of one downlink multiple-input single-output (MISO) broadcasting communication model was formulated and optimized via alternating optimization (AO) and path tracing algorithm. The authors in [15] introduced an RIS-based secure transmission framework aiming to obtain the minimum value of its communication network energy consumption within the scenarios of two different access point (AP-) RIS channel models. In [16], the joint optimization of transmitter beamforming and RIS phase shifts by block coordinate descending (BCD) and majorization-minimization (MM) methods was studied to maximize the system security capacity for the secure wireless communication assisted with RIS. Cui et al. [17] considered a stronger eavesdropping channel than the legal one, where the AO and semidefinite relaxation (SDR) algorithms were explored to derive the maximum value of the security capacity in the secure wireless communication system. Wang et al. [18] investigated an RIS-aided MISO communication network with the unknown eavesdropping CSI, where the oblique manifold (OM) and MM algorithms were designed to promote the security of the communication network.

Note that the above mentioned studies [14–18] simply focused their attention on one function of RIS, i.e., enhancing or jamming, but the unknown situation of RIS was not considered. This paper investigates RIS-assisted secure wireless communication with the main contributions listed as follows:

- (i) We propose a novel PLS communication system with an uncertain RIS aiming to enhance the secrecy rate, where the components of RIS are separated into two functional parts, one assisting the legitimate user and the other suppressing the achievable rate of eavesdropper as a jammer.
- (ii) The optimal result of elements allocation for RIS is analyzed, and its exact closed-form expression is concluded to confirm the superiority of our proposed scheme.
- (iii) Numerical results verifying the correctness of the derived expression show the advantage and validity of the designed scheme compared with its counterparts, especially under a great many reflecting elements.

## 2. System Model

An uncertain RIS-assisted secure communication system is considered, which consists of a source ( $S$ ), a legitimate user ( $B$ ), an eavesdropper (Eve), and an RIS, as depicted in Figure 1. To improve the secure wireless communication, the RIS with  $N$  reflecting elements is placed near by  $S$ ,  $B$ , and Eve which are equipped with a single antenna. For the

uncertain RIS, the size of each element is considered a lot smaller than the wavelength of RF signals resulting in that the reflecting element scatters the received signal in whole directions with approximately constant gain [19], and the phase shift of each passive units is controlled dynamically by RIS via an intelligent controller. It is also supposed that both  $S$  and  $B$  have the knowledge of the location information of RIS, while Eve is unable to acquire the location information of RIS, and that the CSI of the uncertain RIS is entirely known to  $S$  and  $B$ , but not for the Eve.

The direct links of  $S \rightarrow B$  and  $S \rightarrow$  Eve are considered, which are expressed by  $h_{SB} \in \mathbb{C}$  and  $h_{SE} \in \mathbb{C}$ , respectively. Considering that the number of components is fixed to  $N$ , we divide the surface of RIS into two sectors, in which the one with  $\eta$  elements is used to serve as a jammer to Eve and the other with  $(N - \eta)$  elements is applied to promote the information signals from  $S$  to  $B$  via high-quality virtual links. The source transmits the signal to legitimate user via the RIS to promote the information transmission for the legitimate user, and the channel coefficients of the links  $S \rightarrow$  RIS and  $RIS \rightarrow B$  are expressed by  $\mathbf{h}_{SR_1} \in \mathbb{C}^{\eta \times 1}$  and  $\mathbf{h}_{R_1B} \in \mathbb{C}^{\eta \times 1}$ , respectively. Since the eavesdropper attempts to wiretap the signal from  $S$ , it is assumed that the eavesdropper does not know the location of RIS, and hence, the reflected signals from RIS will be treated as jamming signals for the Eve, with the available links  $S \rightarrow$  RIS and  $RIS \rightarrow$  Eve denoted as  $\mathbf{h}_{SR_2} \in \mathbb{C}^{(N-\eta) \times 1}$  and  $\mathbf{h}_{R_2E} \in \mathbb{C}^{(N-\eta) \times 1}$ , respectively. Furthermore, it is assumed that each channel of the system follows independent Rayleigh fading distribution. Thus, the received signals at  $B$  and Eve are separately expressed as

$$y_B = \sqrt{P} \left( h_{SB} + \mathbf{h}_{R_1B}^H \mathbf{\Theta}_1 \mathbf{h}_{SR_1} \right) s + n_1, \quad (1)$$

$$y_E = \sqrt{P} h_{SE} s + \sqrt{P} \mathbf{h}_{R_2E}^H \mathbf{\Theta}_2 \mathbf{h}_{SR_2} s + n_2, \quad (2)$$

where  $s$  and  $P$  mean the data symbol and the transmit power for  $S$  and  $\mathbb{E}[|s|^2] = 1$ .  $\mathbf{\Theta}_1 = \alpha \text{diag} (e^{j\psi_1}, \dots, e^{j\psi_{N-\eta}})$  and  $\alpha \in (0, 1]$  stand for the channel coefficient matrix of the RIS sector serving  $B$  and the amplitude of the reflection coefficient, where  $\alpha = 1$  represents the lossless reflection;  $\psi_i \in [0, 2\pi)$  is the phase-shift produced by the  $i$ -th element for  $i = 1, \dots, N - \eta$ .  $\mathbf{\Theta}_2 = \alpha \text{diag} (e^{j\phi_1}, \dots, e^{j\phi_\eta})$  stands for the channel coefficients matrix of the RIS sector allocated for Eve,  $\phi_j \in [0, 2\pi)$  is the phase-shift produced by the  $j$ -th element for  $j = 1, \dots, \eta$ . In addition,  $n_1$  and  $n_2 \sim \mathcal{CN}(0, \sigma^2)$  are the additive white Gaussian noise (AWGN).

Since the signal from RIS will be treated as interference by the Eve because of the unknown CSI of RIS, the achievable SNR at  $B$  and SINR at Eve can be, respectively, obtained from

$$\gamma_B = \frac{P}{\sigma^2} \left| h_{SB} + \mathbf{h}_{R_1B}^H \mathbf{\Theta}_1 \mathbf{h}_{SR_1} \right|^2, \quad (3)$$

$$\gamma_E = \frac{P |h_{SE}|^2}{P \left| \mathbf{h}_{R_2E}^H \mathbf{\Theta}_2 \mathbf{h}_{SR_2} \right|^2 + \sigma^2}. \quad (4)$$

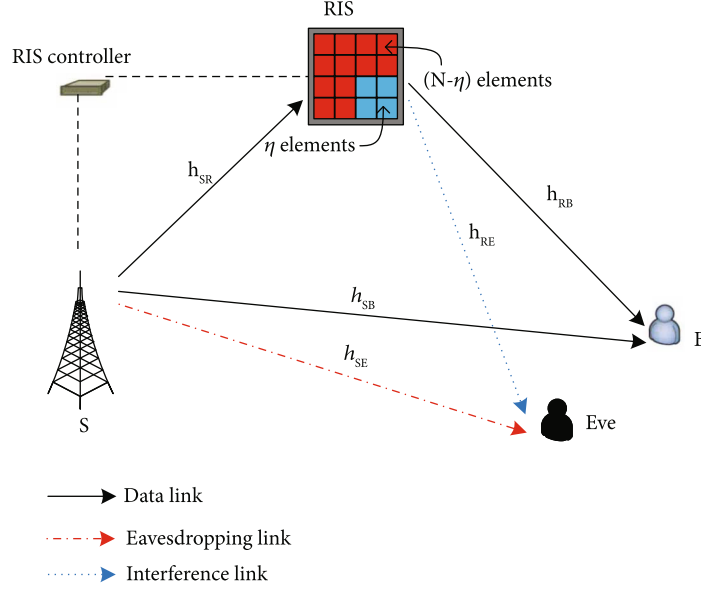


FIGURE 1: The uncertain RIS-assisted PLS system.

For any given  $\Theta_1$  and  $\Theta_2$ , the data rate expressions for  $B$  and Eve can be achieved from the Shannon theory as  $R_B = \log_2(1 + \gamma_B)$  and  $R_E = \log_2(1 + \gamma_E)$ .

From above observations, the secrecy capacity (SC) in our proposed scheme can be expressed as

$$C_S = [R_B - R_E]^+, \quad (5)$$

where  $[x]^+ = \max(0, x)$ . Considering that the deterministic channels for  $B$  and Eve are fully known for  $S$ , with the optimized phase shift variables, the SC optimization of the proposed uncertain RIS-assisted communication system is formulated as

$$\begin{aligned} \mathcal{P}_1 : \max_{\{\Theta_1, \Theta_2\}} C_S, \\ \text{s.t. } \Theta_1 = \text{diag} \left( e^{j\psi_1}, \dots, e^{j\psi_{N-\eta}} \right), \end{aligned} \quad (6a)$$

$$\begin{aligned} |e^{j\psi_i}| = 1, \psi_i \in [0, 2\pi), \quad \forall i \in (1, N - \eta), \\ \Theta_2 = \text{diag} \left( e^{j\phi_1}, \dots, e^{j\phi_\eta} \right), \end{aligned} \quad (6b)$$

$$|e^{j\phi_j}| = 1, \phi_j \in [0, 2\pi), \quad \forall j \in (1, \eta). \quad (6c)$$

Clearly, it is quite challenging to solve this optimization problem straightforwardly. Thus, we turn to analyze its security performance with optimal elements allocation and derive the corresponding optimal solutions.

### 3. Performance Analysis

In the section, the optimization problem of maximizing SC for the proposed system is analyzed under the optimal elements allocation of RIS, and then we derive the optimal solution for such problem. Since the maximum SC must be

nonnegative, the  $[\cdot]^+$  operation can be omitted without loss of optimality. With the monotonic function  $\log(\cdot)$ , the objective function in  $\mathcal{P}_1$  is reexpressed as

$$R = \max_{\{\Theta_1, \Theta_2\}} \frac{1 + \gamma_B}{1 + \gamma_E}. \quad (7)$$

For simplicity, the reflective passive beamforming with perfect CSI of RIS is taken into account, and we also assume entire elements of the RIS are with the same reflection amplitude [20]. Thus, we have  $\mathbf{h}_{R_1B}^H \Theta_1 \mathbf{h}_{SR_1} = \alpha \sum_{i=1}^{N-\eta} e^{j\psi_i} [\mathbf{h}_{SR_1}]_i [\mathbf{h}_{R_1B}]_i$  and  $\mathbf{h}_{R_2E}^H \Theta_2 \mathbf{h}_{SR_2} = \alpha \sum_{j=1}^{\eta} e^{j\phi_j} [\mathbf{h}_{SR_2}]_j [\mathbf{h}_{R_2E}]_j$ . In the case that  $S$  knows the CSI of all channels completely, the optimal reflection phase shift can be obtained through channels. In addition, the optimal phase shift can be derived via the proposition as follows.

**Proposition 1.** *The optimal RIS phase shift of the communication network is expressed as*

$$\theta_n = \arg(h_{SU}) - \arg([\mathbf{h}_{SR}]_n [\mathbf{h}_{RU}]_n), \quad n = 1, \dots, N, \quad (8)$$

where  $[\mathbf{h}_{SR}]_n$  and  $[\mathbf{h}_{RU}]_n$  represent the  $n$ -th element of  $\mathbf{h}_{SR}$  and  $\mathbf{h}_{RU}$ , respectively.  $h_{SU}$  and  $\mathbf{h}_{RU}$  represent the links from BS and RIS to user, respectively, and  $\arg(\cdot)$  denotes the phase operator.

*Proof.* Note that the function  $|h_{SU} + \mathbf{h}_{RU}^H \Theta \mathbf{h}_{SR}|^2$ ,  $U \in \{B, E\}$  can be rewritten as

$$\begin{aligned} |h_{SU} + \mathbf{h}_{RU}^H \Theta \mathbf{h}_{SR}|^2 \\ = |h_{SU}|^2 + |\mathbf{h}_{RU}^H \Theta \mathbf{h}_{SR}|^2 + 2|\mathbf{h}_{RU}^H \Theta \mathbf{h}_{SR}| |h_{SU}| \cos \\ \cdot [\arg(h_{SU}) - \arg(\mathbf{h}_{RU}^H \Theta \mathbf{h}_{SR})], \quad U \in \{B, E\}. \end{aligned} \quad (9)$$

From Equation (9), it can be easily verified that  $|h_{SU} + \mathbf{h}_{RU}^H \mathbf{\Theta} \mathbf{h}_{SR}|^2$  achieves its maximum value for  $\cos[\arg(h_{SU}) - \arg(\mathbf{h}_{RU}^H \mathbf{\Theta} \mathbf{h}_{SR})] = 1$ , which means that the phase shifts of both direct and cascaded links between the  $U$  and  $S$  are identical, i.e.  $\arg(h_{SU}) = \arg(\mathbf{h}_{RU}^H \mathbf{\Theta} \mathbf{h}_{SR})$ . Therefore, Equation (8) can be derived, which completes this proof.  $\square$

As a result, the maximum achievable SNR can be obtained if the phase shifts satisfy  $\psi_i = \arg(h_{SB}) - \arg([\mathbf{h}_{SR_1}]_i [\mathbf{h}_{R_1 B}]_i)$  and  $\phi_j = \arg(h_{SE}) - \arg([\mathbf{h}_{SR_2}]_j [\mathbf{h}_{R_2 E}]_j)$ , which means that the cascaded links via RIS are with the same phase as the corresponding direct links [21]. Therefore, we have

$$\gamma_B = \frac{P}{\sigma^2} \left| h_{SB} + \alpha \sum_{i=1}^{N-\eta} [\mathbf{h}_{SR_1}]_i [\mathbf{h}_{R_1 B}]_i \right|^2, \quad (10)$$

$$\gamma_E = \frac{P|h_{SE}|^2}{P \left| \alpha \sum_{j=1}^{\eta} [\mathbf{h}_{SR_2}]_j [\mathbf{h}_{R_2 E}]_j \right|^2 + \sigma^2}. \quad (11)$$

From Equations (10) and (11), it is easy to see that the SNR of  $B$  and SINR of Eve only rely on the amplitudes of the reflective elements, not on their phases. Therefore, we can reformulate the optimization problem as

$$\mathcal{P}_2 : \max_{\eta} \frac{1 + \gamma_B}{1 + \gamma_E}, \quad (12a)$$

$$\text{s.t. } 0 \leq \eta \leq N. \quad (12b)$$

Clearly,  $\mathcal{P}_2$  is nonconvex and hard to obtain its optimal result. Therefore, we turn to derive its closed form in an approximate way. Assuming the equivalent size for each RIS element, thus, it follows that all elements in  $\mathbf{h}_{SR}$ ,  $\mathbf{h}_{R_1 B}$ , and  $\mathbf{h}_{RE}$  have the same magnitude, respectively. For brevity, we have  $|h_{SB}| = \sqrt{\beta_{SB}}$ ,  $|h_{SE}| = \sqrt{\beta_{SE}}$ ,  $1/(N-\eta) \sum_{i=1}^{N-\eta} |[\mathbf{h}_{SR_1}]_i| = \sqrt{\beta_{RB}}$ , and  $1/N \sum_{j=1}^{\eta} |[\mathbf{h}_{SR_2}]_j| = \sqrt{\beta_{RE}}$ . In addition,  $\beta_{RB} = h_{SR} h_{RB}$  and  $\beta_{RE} = h_{SR} h_{RE}$ , where  $h_{SR}$ ,  $h_{RB}$ , and  $h_{RE}$  are the channel coefficients of  $S$ ,  $B$ , and Eve related to a single element of RIS, respectively [22]. It is known that, for a small number of components, the channel gain provided by RIS should be smaller than that of the direct channel. However, as the amount of reflective components grows, the channel gain provided by RIS significantly improves and becomes much larger than that of direct channel. Then, we can reformulate Equations (10) and (11) in the more compact forms as

$$\gamma_B = \frac{P}{\sigma^2} \left( \sqrt{\beta_{SB}} + (N-\eta) \alpha \sqrt{\beta_{RB}} \right)^2, \quad (13)$$

$$\gamma_E = \frac{P \beta_{SE}}{P \alpha^2 \beta_{RE} \eta^2 + \sigma^2}. \quad (14)$$

Since  $\gamma_B \gg 1$ , letting  $1 + \gamma_B \approx \gamma_B$  and with Equations (13) and (14), the secrecy capacity maximization problem can be

obtained from

$$\begin{aligned} \frac{\gamma_B}{1 + \gamma_E} &= \frac{(P/\sigma^2) \left( \sqrt{\beta_{SB}} + (N-\eta) \alpha \sqrt{\beta_{RB}} \right)^2}{1 + (P \beta_{SE} / (P \beta_{RE} \alpha^2 \eta^2 + \sigma^2))} \\ &= \frac{P}{\sigma^2} (\alpha_{SB} + (N-\eta) \alpha_{RB})^2 \left( 1 - \frac{P \alpha_{SE}^2}{P \alpha_{SE}^2 + P \alpha_{RE}^2 \eta^2 + \sigma^2} \right), \end{aligned} \quad (15)$$

where  $\alpha_{SB} = \sqrt{\beta_{SB}}$ ,  $\alpha_{RB} = \alpha \sqrt{\beta_{RB}}$ ,  $\alpha_{SE} = \sqrt{\beta_{SE}}$ , and  $\alpha_{RE} = \alpha \sqrt{\beta_{RE}}$ . Since the term  $P/\sigma^2$  has no influence on solving the optimization problem in Equation (15) and the approximation  $P \alpha_{SE}^2 + \sigma^2 \approx P \alpha_{SE}^2$  holds due to  $P \alpha_{SE}^2 \gg \sigma^2$ , the above maximization problem can be simplified as

$$R(\eta) = (\alpha_{SB} + \alpha_{RB}(N-\eta))^2 \left( 1 - \frac{\alpha_{SE}^2}{\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2} \right). \quad (16)$$

Since the function in Equation (16) is continuous, we can derive the following expression by its first-derivations as

$$\begin{aligned} \frac{\partial R(\eta)}{\partial \eta} &= \frac{2 \alpha_{SE}^2 (\alpha_{SB} + \alpha_{RB}(N-\eta))}{(\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2)^2} \\ &\quad \cdot (\alpha_{RB} \alpha_{SE}^2 + (\alpha_{SB} + N \alpha_{RB}) \alpha_{RE}^2 \eta) \\ &\quad - 2 \alpha_{RB} (\alpha_{SB} + \alpha_{RB}(N-\eta)). \end{aligned} \quad (17)$$

It is quite difficult to analyze its trend directly from the above formula; then, the second-derivation is derived and simplified as

$$\begin{aligned} \frac{\partial^2 R(\eta)}{\partial \eta^2} &= \frac{2}{(\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2)^3} \left( (\alpha_{SE}^2 \alpha_{RE}^2 (\alpha_{SB} + \alpha_{RB}(N-2\eta)) - \alpha_{RB} \alpha_{RE}^4 \eta^3) \right. \\ &\quad \cdot (\alpha_{RB} \alpha_{RE}^2 \eta^2 (2\eta - 3N) - 3 \alpha_{SE} \alpha_{RE}^2 \eta^2 + \alpha_{RB} \alpha_{SE}^2 (N-2\eta) \\ &\quad + \alpha_{SB} \alpha_{SE}^2) - (\alpha_{SB} + \alpha_{RB}(N-\eta)) (\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2) \\ &\quad \left. \cdot (2 \alpha_{RB} \alpha_{SE}^2 \alpha_{RE}^2 \eta + 3 \alpha_{RB} \alpha_{RE}^4 \eta^3) \right). \end{aligned} \quad (18)$$

Through the positive and negative judgment in Equation (18), we can acquire the monotonicity of the first-derivation expression and further determine whether it has a zero point in the value range. For simplicity let  $G_1(\eta) = \alpha_{SE}^2 \alpha_{RE}^2 (\alpha_{SB} + \alpha_{RB}(N-2\eta)) - \alpha_{RB} \alpha_{RE}^4 \eta^3$ ,  $G_2(\eta) = \alpha_{RB} \alpha_{RE}^2 \eta^2 (2\eta - 3N) - 3 \alpha_{SE} \alpha_{RE}^2 \eta^2 + \alpha_{RB} \alpha_{SE}^2 (N-2\eta) + \alpha_{SB} \alpha_{SE}^2$ , and  $G_3(\eta) = -(\alpha_{SB} + \alpha_{RB}(N-\eta)) (\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2) (2 \alpha_{RB} \alpha_{SE}^2 \alpha_{RE}^2 \eta + 3 \alpha_{RB} \alpha_{RE}^4 \eta^3)$ . Due to the positive characteristic of  $2/(\alpha_{SE}^2 + \alpha_{RE}^2 \eta^2)^3$ , Equation (18) can be simplified as

$$\frac{\partial^2 R(\eta)}{\partial \eta^2} = G_1(\eta) G_2(\eta) + G_3(\eta). \quad (19)$$

Clearly,  $G_1(\eta)$ ,  $G_2(\eta)$ , and  $G_3(\eta)$  are monotonically decreasing with  $\eta$ , where  $G_1(0) > 0$ ,  $G_1(N) < 0$ ,  $G_2(0) > 0$ ,  $G_2(N) < 0$ , and  $G_3(\eta)$  are always smaller than 0 with the increase of  $\eta$ . Therefore, one can conclude that  $G_1(\eta)$  and  $G_2(\eta)$  have

zero points in their effective domain, which are denoted by  $\xi_1$  and  $\xi_2$ , respectively. For the sake of illustration, letting  $\xi_1 < \xi_2$ , we have

$$G_1(\eta)G_2(\eta) \begin{cases} >0, & \text{for } 0 \leq \eta < \xi_1, \\ <0, & \text{for } \xi_1 \leq \eta < \xi_2, \\ >0, & \text{for } \xi_2 \leq \eta \leq N. \end{cases} \quad (20)$$

Since  $G_3(\eta)$  is always smaller than 0 with the increase of  $\eta$ , it can be regarded as a constant and has no effect on the trend of  $G_1(\eta)G_2(\eta)$ , but it determines their zero points. According to Equation (18),  $\partial^2 R(\eta)/\partial \eta^2|_{\eta=0} = 0$ . However, the positive or negative of Equation (18) cannot be determined; thus we discuss each case separately. For the case  $\partial^2 R(\eta)/\partial \eta^2|_{\eta=N} > 0$ , the monotonicity of Equation (18) is similar to that of  $G_1(\eta)G_2(\eta)$ . However, when  $\partial^2 R(\eta)/\partial \eta^2|_{\eta=N} < 0$ , the zero point of Equation (18) denoted by  $\zeta$  exists, and hence, we can conclude

$$\frac{\partial^2 R(\eta)}{\partial \eta^2} \begin{cases} >0, & \text{for } 0 \leq \eta < \zeta, \\ <0, & \text{for } \zeta \leq \eta < N. \end{cases} \quad (21)$$

Through observing the positive and negative of  $\partial^2 R(\eta)/\partial \eta^2$ , we know that the first-derivation of  $R(\eta)$  tends to increase and then decreases and then maybe increase again with the increase of  $\eta$ , where  $\partial R(\eta)/\partial \eta|_{\eta=0} = 0$  and  $\partial R(\eta)/\partial \eta|_{\eta=N} < 0$ . Therefore, there exists a maximum value of the secrecy rate for the formulated optimization problem, within the scope of its first derivative in solving extreme result. Thus, we can derive the optimal solution by calculating the equation of  $\partial R(\eta)/\partial \eta = 0$ , and the maximum secrecy capacity is also determined. Letting  $\partial R(\eta)/\partial \eta = 0$ , the objective function can be simplified as

$$\alpha_{RB}\alpha_{RE}^2\eta^3 + 2\alpha_{RB}\alpha_{SE}^2\eta - \alpha_{SE}^2(\alpha_{SB} + N\alpha_{RB}) = 0. \quad (22)$$

Since  $\alpha_{RB}\alpha_{RE}^2 \neq 0$ , we can eliminate the coefficient of the first term in Equation (22) and obtain the following equation:

$$\eta^3 + 2\frac{\alpha_{SE}^2}{\alpha_{RE}^2}\eta - \frac{\alpha_{SE}^2}{\alpha_{RB}\alpha_{RE}^2}(\alpha_{SB} + N\alpha_{RB}) = 0. \quad (23)$$

Clearly, Equation (23) is a special cubic equation, which can be solved by the Cardano formula method. From Equation (23), we can find that  $\alpha_{SE}^2/\alpha_{RE}^2 \geq 0$  and  $\alpha_{SE}^2/\alpha_{RB}\alpha_{RE}^2(\alpha_{SB} + N\alpha_{RB}) \geq 0$ , it can be converted into the form of

$$\eta^3 + p\eta + q = 0, \quad (24)$$

where  $p = 2\alpha_{SE}^2/3\alpha_{RE}^2$  and  $q = -\alpha_{SE}^2/\alpha_{RB}\alpha_{RE}^2(\alpha_{SB} + N\alpha_{RB})$ . In addition, due to  $p \neq 0$  and  $q \neq 0$ , letting  $\eta = u + v$ , we have

$$\eta^3 = (u + v)^3 = u^3 + v^3 + 3uv(u + v) = u^3 + v^3 + 3uv\eta. \quad (25)$$

Comparing the coefficients of Equations (24) and (25) and

after some mathematical manipulations, we can obtain

$$\begin{cases} u^3 + v^3 = -q, \\ u^3 v^3 = \left(\frac{-p}{3}\right)^3. \end{cases} \quad (26)$$

According to  $q^2 + (4p^3/27) > 0$ ,  $u^3$  and  $v^3$  are two roots of the equation  $x^2 + qx - (p/3)^3 = 0$ , which can be found by the root formula; the corresponding closed-form expressions can be obtained from

$$\begin{cases} u^* = \sqrt[3]{\frac{-q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}, \\ v^* = \sqrt[3]{\frac{-q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \end{cases} \quad (27)$$

In this manner, the optimal solution of Equation (15) can be finally expressed as

$$\eta^* = \lceil u^* + v^* \rceil, \quad (28)$$

where  $\lceil \cdot \rceil$  is the ceiling function. Thus, from the closed-form expression of  $\eta^*$ ,  $N - \eta^*$  can thus be obtained as the optimal number of elements allocated for serving  $B$ .

#### 4. Numerical Result Comparison

In the section, we validate the expression of the analysis and evaluate the performance of our considered system with the assistance of the uncertain RIS via simulation results. All links in the system are assumed to follow independent Rayleigh fading with the path loss exponent of 2.2, and the noise variance is -94 dBm. For simplicity, the following node locations are considered:  $(x_S, y_S) = (0, 0)$ ,  $(x_R, y_R) = (40, 10)$ ,  $(x_B, y_B) = (50, 0)$ , and  $(x_E, y_E) = (45, -5)$ .

Figure 2 illustrates the variation of achieved rate as the number of elements for Eve with the fixed transmit power  $P = 1$  W varies, and the total number of elements is set to 1,000. The achievable rates of  $B$  and Eve are decreasing when the number of the elements for Eve increases, as shown in the figure. However, the achievable rate at Eve decreases significantly than that of the one at  $B$ . It also illustrates that the achievable secrecy rate is first increasing and then decreasing with an increasing number of elements for Eve. Thus, there exists an optimal value of elements for  $B$  and Eve as the result of 10.0392 bps/Hz. Through observation, it is known that 450 is the optimal value in the simulation results which are equal to the optimal value obtained from our analysis in Equation (28).

In Figure 3, we compare different schemes of elements allocation versus the transmit power with the power from 1 W to 30 W. By means of the simulation, we compare several models including the proposed scheme, all elements of RIS to enhance secure communication at  $B$ , all elements of RIS to interfere Eve, and the model without RIS. For the system without RIS, its achievable secrecy rate is 0 since the

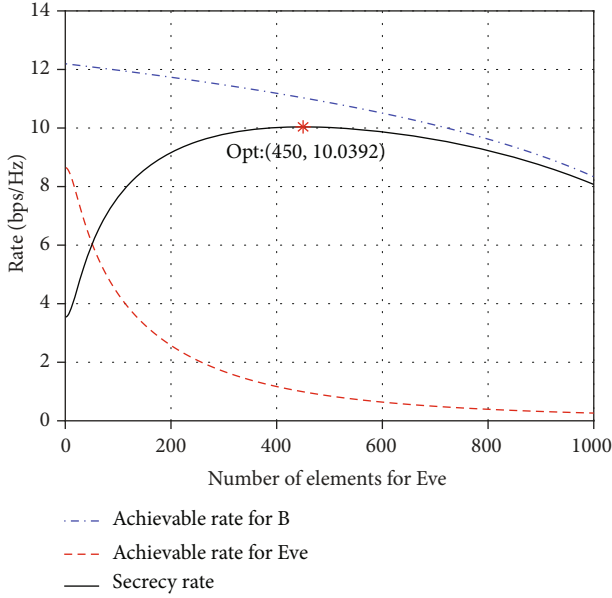


FIGURE 2: Achievable rate vs. the number of elements for jamming Eve  $\eta$ .

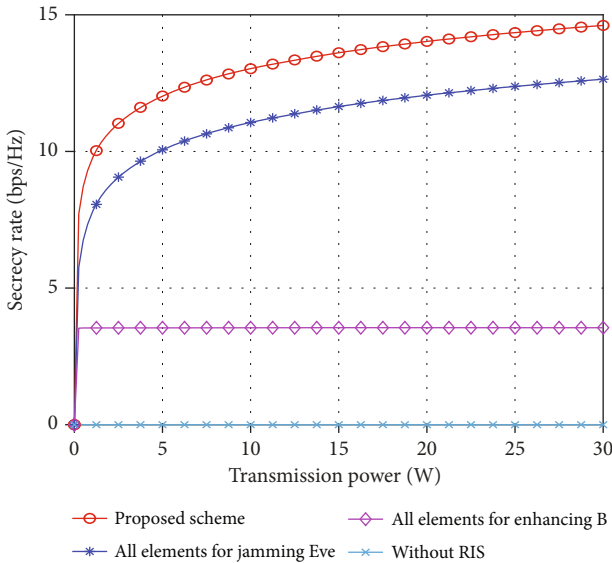


FIGURE 3: Comparisons of the proposed scheme and other schemes.

achievable rate at Eve is higher than that at B. And when all the components of RIS are used to improve the legitimate communication, the secrecy rate is determined by the CSI instead of the transmission power of base station; thus, the secrecy rate keeps constant under varying transmit power. The achievable secrecy capacity for our considered scheme and the one that all elements to interfere Eve monotonically increase with the increasing transmit power, and our proposed scheme achieves better results. Moreover, it also verifies that our proposed scheme has a higher secrecy capacity than other schemes, which makes it superior compared with than other schemes.

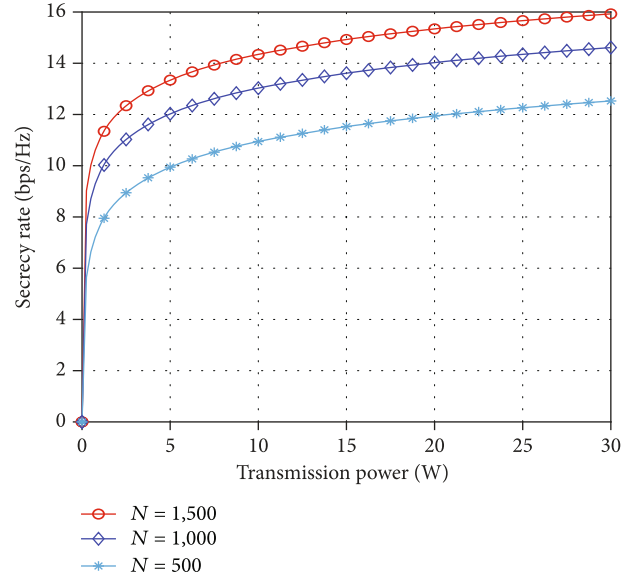


FIGURE 4: Secrecy rate vs. transmission power  $P$  with  $N = 500, 1000, 1500$ .

Figure 4 demonstrates the variation of achievable secrecy rate with total number of elements for RIS, in which  $N$  is set to be 500, 1,000, and 1,500. According to the figure, the achievable secrecy capacity is increasing as the transmit power increases. In addition, it reveals that the secrecy rate in the uncertain RIS-assisted communication network increases with an increasing number of components.

## 5. Conclusions

This paper investigated uncertain RIS-assisted PLS, with the presence of an eavesdropper. The uncertain RIS can enhance the signal for the legitimate user while jamming the eavesdropper by separating the components of RIS into two functional sections. The optimal element allocation and achievable secrecy rate can be obtained in their closed forms by solving the formulated problem. Numerical results verifying the accuracy of the derived expression show the advantage of our considered scheme compared with its counterparts.

## Data Availability

The simulation data used to support the findings of this study are available from the author upon request, who can be contacted at 2010310034@stmail.ntu.edu.cn

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61971245 and 61801249.

## References

- [1] W. Duan, J. Gu, M. Wen, G. Zhang, Y. Ji, and S. Mumtaz, "Emerging technologies for 5G-IoV networks: applications, trends and opportunities," *IEEE Network*, vol. 34, no. 5, pp. 283–289, 2020.
- [2] X. Pei, Y. Chen, M. Wen, H. Yu, E. Panayirci, and H. V. Poor, "Next-generation multiple access based on NOMA with power level modulation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1072–1083, 2022.
- [3] W. Xu, F. Gao, J. Zhang, X. Tao, and A. Alkhateeb, "Deep learning based channel covariance matrix estimation with user location and scene images," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8145–8158, 2021.
- [4] F. Gao, B. Lin, C. Bian, T. Zhou, J. Qian, and H. Wang, "FusionNet: enhanced beam prediction for mmWave communications using Sub-6 GHz channel and a few pilots," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8488–8500, 2021.
- [5] M. Wen, Q. Li, K. J. Kim et al., "Private 5G networks: concepts, architectures, and research landscape," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 7–25, 2022.
- [6] L. Sun and Q. Du, "Physical layer security with its applications in 5G networks: a review," *China Communications*, vol. 14, no. 12, pp. 1–14, 2017.
- [7] C. Huang, S. Hu, G. C. Alexandropoulos et al., "Holographic MIMO surfaces for 6G wireless networks: opportunities, challenges, and trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, 2020.
- [8] W. Duan, Y. Ji, J. Hou, B. Zhuo, M. Wen, and G. Zhang, "Partial-DF full-duplex D2D-NOMA systems for IoT with/without an eavesdropper," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6154–6166, 2021.
- [9] D.-H. Ha, T. T. Duy, P. N. Son, T. Le-Tien, and M. Voznak, "Security-reliability trade-off analysis for rateless codes-based relaying protocols using NOMA, cooperative jamming and partial relay selection," *IEEE Access*, vol. 9, pp. 131087–131108, 2021.
- [10] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.
- [11] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.
- [12] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.
- [13] Z. Chen, X. Ma, C. Han, and Q. Wen, "Towards intelligent reflecting surface empowered 6G terahertz communications: a survey," *China Communications*, vol. 18, no. 5, pp. 93–119, 2021.
- [14] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: a programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, 2019.
- [15] B. Feng, Y. Wu, and M. Zheng, "Secure transmission strategy for intelligent reflecting surface enhanced wireless system," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Xi'an, China, Oct. 2019.
- [16] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Waikoloa, HI, USA, Dec. 2019.
- [17] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1410–1414, 2019.
- [18] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Processing Letters*, vol. 27, pp. 1300–1304, 2020.
- [19] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2020.
- [20] M. Wang, W. Duan, G. Zhang, M. Wen, J. Choi, and P. -H. Ho, "On the achievable capacity of cooperative NOMA networks: RIS or relay?," *IEEE Wireless Communications Letters*, vol. 11, no. 8, 2022.
- [21] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [22] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: how large surfaces are needed to beat relaying?," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020.



## Research Article

# Reconfigurable Intelligent Surface Assisted Non-Terrestrial NOMA Networks

Xuanhao Lian <sup>1</sup>, Xinwei Yue <sup>1</sup>, Xuehua Li <sup>1</sup>, Xiang Yun,<sup>2</sup> Tian Li,<sup>3</sup> and Dehan Wan <sup>4</sup>

<sup>1</sup>Key Laboratory of Information and Communication Systems, Ministry of Information Industry and Also with the Key Laboratory of Modern Measurement & Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 100101, China

<sup>2</sup>Baicells Technologies co. Ltd, Beijing 100095, China

<sup>3</sup>The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang Hebei 050081, China

<sup>4</sup>Guangdong University of Finance, Guangzhou 510521, China

Correspondence should be addressed to Xuehua Li; [lixuehua@bistu.edu.cn](mailto:lixuehua@bistu.edu.cn)

Received 30 June 2022; Revised 27 August 2022; Accepted 31 August 2022; Published 19 September 2022

Academic Editor: Jun Li

Copyright © 2022 Xuanhao Lian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper considers the application of reconfigurable intelligent surface (RIS) to non-terrestrial non-orthogonal multiple access (NOMA) networks. More specifically, the performance of a pair of non-orthogonal users for RIS assisted non-terrestrial NOMA networks is investigated over large-scale fading and Nakagami- $m$  fading cascaded channel. The exact and asymptotic expressions of outage probability are derived for the nearby user and distant user with the imperfect successive interference cancellation (SIC) and perfect SIC schemes. Based on the approximated results, the diversity orders of these two users are obtained in the high signal-to-noise ratios. The simulation results are used to verify the theoretical derivations and find that: 1) The outage behaviors of RIS assisted non-terrestrial NOMA networks outperforms than that of orthogonal multiple access; 2) By increasing the number of reflecting elements of RIS and Nakagami- $m$  fading factors  $m'$  and  $\Omega$ , RIS-assisted non-terrestrial NOMA networks are able to achieve the enhanced outage performance.

## 1. Introduction

With the development of the fifth-generation (5G) communication technologies, mobile communication has gradually entered the era of high-speed interconnections [1], and research on the sixth-generation (6G) communication technologies has gradually been carried out to aim at the achievement of global coverage, enhanced spectrum efficiency, better security and higher intelligence level [2]. As one of the key technologies of 6G, non-orthogonal multiple access (NOMA) has ability to support multiple users' information which is linearly superposed at the same physical resource over different power levels. The authors of [3] well explained the development status and working principle of NOMA. And in existing researches, NOMA technologies have been applied in the Internet of Things industry for improving the spectral efficiency to make up for the shortage

of spectrum resources [4]. Also NOMA can not only meet the requirements of communication transmission, but also have more prominent advantages in energy consumption [5].

Recent years have witnessed a spurt of progress in NOMA technologies and studies about terrestrial scenarios have provided new insights into NOMA application. The specific concept of NOMA is that multiple users' signals are overlaid at the transmitting end by employing the superposition coding scheme. At receiving end, the successive interference cancellation (SIC) is carried out to peel off the desired signal [6]. In the SIC process, the strong users firstly decode all weak users' signals and remove them for the preparation to decode their own signals. The weak users directly decode their own signals by treating the signals of all strong users as interference. It has shown that NOMA is capable of providing the enhanced system efficiency and user fairness

with respect to orthogonal multiple access (OMA), and the authors of [7] provided a systematic treatment of NOMA from its combination with multiple-input multiple-output technologies to cooperative NOMA. The performance of unified NOMA framework was surveyed in terms of outage probability by taking into account both perfect SIC (pSIC) and imperfect SIC (ipSIC) [8]. Furthermore, the ideology of NOMA was extended to cooperative communications [9], where the nearby user with better channel conditions was reviewed as the relaying to forward information for distant user. In [10], the authors analyzed the ergodic rate of decode-and-forward (DF) relaying based cooperative NOMA systems. With the emphasis on physical layer communications, the authors of [11] investigated the secrecy outage behaviors of NOMA networks with the aid of stochastic geometry. Additionally, the authors in [12] proposed a backscatter assisted NOMA network and confirmed that the transmission reliability and effectiveness can be enhanced without impairing the spectrum efficiency. Applying NOMA to random access, the NOMA assisted semi-grant free transmission was proposed in [13], which solves the problems of no upper limit on the number of admitted users. A channel differences exploiting scheme for NOMA was discussed in [14], where a dual-hop cooperative relaying network was proposed and a best relay was picked up as an active node for transmission. For better reliability of NOMA based networks, the authors of [15] suggested a multi-band scheduling policy which accommodates the near and far users through sub-band exploring.

Non-terrestrial communication network is one of the important application scenarios for 6G networks and a lot of attention have been received on it [16]. The NOMA technology can also be incorporated into the space-air-ground networks and play a pivotal role in the non-terrestrial communication scenarios [17]. Cooperative NOMA communication protocol was applied to a satellite-terrestrial networks in [18], in which outage performance was notably improved because of the application of NOMA scheme. An unmanned aerial vehicle (UAV) was used as a relay in [19], where NOMA scheme was utilized to transmit signals to solve the detection-vector optimization problem about UAV signal collection. Additionally for NOMA-UAV networks in [20], a path following algorithm was proposed to formulate the max-min rate optimization under NOMA scheme with a UAV serving a large number of ground users. Multi-layer NOMA satellite network was studied in [21], joint user pairing and power allocation scheme were introduced to deal the frequency interference coordination between different orbits. Different from studying on geostationary orbit scenarios, the authors of [22] studied a low earth orbit satellite NOMA system, huge time delay and Doppler shift were taken into consideration carefully. In terms of performance analysis, the authors in [23] analyzed the outage behaviors of satellite NOMA networks based on Shadowed-Rician fading channels, the diversity order of the users have achieved with ipSIC and pSIC scheme, respectively. With the perspective on decoding and transmitting mode in [24], the outage performance for amplify-and-forward (AF) relaying transmission protocols under

the hybrid satellite-terrestrial overlay NOMA network was analyzed in detail. Meanwhile in [25], a two-user NOMA-based hybrid satellite-terrestrial relay network was investigated, where the secondary user acquire its desired signal through a DF relay.

Reconfigurable intelligent surface (RIS), as a new type of transmission relay, has ability to correct the wireless channel through a highly controllable software, which paves the way for an intelligent and programmable wireless environment [26–28]. Because of its novelty and greater gain, the RIS-assisted NOMA networks have attracted a large part of research effort [29–31]. In [29], the performance of theoretical framework for the RIS and AF relaying were compared and results showed that RIS-assisted wireless systems outperform the corresponding AF-relaying ones. Further considered in [30], a RIS-aided NOMA network with stochastic geometry model was discussed and the SIC order is proofed to be altered since RISs were able to change the channel quality of users. Multiple RISs assisted NOMA networks was investigated in [31], two scenarios were taken into consideration on whether or not there was a direct link between base station and each user. Additionally discussed in [32], outage performance of RIS-assisted NOMA systems was analyzed with the help of 1-bit coding, in which the performance analysis of cascaded channel condition has been taken into consideration. Moreover in [33], a RIS-assisted NOMA scheme was proposed to demonstrate that more users can be served on each orthogonal spatial direction with the help of RIS. Eventually, as the latest research direction, the authors of [34] have proposed a simultaneously transmitting and reflecting RIS aided NOMA systems, which shared new ideas for RIS deployment in multiple scenarios. For the perspective of grant-free massive access, the authors of [35] have proposed a design to leverage the RIS to for grant-free massive access at millimeter-wave frequency to boost the reliability of communication systems. To solve the multiplicative fading effect introduced by passive RIS, active RIS [36] has been proposed to compensate for it by reflecting and amplifying the signals, more research on active RIS is also underway. In non-terrestrial networks, the use of RIS technology can make signal transmission more flexible to make up for the problem of satellite communication encountering blind spots. At the same time, the use of RIS technology also provides an idea for the modeling and calculation of cascaded channels, which is worth more in the future research.

*1.1. Motivations and Related Works.* Based on the above-mentioned papers, it lays a solid foundation for NOMA and RIS technologies, the research for integrating the two promising technologies in non-terrestrial networks still gets a lot of upside. RIS can effectively change the transmission mode of electromagnetic waves, including direction, phase and polarization mode, etc. Therefore, introducing RIS into the non-terrestrial communication system to change the phase shifts of the satellite signal and propagate it to the ground users has attracted our attention. It should be noticed that compared with the RIS-assisted terrestrial networks, the RIS-assisted non-terrestrial network has a longer transmission distance and the channel fading needs

reconsideration to fit the scenario. Path Loss gets complicated and channel model needs to be considered for long-distance transmission. Simulation analysis of the transmission power also requires differential modeling. Recently, investigators have examined the effects of RIS communications. To broaden the application of channel mathematical calculation, the probability distribution expression for Rayleigh/Nakagami- $m$  composite fading were deduced in [37], which provides a mathematical method for the computation of cascaded channel models. Simultaneously, the authors of [38] proposed a coherent phase shifting method to deal with the phase shift matrix of RIS, which paves way for the RIS simulation calculation. Using the mathematical models built on the basis of these researches, we aspire to establish a RIS-assisted non-terrestrial NOMA networks with channel model of large-scale fading and Nakagami- $m$  fading cascaded channels, which has not been studied before to the best of our knowledge. More specifically, a pair of users, the nearby user  $n$  and distant user  $m$ , are set to be investigated according to outage probability. The non-terrestrial network communication model based on the transmission of RIS and the channel calculation method under the large-scale fading and Nakagami- $m$  fading cascaded channel model have become the innovations of this paper. Additionally, OMA condition are set to be benchmark and also evaluated seriously. In terms of the above details, the main related work in the paper is described as follows:

- (1) We investigate the performance of RIS-assisted non-terrestrial NOMA networks over large-scale fading and Nakagami- $m$  fading cascaded channels. Specifically, cascaded channels are applied in the networks, large-scale fading and Nakagami- $m$  fading distribution for the channel molding are taken into account. The approximated scheme of a Laguerre expansion is applied for the cascaded channel modeling and approximate expression of the cascaded channel is analyzed in detail
- (2) We derive the approximate and asymptotic expressions of outage probability for the nearby user  $n$  with ipSIC/pSIC and distant user  $m$ . To get more insights, the diversity orders of the nearby user  $n$  with ipSIC/pSIC and distant user  $m$  are obtained on the basis of the approximated outage probability expressions. We observe that the diversity orders of user  $n$  with ipSIC and user  $m$  are related to the configure elements, user order, Rayleigh and Nakagami- $m$  factors
- (3) We analyze the outage behaviors of the two users for OMA. It can be further confirmed that outage behaviors of NOMA with pSIC outperforms than that of OMA. Detailed comparisons about the impact of changes in individual variables on outage performance are analyzed as well. The simulation results illustrate that when the configure elements  $K$ , Nakagami- $m$  shaping parameter  $m'$  and scale parameter  $\Omega$  increase, the outage performance of the networks turns to be enhanced

*1.2. Organizations and Notations.* The remainder of this paper is structured as follows. System model of the RIS-assisted non-terrestrial NOMA networks is depicted in Section II. In Section III, outage behaviors of the networks is analyzed and expressions of outage probability for the two non-orthogonal users are derived meticulously. In Section IV, simulation results and detailed analysis are presented to consolidate the conclusions obtained in the above sections, while summarized records in Section V.

Notations in this paper describes mainly as follows:  $f_X(\cdot)$  and  $F_X(\cdot)$  are denote the probability density function (PDF) and the cumulative distribution function (CDF) of a random variable  $X$ ;  $\mathbb{E}[\cdot]$  denotes the expectation operator;  $\propto$  denotes “be proportional to”.

## 2. System Model

In this paper, we consider a RIS-assisted non-terrestrial NOMA network with a pair of non-orthogonal users among multiple users as shown in Figure 1 and two user equipments (UEs), nearby user  $n$  and distant user  $m$ , are denoted by  $D_n$  and  $D_m$ , respectively. Moreover, taking into account the large transmission distance in the non-terrestrial communication model, the signal of direct connection between each UE and the satellite is strongly attenuated, so we assume that the UEs only communicate with the satellite through the RIS. In order to get the succinct analysis, we only consider that each user is equipped with a single antenna. As a novel type of relay for transmitting signals, the RIS has  $K$  discrete reconfigurable elements, each element is controlled by the software systems to affect the phase shifting of signal reflecting, and the reflection-coefficient matrix can be denoted as  $\Theta = \text{diag}(\beta e^{j\theta_1}, \beta e^{j\theta_2}, \dots, \beta e^{j\theta_K})$ , where  $\beta \in [0, 1]$  and  $\theta_k \in [0, 2\pi)$  are the amplitude-reflection coefficient and the phase shift variable of  $k$ -th element of RIS, respectively. Same as above, the reflection-coefficient matrix can be intelligently adjusted for a better way on transmission. The complex channel coefficient between the satellite and RIS is denoted by  $h \in \mathbb{C}^{1 \times K}$ , and the complex channel coefficients between the RIS and  $p$ -th UE are denoted by  $\mathbf{g}_p \in \mathbb{C}^{1 \times K}$ , where  $p \in \{m, n\}$ . Specifically,  $h = [h_1, h_2, \dots, h_K]$ ,  $\mathbf{g}_p = [g_1^p, g_2^p, \dots, g_K^p]$ , and the channels between each element and the satellite can be denoted by [39]  $h_k = \rho \sqrt{G_t} \sqrt{\lambda(4\pi d)^2} \sqrt{G_i}$ , where  $\rho \sim CN(0, 1)$  represents the small-scale fading,  $G_t$  and  $G_i$  are lobe gains of antennas at satellite end and users end, respectively.  $\sqrt{\lambda(4\pi d)^2}$  is the free space path loss,  $\lambda$  and  $d$  are the wave length and the distance between the satellite and RIS, respectively. In order to facilitate subsequent calculations, we denote  $\bar{\lambda} = \sqrt{G_t} \sqrt{\lambda(4\pi d)^2} \sqrt{G_i}$ . Without loss of precision, in order to visually represent the performance of users' channel under different transmission distances, the ordered cascaded channel gains has been taken into account, so we consider  $|h\Theta\mathbf{g}_m^T|^2 \leq |h\Theta\mathbf{g}_n^T|^2$ . It is worth noting that the ordering of channels also determines the order of nearby and distant users.

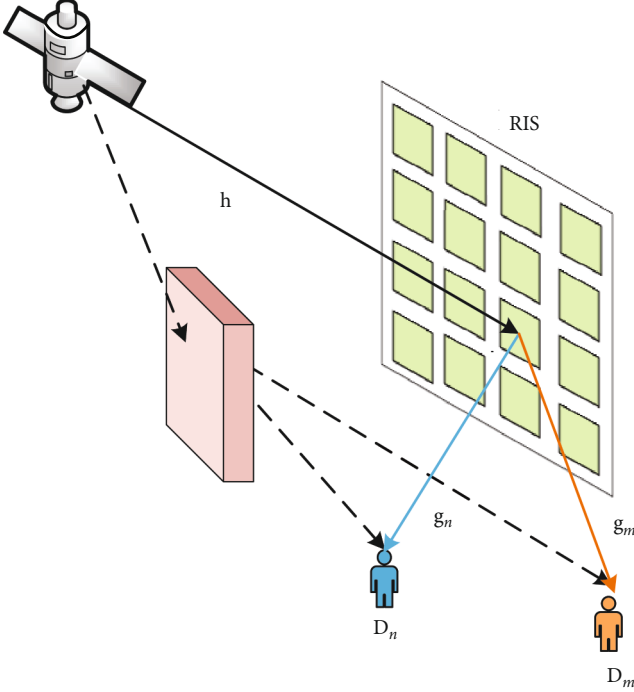


FIGURE 1: System model of RIS-assisted non-terrestrial NOMA networks.

**2.1. Signal Model.** The superposed signals are broadcast to the two UEs by reflection of the RIS, and the received signal  $y_p$  at the  $p$ -th UE reflected by RIS is given by

$$y_p = h\Theta\mathbf{g}_p^T \left( \sqrt{a_n P}x_n + \sqrt{a_m P}x_m \right) + n_p, \quad (1)$$

where  $P$  is the transmit power,  $x_n$  and  $x_m$  are the transmitted normalized power signal at  $D_n$  and  $D_m$ . For elaboration,  $\mathbb{E}\{x_n^2\} = \mathbb{E}\{x_m^2\} = 1$ .  $a_n$  and  $a_m$  are power allocation coefficients,  $a_n \leq a_m$  with  $a_n + a_m = 1$ , which is for the user's fairness of the NOMA transmission. Meanwhile,  $n_p$  is the additive white Gaussian noise (AWGN) at  $D_p$ .

In NOMA communication protocols, the transmitting end use the non-orthogonal transmission scheme to allocate users' transmission power. At the receiving end, the SIC receiver decode the signal for each of the users' part, and the process of SIC is the user signal with weaker power allocation regards the user signal with stronger power as interference and decodes its own signal, while the user signal with stronger power allocation needs to decode and subtract the user signal with weaker power allocation first, and then decode its own signal. Based on the basic principles of NOMA, the received signal-to-interference-plus-noise ratio (SINR) at  $D_n$  to detected  $D_m$ 's information  $x_m$  is given by

$$\gamma_{D_n \rightarrow D_m} = \frac{\rho |\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_m}{\rho |\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_n + 1}, \quad (2)$$

where  $\rho = P/N_0$  denotes the transmit signal-to-noise ratio (SNR).

After SIC process, the received signal at  $D_n$  to detect its own information  $x_n$  is given by

$$\gamma_{D_n} = \frac{\rho |\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_n}{\omega \rho |h_I|^2 + 1}, \quad (3)$$

where  $\omega \in \{0, 1\}$ ,  $\omega = 0$  and  $\omega = 1$  denotes the pSIC and ipSIC operation, respectively. Without loss of generality,  $h_I \sim CN(0, \Omega_I)$  is the corresponding complex channel coefficient of the residual interference form ipSIC [40, 41], which is modeled as the Rayleigh fading.

When the receiver at  $D_m$  get the superposed signal, the decoding process is to treat the signal of  $D_n$  part as interference, and treat the signal of  $D_m$  part as desired signal. Therefore, the received SINR at  $D_m$  to detected its own information  $x_m$  by treating the signal  $x_n$  as interference is given by

$$\gamma_{D_m} = \frac{\rho |\mathbf{h}\Theta\mathbf{g}_m^T|^2 a_m}{\rho |\mathbf{h}\Theta\mathbf{g}_m^T|^2 a_n + 1}. \quad (4)$$

For the baseline of OMA scheme, which is chosen for the purpose of comparing to the NOMA scheme, the SNR of the  $p$ -th user under the OMA scheme can be given by

$$\gamma_p^{OMA} = \rho |h\Theta\mathbf{g}_p^T|^2. \quad (5)$$

**2.2. Channel Model.** In non-terrestrial communication systems, there is an extreme long distance from the satellite to the RIS and the signals are mainly attenuated by free-space path loss. Therefore, a free space transmission model is often used to predict the received signals in a line-of-sight environment [42], and the signals should mainly experience large-scale fading [39], so we consider the channels between satellite and RIS are modeled as large-scale fading channels. The channels between RIS and the UEs are mainly None Line of Sight, then we consider the RIS-UEs channels to be modeled as Nakagami- $m$  fading channels. Noted that large-scale fading is a kind of Rayleigh fading, our analysis of large-scale fading is based on the analysis of Rayleigh fading. So we denote that  $[h_1, h_2, \dots, h_K]$  are independent identically distributed (i.i.d.) Rayleigh random variables (RVs), and  $[g_1, g_2, \dots, g_K]$  are i.i.d. Nakagami- $m$  RVs with shape parameter  $m'$  and spread parameter  $\Omega$ , and all of the  $h_n$  and  $g_n$  are independent, either.

From the perspective of enhancing the network performance, coherent phase shifting design are selected to optimize the phase shifting process for RIS-assisted non-terrestrial NOMA networks. In coherent phase shifting scheme, the phase shift of each reflecting and transmitting element is matched with the phases of its incoming and outgoing fading channels, and coherent phase shifting scheme has better analytical performance than that of the random phase shifting scheme. it can simulate a more idealized and better channel state, and can reflect the system characteristics more intuitively and conveniently. For the coherent

phase shifting scheme [38], we can define  $h\Theta\mathbf{g}_p^T = \sum_{n=1}^K e^{-j\theta_n} h_n g_n$ , where  $h_n$  and  $g_n$  denote the  $n$ -th element of  $\mathbf{h}$  and  $\mathbf{g}_p$ ,  $\theta_n$  denotes the phase shift of the  $n$ -th reflecting element of the RIS. Match the phase shifts of the RIS and the phases of the RIS fading gains, we can obtain  $h\Theta\mathbf{g}_p^T = \sum_{n=1}^K |h_n g_n|$

To take explicit insights for understanding, we first consider one single cascaded channel from the satellite to the RIS to the UE, then analyze the multiple parallel cascaded channels on the basis of single cascaded channel. The PDF of a single Rayleigh cascade Nakagami- $m$  fading channel can be given by [37].

$$f_{RN}(x) = \frac{4x^{m'}}{\Gamma(m')} \left( \frac{m'}{\hat{r}_1^2 \hat{r}_2^2} \right)^{(1+m')/2} K_{1-m'} \left( 2x \sqrt{\frac{m'}{\hat{r}_1^2 \hat{r}_2^2}} \right), \quad (6)$$

where  $m'$  is the shaping parameter of the Nakagami- $m$  RV,  $\hat{r}_i$  is the root-mean square value of the received signal envelope  $R_i$ , from [43],

$$\hat{r}_i \triangleq \sqrt{\mathbb{E}[R_i^2]} = \frac{\bar{r}_i \Gamma(\mu_i) [(1 + \kappa_i) \mu_i]^{1/2}}{\Gamma(\mu_i + 1/2) e^{-\kappa_i \mu_i} {}_1F_1(\mu_i + 1/2; \mu_i; \kappa_i \mu_i)}, \quad (7)$$

where  $i \in \{1, 2\}$ ,  $\mathbb{E}[R_i^2]$  represents the second moment of  $R_i$ , and  $R_i$  is the received signal envelope.  $\bar{r}_i = \mathbb{E}[R_i]$ ,  $\Gamma(\cdot)$  is the gamma function ([44], Eq. (8.310.1)),  ${}_1F_1(a; b; z)$  denotes the confluent hypergeometric function ([44], Eq. (9.210.1)).  $K_\phi(\cdot)$  is the  $\phi$ -th-order modified Bessel function of the second kind ([44], Eq. (8.432)).

To simplify expression, we denote  $\varphi_n = |h_n g_n|$ . It can be derived that the mean and variance of  $\varphi_n$  are, respectively

$$E(\varphi_n) = \bar{\lambda} \sigma \sqrt{\frac{\pi}{2}} \frac{\Gamma(m' + 1/2)}{\Gamma(m')} \left( \frac{\Omega}{m'} \right)^{1/2}, \quad (8)$$

$$Var(\varphi_n) = \bar{\lambda}^2 \left[ 2\Omega\sigma^2 - \frac{\pi\Omega\sigma^2}{2m'} \frac{\Gamma^2(m' + 1/2)}{\Gamma^2(m')} \right]. \quad (9)$$

However, the signal is transmitted by the reflecting of  $K$  elements of RIS from each reflecting element to each user. Therefore, there is  $K$  parallel separate cascaded channels that should be taken into consideration for the signals' transmission.

Let  $\varphi = \sum_{n=1}^K \varphi_n$  to denote the  $K$  parallel cascaded channels. According to ([45], Sec. 2.2.2), the first term of the Laguerre polynomials can be used to approximate the PDF of  $\varphi$  and  $\varphi^2$ , which are, respectively, given by

$$f_\varphi(x) = \frac{x^a}{b^{a+1} \Gamma(a+1)} \exp\left(-\frac{x}{b}\right), \quad (10)$$

$$f_{\varphi^2}(x) = \frac{1}{2\sqrt{x} b^{a+1} \Gamma(a+1)} \exp\left(-\frac{\sqrt{x}}{b}\right), \quad (11)$$

, respectively. Where  $a = K[E(\varphi_n)]^2 / Var(\varphi_n) - 1$ ,  $b = Var(\varphi_n) / E(\varphi_n)$ . Calculated on the basis of (11), and with the definition of the lower incomplete gamma function  $\gamma(\alpha, x) = \int_0^x e^{-t} t^{\alpha-1} dt$  ([44], Eq. (8.350.1)), the cumulative distribution function (PDF) of  $\varphi^2$  can be given by

$$F_{\varphi^2}(y) = \frac{\gamma(a+1, (\sqrt{y}/b))}{\Gamma(a+1)}. \quad (12)$$

### 3. Outage Probability

Outage probability is an important performance evaluation metric for wireless communication systems. In communications, if the signals' transmission rate is higher than the channel capacity, then the transmitted signal cannot be received completely and correctly. In another word, when the signals' transmission rate is lower than the service reliability transmission rate, the outage occurs. In this section, the outage performance of the RIS-assisted non-terrestrial NOMA networks is analyzed, the outage probability for the nearby user  $n$  with ipSIC/pSIC and distant user  $m$  are discussed in detail.

**3.1. The Outage Probability of Nearby User  $n$ .** According to NOMA protocol, for the nearby user  $n$ , during the SIC process, the SIC receiver should first detect and decode the signal of the distant user  $m$ , and then decode its own signal. Therefore, outage occurs when 1) user  $n$  cannot detect the signal  $x_m$ ; 2) user  $n$  can detect the signal  $x_m$  but cannot detect the signal  $x_n$ . The outage probability of the nearby user  $n$  can be written as

$$P_{D_n}^{ipSIC} = \Pr(\gamma_{D_n \rightarrow D_m} < \gamma_{th_m}) + \Pr(\gamma_{D_n \rightarrow D_m} > \gamma_{th_m}, \gamma_{D_n} < \gamma_{th_n}), \quad (13)$$

where  $\gamma_{th_n} = 2^{R_n} - 1$  and  $\gamma_{th_m} = 2^{R_m} - 1$  represent the target SNRs of user  $n$  and user  $m$  for detecting and decoding the signals  $x_n$  and  $x_m$ , with  $R_n$  and  $R_m$  being the target rate for  $D_n$  to detect  $x_n$  and  $x_m$ , respectively. For further explanation, the following theorem illustrates the outage probability of nearby user  $n$

**Theorem 1.** *Under large-scale fading and Nakagami- $m$  fading cascaded channels, the approximate expression for outage probability of the nearby user  $n$  with ipSIC for RIS-assisted non-terrestrial NOMA networks is given by*

$$P_{D_n}^{ipSIC} \approx S_n \sum_{l=0}^{M-n} \sum_{u=0}^U M-n_l \times \frac{(-1)^l A_u}{(n+l) \Gamma(a+1)^{n+l}} \left[ \gamma\left(a+1, \frac{G_u}{\bar{\lambda} b}\right) \right]^{n+l}, \quad (14)$$

where  $\bar{\omega} = 1$ ,  $G_u = \sqrt{(\Omega_l \bar{\omega} \rho x_u + 1) \theta}$ ,  $S_n = M! / (M-n)! (n-1)!$ ,  $\theta = \gamma_{th_n} / a_n \rho$ . With the help of Gauss-Laguerre quadrature,  $x_u$  is the abscissas and the  $u$ -th zero point of Laguerre polynomial  $L_u(x_u)$ ,  $A_u$  is the  $u$ -th weight, and can be denoted by  $A_u = [(U+1)! / x_u [L'_{U+1}(x_u)]^2]$ ,  $u = 0, 1, \dots, U$ .

Specifically,  $M$  is the total number of users,  $n$  is the order of the current user.

*Proof.* By using the definition expression of the outage event of the nearby user  $n$  and substituting (2) and (3) into (13), the outage probability of user  $n$  with ipSIC can be given by (15) at the top of the next page.

$$P_{D_n}^{ipSIC} = \Pr\left(\frac{\rho|\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_m}{\rho|\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_n + 1} < \gamma_{th_m}\right) + \Pr\left(\frac{\rho|\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_m}{\rho|\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_n + 1} > \gamma_{th_m}, \frac{\rho|\mathbf{h}\Theta\mathbf{g}_n^T|^2 a_n}{\omega\rho|h_I|^2 + 1} < \gamma_{th_n}\right). \quad (15)$$

And do further calculations, it can be derived as follows:

$$P_{D_n}^{ipSIC} = \Pr\left[0 < |\mathbf{h}\Theta\mathbf{g}_n^T|^2 < (\omega|h_I|^2\rho + 1)\theta\right] = \int_0^\infty \int_0^{(x\omega\rho+1)\theta} f_{|h_{LI}|^2}(x)\hat{f}_{\varphi^2}(y)dx dy \quad (16) = \int_0^\infty \hat{F}_{\varphi^2}[(x\omega\rho + 1)\theta] \frac{1}{\Omega_I} e^{-(x/\Omega_I)} dx,$$

where  $\hat{F}_{\varphi^2}(\cdot)$  means the CDF of the sorted channel. According to [46], the relationship between the PDFs of the post-sorted and pre-sorted channels can be expressed as follows

$$\hat{F}_{\varphi^2}(x) = \frac{M!}{(M-p)!(p-1)!} \sum_{l=0}^{M-p} M-p_l \times \frac{(-1)^l}{p+l} [F_{\varphi^2}(x)]^{p+l}. \quad (17)$$

In this formula,  $M$  represents the total number of users,  $p$  is the order of the current user. Finally, substituting (12) into (17) and recombining (16), (14) can be derived. The proof is completed.  $\square$

**Corollary 2.** When it comes to the special condition of  $\omega = 0$ , the approximate expression for outage probability of the nearby user  $n$  with pSIC scheme for RIS-assisted non-terrestrial NOMA networks can be expressed by

$$P_{D_n}^{pSIC} \approx S_n \sum_{l=0}^{M-n} M-n_l \frac{(-1)^l}{n+l} \left[ \frac{\gamma(a+1, (\sqrt{\theta}\lambda b))}{\Gamma(a+1)} \right]^{n+l}. \quad (18)$$

**3.2. The Outage Probability of Distant User  $m$ .** For the distant user  $m$ , it only needs to detect and decode its own signal and treat the signal from the nearby user  $n$  as interference. So, when the distant user  $m$  cannot detect and decode its own signal, the outage occurs. Therefore, the outage probability of the distant user  $m$  can be expressed as

$$P_{D_m} = \Pr(\gamma_{D_m} < \gamma_{th_m}). \quad (19)$$

**Theorem 3.** Under large-scale fading and Nakagami- $m$  fading cascaded channels, the approximate expression for outage probability of the distant user  $m$  for RIS-assisted non-terrestrial NOMA networks is given by

$$P_{D_m} \approx S_m \sum_{l=0}^{M-m} M-m_l \frac{(-1)^l}{(n+l)\Gamma(a+1)^{m+l}} \times \left[ \gamma\left(a+1, \frac{\sqrt{\gamma^*}}{\lambda b}\right) \right]^{m+l}, \quad (20)$$

where  $S_m = M!/(M-m)!(m-1)!$ ,  $\gamma^* = \gamma_{th_m} / \rho(a_m - \gamma_{th_m} a_n)$

*Proof.* By substituting (4) into (19), the outage probability of  $D_n$  can be calculated as

$$P_{D_m} = \Pr\left(\frac{\rho|\mathbf{h}\Theta\mathbf{g}_m^T|^2 a_m}{\rho|\mathbf{h}\Theta\mathbf{g}_m^T|^2 a_n + 1} < \gamma_{th_m}\right) = \Pr\left[|\mathbf{h}\Theta\mathbf{g}_m^T|^2 < \frac{\gamma_{th_m}}{\rho(a_m - \gamma_{th_m} a_n)}\right] = \int_0^{\gamma^*} \hat{f}_{\varphi^2}(y) dy, \quad (21)$$

where  $\hat{f}_{\varphi^2}(y)$  is the PDF of the sorted channel gain, according to [46], it can be defined as

$$\hat{f}_{\varphi^2}(x) = \frac{M!}{(M-p)!(p-1)!} f_{\varphi^2}(x) \times [F_{\varphi^2}(x)]^{p-1} [1 - F_{\varphi^2}(x)]^{M-p}. \quad (22)$$

Specifically,  $\int_0^x \hat{f}_{\varphi^2}(t) dt = \hat{F}_{\varphi^2}(x)$ . Then, combining (12), (17) and (21), (20) can be derived.  $\square$

**3.3. The Outage Probability of the OMA Benchmark.** For RIS-assisted non-terrestrial OMA networks, there is one time slot in the transmission of communications. Specifically, with the assistance of the RIS, the entire transmission process consists of the satellite sending the signals to two users in the same time slot. Hence the outage probability of the  $p$ -th user is defined to occur as the instantaneous SNR of the network, represented by  $\gamma_p^{OMA}$ , is less than a certain threshold in the slot while communicating. The outage probability of the RIS-assisted non-terrestrial NOMA networks via OMA benchmark can be expressed as

$$P_{D_p}^{OMA} = \Pr(\gamma_p^{OMA} < \gamma_{th}^{OMA}), \quad (23)$$

where  $\gamma_{th}^{OMA} = 2^{R_p} - 1$  is the target SNR of the  $p$ -th user to detect and decode the signals.

**Theorem 4.** Under large-scale fading and Nakagami- $m$  fading cascaded channels, the approximate expression for outage probability of the  $p$ -th user for RIS-assisted non-terrestrial

TABLE 1: The fixed numerical values of the parameters.

Monte Carlo simulation repeated	$10^6$ iterations
Two users' power allocations	$a_n = 0.4$ $a_m = 0.6$
Two users' target rates	$R_n = 0.05\text{BPCU}$ $R_m = 0.04\text{BPCU}$
The lobe gains of Users' antennas	$G_i = 12\text{dB}$
The lobe gains of satellite's antennas	$G_t = 80\text{dB}$
Carrier frequency	$f_c = 3 \times 10^8\text{Hz}$

OMA networks is given by

$$P_{D_p}^{OMA} \approx S_p \sum_{l=0}^{M-p} M-p_l \frac{(-1)^l}{(p+l)[\Gamma(a+1)]^{p+l}} \times \left[ \gamma \left( a+1, \frac{\sqrt{\Theta_{oma}}}{\lambda b} \right) \right]^{p+l}, \quad (24)$$

where  $S_p = M!/(M-p)!(p-1)!$ ,  $\Theta_{oma} = \gamma_{th}^{OMA}/\rho$

**3.4. Diversity Analysis.** For further evaluation of system performance, the diversity order is taken to be analysed. The diversity order is the limit value when the SNR tends to be infinity, and is particularly important to evaluate system performance under high SNR, because it is approximately the slope of the outage probability curve at high SNR regime. Thus, it can be used to assess the decreasing rate of the outage probability for communication systems. The diversity order can be expressed as

$$d = - \lim_{\rho \rightarrow \infty} \frac{\log [P^{\infty}(\rho)]}{\log \rho}, \quad (25)$$

where  $P^{\infty}(\rho)$  denotes the asymptotic outage probability.

**Corollary 5.** *The asymptotic outage probability expression of the nearby user n with ipSIC for RIS-assisted non-terrestrial NOMA networks when  $\rho \rightarrow \infty$  can be given by*

$$P_{D_n}^{\infty, ipSIC} = S_n \sum_{l=0}^{M-n} \sum_{u=0}^U M-n_l \frac{(-1)^l A_u}{(n+l)[\Gamma(a+1)]^{n+l}} \times \left[ \gamma \left( a+1, \frac{\sqrt{\varpi \Omega_t \theta_n^p x_u}}{\lambda b} \right) \right]^{n+l}, \quad (26)$$

where  $\varpi = 1$ ,  $\theta_n^p = \gamma_{th_n}/a_n$

**Remark 6.** Upon substituting (26) into (25), a zero diversity order for the nearby user n with ipSIC can be derived, this signifies that the outage probability curve eventually converges to a certain value in the high SNR regime. This phenomenon is due to the influence of the residual interference from ipSIC during the detecting and decoding process.

**Corollary 7.** *The asymptotic outage probability expression of the nearby user n with pSIC for RIS-assisted non-terrestrial NOMA networks when  $\rho \rightarrow \infty$  can be given by*

$$P_{D_n}^{\infty, pSIC} = \frac{S_n^{\infty} \theta^{n(a+1)/2}}{(\lambda b)^{n(a+1)} [(a+1)\Gamma(a+1)]^n} \propto \frac{1}{\rho^{n(a+1)/2}}, \quad (27)$$

where  $S_n^{\infty} = M!/(M-n)!n!$

*Proof.* According to the series representation of the lower incomplete gamma function ([44], Eq. (8.354.1)), the gamma function part of  $P_{D_n}^{pSIC}$  can be further rewritten as

$$\gamma \left( a+1, \frac{\sqrt{\theta}}{\lambda b} \right) = \sum_{n=0}^{\infty} \frac{(-1)^n \left( \frac{\sqrt{\theta}}{\lambda b} \right)^{a+1+n}}{n!(a+1+n)}. \quad (28)$$

When  $\rho \rightarrow \infty$ ,  $\theta \rightarrow 0$ . Then, by substituting (28) into (18), and further taking the first term of each series representation, i.e.,  $n=0$  and  $l=0$ , we can obtain (27). The proof is completed.  $\square$

**Remark 8.** Upon substituting (27) into (25), the diversity order of the nearby user n with pSIC is equal to  $nK [E(\varphi_n)]^2/2\text{Var}(\varphi_n)$ , which is related to the order of the nearby user, and Nakagami-m parameter  $m'$  and  $\Omega$

**Corollary 9.** *The asymptotic outage probability expression of the distant user m for RIS-assisted non-terrestrial NOMA network when  $\rho \rightarrow \infty$  can be given by*

$$P_{D_m}^{\infty} = \frac{S_m^{\infty} \gamma^{*m(a+1)/2}}{(\lambda b)^{m(a+1)} [(a+1)\Gamma(a+1)]^m} \propto \frac{1}{\rho^{m(a+1)/2}}, \quad (29)$$

where  $S_m^{\infty} = M!/(M-m)!m!$

**Remark 10.** By substituting (29) into (25), it can be derived that the diversity order of the distant user m is equal to  $mK [E(\varphi_n)]^2/2\text{Var}(\varphi_n)$ , which is also related to the order of the distant user, and Nakagami-m parameter  $m'$  and  $\Omega$

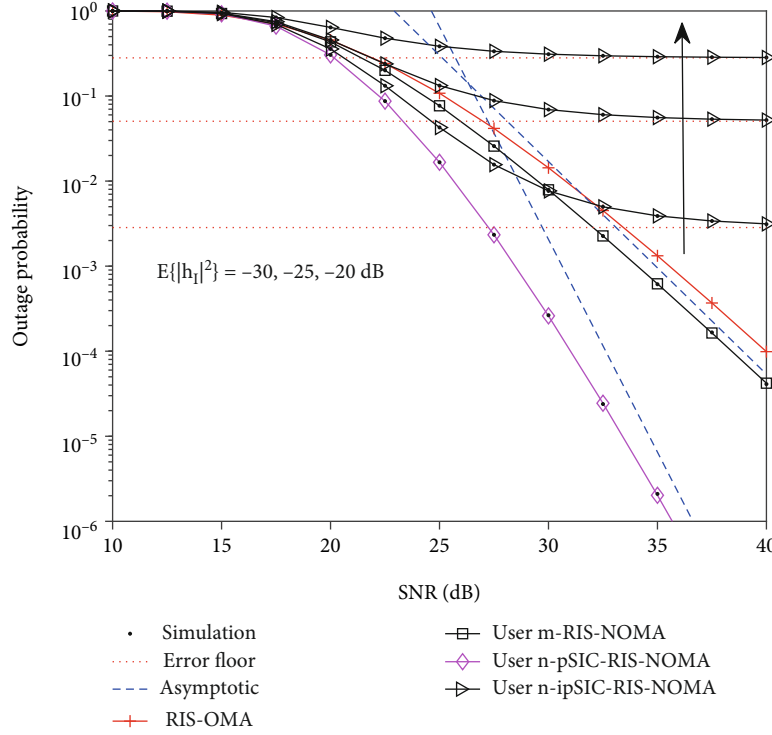


FIGURE 2: Outage probability versus the transmit SNR, with  $K = 5$ ,  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

#### 4. Simulation and Numerical Results

In this section, simulation analysis of the outage probability for RIS-assisted non-terrestrial NOMA networks is provided to adequately confirm the previous conclusions from the above subsections. Listing statistics for convenience, Table 1 shows the fixed numerical values of the parameters used in this paper [9, 39, 43] and BPCU is the abbreviation of bit per channel use. Without loss of generality, the method of controlling variables is adopted, we just change only a certain parameter to analyze the probability of outage in this scenario for a more intuitive expression. To make the scenario more specific to evaluate performance, in the simulations, the total number of users  $M$  is set as three, the  $n$ -th user is set as the first user and the  $m$ -th user is set as the second user (It is set that  $M = 3$ ,  $m = 2$ ,  $n = 1$ ). In addition, RIS-assisted non-terrestrial OMA is also select to be included in the analysis as a benchmark for comparison. It is worth pointing that the detecting target rate for OMA benchmark of the entire networks is set to be  $R_m + R_n$

Figure 2 illustrates the outage probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $K = 5$ ,  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the condition of different value of residual interference for ipSIC scheme at the same time. As shown in the figure, according to (14) and (18), the expression of the outage probability for user  $n$  with ipSIC/pSIC can be plotted as the right triangle and diamond solid curves, respectively. According to (20), the exact expression of the outage probability for user  $m$  can be plotted as the square solid curve, and the cross curve

for outage probability of the RIS-assisted non-terrestrial OMA is plotted based on (24). The outage probability curve obtained from the numerical simulation results fits perfectly with the curve drawn by the formula derived from the above subsections. It can be observed that for the nearby user  $n$ , the outage performance is better in the pSIC scheme than ipSIC scheme because of the existence of residual interference, and when the residual interference gradually decreases, under a certain range of SNR, the curves gradually fit. For the comparison between RIS-assisted non-terrestrial NOMA and OMA, it can be proofed that NOMA scheme performs better than OMA scheme, it is because that in NOMA scheme, multiple users receive the spectrum resource with power allocation to bring about better fairness relative to OMA [7], and SIC process at the receive-end can also help to increase the received SINR. According to (26), (27) and (29), the exact expression of the asymptotic outage probability for user  $n$  with pSIC and user  $m$  can be plotted as the blue dotted curve and that of the user  $n$  with ipSIC can be plotted as the red dotted curve, respectively. The asymptotic curves can well represent the performance of the networks under high SNR, so that the results can be displayed more intuitively. It can be observed that through the comparison between the asymptotic outage probability of user  $n$  with pSIC and user  $m$ , the former performs better than the latter. Confirmed by Remark 6, the outage probability of user  $n$  with ipSIC eventually gets a zero diversity order and converges to an error floor because of the impact of the residual interference. Hence, the outage performance of the nearby user  $n$  with ipSIC is worse than that of other



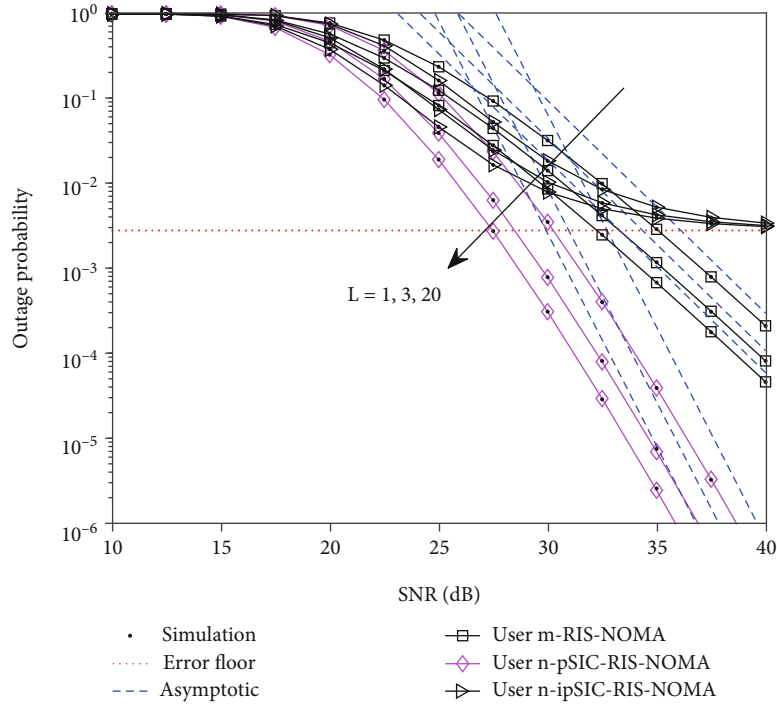


FIGURE 3: Outage probability versus the transmit SNR for various length of training symbols, with  $K = 5$ ,  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_I|^2\} = -30$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

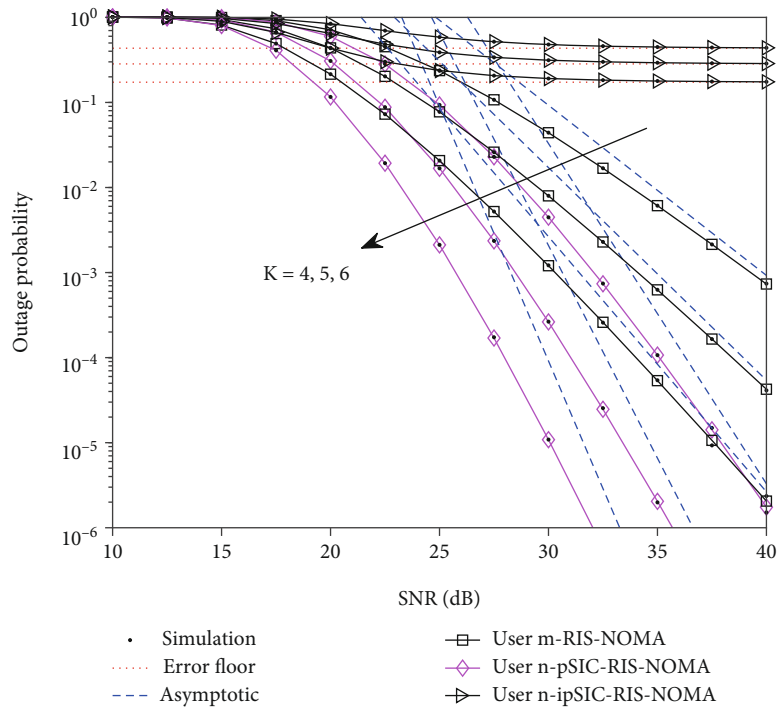


FIGURE 4: Outage probability versus the transmit SNR, with  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_I|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

users' conditions, and it is kind crucial to take the residual interference into consideration when it comes to a practical communication scenario.

Under actual channel environment conditions, it is not easy to estimate channel state information (CSI) between the satellite and terrestrial nodes, so we also consider the imperfect

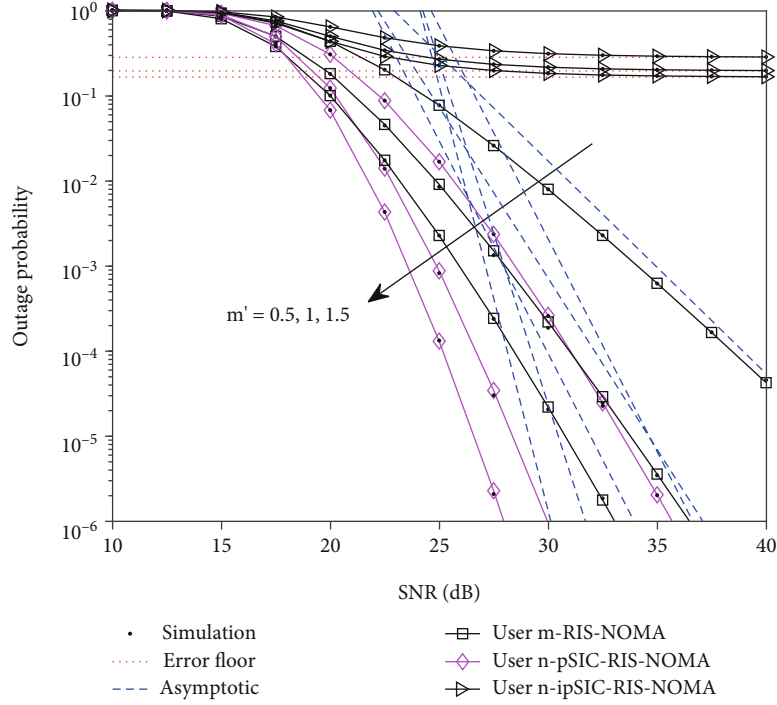


FIGURE 5: Outage probability versus the transmit SNR, with  $K = 5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

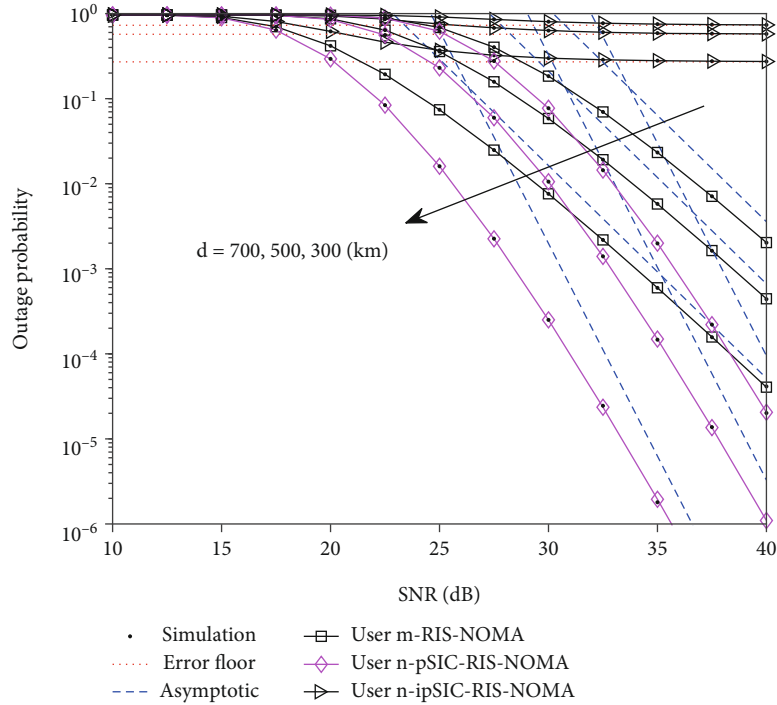


FIGURE 6: Outage probability versus the transmit SNR, with  $K = 5$ ,  $m' = 0.5$ ,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

CSI scenario. The receiver uses the minimum mean square error to estimate the channel coefficients, which can be modeled as  $\bar{h}\Theta\mathbf{g}_p^T = h\Theta\mathbf{g}_p^T + e_k$ , where  $e_k$  denote the estimated channel error with  $e_k \sim CN(0, \sigma_{ek}^2)$ .  $L$  represents the lengths of training symbols for CSI estimation and the variance of esti-

ated channel error can be denoted by  $\sigma_{ek}^2 = 1/\rho L$ . Figure 3 illustrates the outage probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $K = 5$ ,  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -30$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the

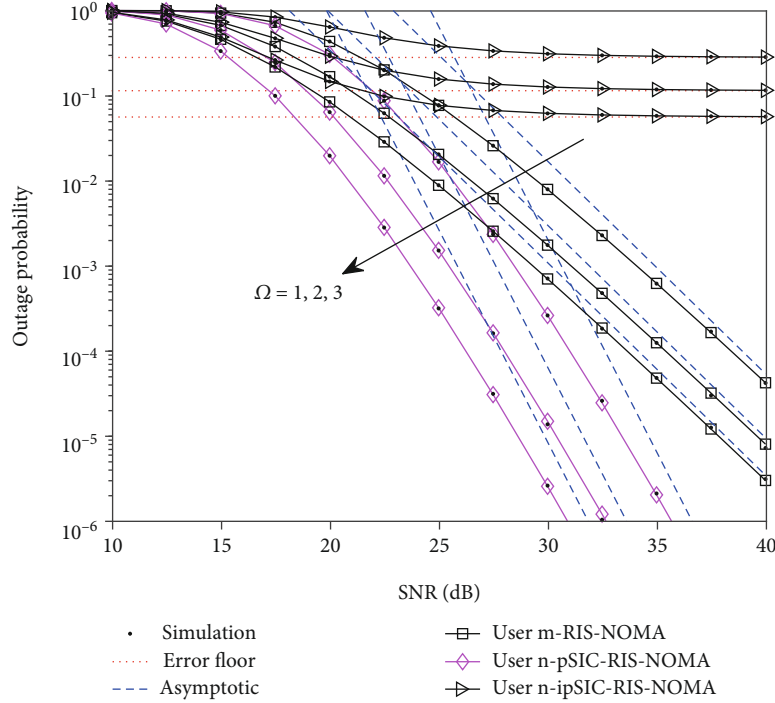


FIGURE 7: Outage probability versus the transmit SNR, with  $K = 5$ ,  $d = 300$  km,  $m' = 0.5$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU

condition of different value of training times at the same time. The situation becomes ideal CSI as  $L$  approaches infinity. Under high SNR, the outage probability of nearby user with ipSIC scheme tends to be consistent with different training amounts due to the impact of the residual interference. It can be observed from the simulation that with the increase of  $L$ , the improvement of outage performance tends to be slow. Therefore in practical applications, we can also increase the lengths of the training symbols to approach better performance.

The impact of the number of reflecting elements on RIS  $K$  for the performance of the networks is significant, and it is shown in Figure 4, which plots the outage probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $m' = 0.5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the condition of different number of reflecting elements on RIS,  $K$ , at the same time. It can be seen intuitively from the figure, the outage performance is significantly affected by the number of reflecting elements on RIS for a specification of more reflecting elements, smaller outage probability of user  $n$  and user  $m$  in the RIS-assisted non-terrestrial NOMA networks. Meanwhile, in line with Remark 8 and Remark 10, it can be observed that diversity orders of non-orthogonal users are affected by the reflecting elements on RIS,  $K$ . When  $K$  grows, the slope of the diversity orders for both user  $n$  and user  $m$  increase. It is worth mentioning that when it comes to the sorted channel, the diversity order will be related to the ordinal number of the current user. Therefore, in practical applications, we can change the value of  $K$  according to the change of the actual

channel environment in order to achieve a better communication effect. Figure 5 illustrates the outage probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $K = 5$ ,  $d = 300$  km,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the condition of different value of Nakagami- $m$  shaping parameter  $m'$  at the same time. As we can see from the figure, outage performance gradually gets better as Nakagami- $m$  shaping parameter  $m'$  increases. For Nakagami- $m$  distribution, Nakagami- $m$  shaping parameter  $m'$  represents the fading environment, when  $m'$  grows, the fading environment gets better, when  $m$  approaches infinity, it means no fading. Hence it is needed to carefully consider the shaping parameter  $m'$  of the environmental channel when we apply it to the actual scene.

Figure 6 illustrates the outage probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $K = 5$ ,  $m' = 0.5$ ,  $\Omega = 1$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the condition of different value of distance from satellite to ground at the same time. The distance from satellite to ground determines the path loss factor in large-scale fading scenarios, when the distance increases, the path loss factor increases. Therefore, it can be clearly seen from the figure that when the distance decreases, the outage performance of the non-orthogonal users gradually improves. However, because the distance to the ground of the satellite is standardized, it should be considered according to practical scenarios. Taking consideration of another factor for Nakagami- $m$  fading, Figure 7 illustrates the outage

probability of the RIS-assisted non-terrestrial NOMA networks versus the transmit SNR with settings of  $K = 5$ ,  $d = 300$  km,  $m' = 0.5$ ,  $E\{|h_l|^2\} = -20$  dB,  $R_n = 0.05$  and  $R_m = 0.04$  BPCU, and compares the condition of different value of scale parameter of Nakagami- $m$  fading  $\Omega$  at the same time. The actual meaning of  $\Omega$  is the fading power of Nakagami- $m$  fading, its physical meaning is the mean of the square of the signal envelope, when the value of  $\Omega$  increases, the channel condition will become better, and the outage performance of the users become better.

## 5. Conclusion

In this paper, a RIS-assisted non-terrestrial NOMA networks with large-scale fading and Nakagami- $m$  fading cascaded channels has been discussed in detail. More specifically, the approximate expression for outage probability of nearby user  $n$  with ipSIC/pSIC and distant user  $m$ , and asymptotic expression for outage probability of user  $n$  with pSIC and user  $m$  under the sorted user channel schemes have been derived carefully. Based on the analysis of the outage performance, the diversity order is also obtained to evaluate the outage performance. Meanwhile, the influence of each variable to the system performance is analysed. Eventually, through detailed study of this paper, we can find that the performance of RIS-assisted non-terrestrial NOMA is better than that of OMA, and the cascaded channel scenario can obtain a more ideal transmission effect through the propagation of RIS and the optimization of parameters in the future studying.

## Data Availability

The calculation and simulation data used to support the findings of this study are included within the article.

## Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant 62071052 and grant 62201533 and the R&D Program of Beijing Municipal Education Commission under grant KM202011232003.

## References

- [1] Y. Ji, J. Zhang, Y. Xiao, and Z. Liu, "5G flexible optical transport networks with large-capacity, low-latency and high-efficiency," *China Communications*, vol. 16, no. 5, pp. 19–32, 2019.
- [2] X. You, C. X. Wang, J. Huang et al., "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, 2021.
- [3] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.
- [4] X. Liu, B. Lin, M. Zhou, and M. Jia, "NOMA-Based cognitive spectrum access for 5G-enabled internet of things," *IEEE Network*, vol. 35, no. 5, pp. 290–297, 2021.
- [5] X. Pei, Y. Chen, M. Wen, H. Yu, E. Panayirci, and H. V. Poor, "Next-Generation multiple access based on NOMA with power level modulation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1072–1083, 2022.
- [6] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, 2014.
- [7] Z. Ding, Y. Liu, J. Choi et al., "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.
- [8] X. Yue, Z. Qin, Y. Liu, S. Kang, and Y. Chen, "A unified framework for non-orthogonal multiple access," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5346–5359, 2018.
- [9] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Exploiting full/half-duplex user relaying in NOMA systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 560–575, 2018.
- [10] D. Wan, M. Wen, F. Ji, Y. Liu, and Y. Huang, "Cooperative NOMA systems with partial channel state information over nakagami- $m$  fading channels," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 947–958, 2018.
- [11] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1656–1672, 2017.
- [12] W. Chen, H. Ding, S. Wang, D. B. da Costa, F. Gong, and P. H. Juliano Nardelli, "Backscatter cooperation in NOMA communications systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3458–3474, 2021.
- [13] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple Semi-Grant-Free transmission strategies assisted by non-orthogonal multiple access," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4464–4478, 2019.
- [14] R. Huang, D. Wan, F. Ji et al., "Performance analysis of NOMA-based cooperative networks with relay selection," *China Communications*, vol. 17, no. 11, pp. 111–119, 2020.
- [15] D. Wan, R. Huang, M. Wen, G. Chen, F. Ji, and J. Li, "A simple multicarrier transmission technique combining transmit diversity and data multiplexing for non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 7216–7220, 2021.
- [16] G. Araniti, A. Iera, S. Pizzi, and F. Rinaldi, "Toward 6G non-terrestrial networks," *IEEE Network*, vol. 36, no. 1, pp. 113–120, 2022.
- [17] S. Chen, S. Sun, and S. Kang, "System integration of terrestrial mobile communication and satellite communication — the trends, challenges and key technologies in B5G and 6G," *China Communications*, vol. 17, no. 12, pp. 156–171, 2020.
- [18] V. Singh and P. K. Upadhyay, "Exploiting FD/HD cooperative-NOMA in underlay cognitive hybrid satellite-terrestrial networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 246–262, 2022.

- [19] N. Wang, F. Li, D. Chen, L. Liu, and Z. Bao, "NOMA-based Energy-Efficiency optimization for UAV enabled space-air-ground integrated relay networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4129–4141, 2022.
- [20] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "UAV-Enabled communication using NOMA," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5126–5138, 2019.
- [21] R. Ge, D. Bian, J. Cheng, K. An, J. Hu, and G. Li, "Joint user Pairing and power allocation for NOMA-based GEO and LEO satellite network," *IEEE Access*, vol. 9, pp. 93255–93266, 2021.
- [22] Z. Gao, A. Liu, and X. Liang, "The performance analysis of downlink NOMA in LEO satellite communication system," *IEEE Access*, vol. 8, pp. 93723–93732, 2020.
- [23] X. Yue, Y. Liu, Y. Yao et al., "Outage behaviors of NOMA-based satellite network over Shadowed-Rician fading channels," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6818–6821, 2020.
- [24] X. Zhang, B. Zhang, K. An et al., "Outage performance of NOMA-based cognitive hybrid satellite-terrestrial overlay networks by amplify-and-forward protocols," *IEEE Access*, vol. 7, pp. 85372–85381, 2019.
- [25] L. Han, W.-P. Zhu, and M. Lin, "Outage of NOMA-based hybrid satellite-terrestrial multi-antenna DF relay networks," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1083–1087, 2021.
- [26] Y. Liu, X. Liu, X. Mu et al., "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Communication Surveys and Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.
- [27] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. D. Renzo, and F. Chen, "Reconfigurable intelligent surfaces with reflection pattern modulation: beamforming design and performance analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 741–754, 2021.
- [28] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surfaces: potentials, applications, and challenges for 6G wireless networks," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 184–191, 2021.
- [29] A.-A. A. Boulogeorgos and A. Alexiou, "Performance analysis of reconfigurable intelligent surface-assisted wireless systems and comparison with relaying," *IEEE Access*, vol. 8, pp. 94463–94483, 2020.
- [30] C. Zhang, W. Yi, Y. Liu, K. Yang, and Z. Ding, "Reconfigurable Intelligent surfaces aided multi-cell NOMA networks: a stochastic geometry model," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 951–966, 2022.
- [31] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and G. K. Karagiannidis, "Nonorthogonal multiple access (NOMA) with multiple intelligent reflecting surfaces," *IEEE Transactions on Communications*, vol. 20, no. 11, pp. 7184–7195, 2021.
- [32] X. Yue and Y. Liu, "Performance analysis of intelligent reflecting surface assisted NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2623–2636, 2022.
- [33] Z. Ding and H. Vincent Poor, "A simple design of IRS-NOMA transmission," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1119–1123, 2020.
- [34] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, 2022.
- [35] X. Zhou, K. Ying, S. Liu, M. Ke, Z. Gao, and M.-S. Alouini, "Reconfigurable intelligent surface assisted grant-free massive access," *Intell and Converged Network*, vol. 3, no. 1, pp. 134–143, 2022.
- [36] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4962–4975, 2021.
- [37] N. Bhargav, C. R. N. da Silva, Y. J. Chun, E. J. Leonardo, S. L. Cotton, and M. D. Yacoub, "On the product of two  $\kappa$  random variables and its application to double and composite fading channels," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2457–2470, 2018.
- [38] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1596–1600, 2020.
- [39] T. Li, X. Hao, and X. Yue, "A power domain multiplexing based cocarrier transmission method in hybrid satellite communication networks," *IEEE Access*, vol. 8, pp. 120036–120043, 2020.
- [40] X. Yue, J. Xie, Y. Liu, Z. Han, R. Liu, and Z. Ding, "Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface Assisted NOMA Networks," (2021), <https://arxiv.org/abs/2112.01336>.
- [41] H. Wang, Z. Shi, Y. Fu, and S. Fu, "On intelligent reflecting Surface-Assisted NOMA uplinks with imperfect SIC," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1518–1522, 2022.
- [42] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, Publishing House of Electronics Industry, Beijing, China, 2013.
- [43] M. D. Yacoub, "The  $\kappa$ - $\mu$  distribution and the  $\eta$ - $\mu$  distribution," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 1, pp. 68–81, 2007.
- [44] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New York, NY, USA, 6th edition, 2000.
- [45] V. Primak and V. Lyandres, *Stochastic Methods and their Applications to Communications: Stochastic Differential Equations Approach*, Wiley, West Sussex, U.K., 2004.
- [46] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley, Hoboken, NJ, USA, 3rd edition, 2003.

## Research Article

# Multiple Dimensional Encoding/Modulation Shift-and-Addition Design for Distributed Systems

Mingjun Dai,<sup>1</sup> Chanting Zhang,<sup>1</sup> Zhaoyan Hong ,<sup>2</sup> Paweł Wawrzyński,<sup>3</sup> Tomasz Trzciński,<sup>3</sup> and Xiaohui Lin<sup>1</sup>

<sup>1</sup>College of Electronic and Information Engineering, Shenzhen University, 518000 Shenzhen, China

<sup>2</sup>College of Mathematics and Statistics, Shenzhen University, 518000 Shenzhen, China

<sup>3</sup>Institute of Computer Science, Warsaw University of Technology, 00-665 Warsaw, Poland

Correspondence should be addressed to Zhaoyan Hong; 1546623526@qq.com

Received 20 July 2022; Accepted 27 August 2022; Published 15 September 2022

Academic Editor: Jun Li

Copyright © 2022 Mingjun Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In distributed computing/storage/machine learning system, the method of encoding and decoding combining shift-and-addition (SA) and zigzag decoding (ZD) is proposed to solve the problem of high computational complexity. However, in each encoded packet, one element takes part in the encoding only once, so the obtained overhead is extremely high. In this work, based on the idea of multidimensional encoding/modulation, we propose to employ one element of the encoding process multiple times when constructing one encoded packet based on the Cauchy matrix, thereby leveraging the favourable properties of the code based on Cauchy matrix. The overhead is reduced from square to logarithmic in certain parameters. Compared with the overhead of the existing square computational complexity, it is greatly reduced.

## 1. Introduction

In the era of big data [1], the amount of data is growing at a doubling rate annually. The way of data processing has been shifted from the centralized data processing to the distributed data processing. However, in distributed applications, not all devices are reliable. Some devices may fail to work, or the performance of the devices is not consistent. In any task of data processing, there will be some unreliable devices whose computing speed is slower than the average speed, which are called stragglers [2, 3]. For example, in a data center of Facebook, more than 100 nodes may fail per day [4, 5]. The completion time of data processing is constrained by the slowest working node. Therefore, how to deal with stragglers becomes a challenge for data processing. To solve this problem, network coding techniques have been developed, and the code with combination property (CP) is proposed:  $k$  original packets are encoded into  $n$  packets, where  $n > k$ , and any  $k$  out of these  $n$  packets are able to recover the original data. The code with CP has been widely used in distributed systems, including distributed storage (DS)

[6–10], distributed computing (DC) [11–16], and distributed machine learning [17].

In distributed systems, linear code is adopted in most of the coding technologies, but linear code involves a lot of multiplication and division operations, which greatly increase the complexity of coding and decoding. For the sake of low computation complexity, a kind of CP-ZD code [18–20] with CP is proposed, which combines shift-and-addition (SA) and zigzag decoding (ZD) [21]. However, for the case where one element takes part in the encoding only once when constructing an encoded packet, the overhead is as high as square of the parameters' number (*not*  $k$ ).

Multidimensional encoding/modulation promises high data rate [18, 19], which has been used as promising technique in communication [20] and distributed systems [22, 23]. As a result, to further reduce the overhead, based on the idea of multidimensional encoding/modulation, we propose the idea of one element taking part in the encoding process multiple times when constructing one encoded packet. Using the properties of the code based on Cauchy matrix in finite field, we design a framework of one element

taking part in the encoding process multiple times for constructing an encoded packet. The overhead of this coding framework reduces from square to logarithmic with respect to the parameter. Specifically, the idea that one element takes part in the encoding only once in each encoded packet can be interpreted as each source packet being treated as an element and occurring at most once in a coded packet. Similarly, the idea that one element takes part in the encoding process multiple times when constructing one encoded packet is that each source packet occurs multiple times in an encoded packet, which is added to its own multiple distinct shifts.

## 2. Preliminary

**2.1. Definition of Cauchy Matrix.** Given  $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ , let  $c_{(i,j)} = 1/(x_i + y_j) (1 \leq i, j \leq n)$ , then the matrix  $C = (c_{i,j})$  is called the Cauchy matrix [24], and its determinant is as follows:

$$\det(C) = \frac{\prod_{1 \leq i < j \leq n} (x_j - x_i)(y_j - y_i)}{\prod_{i=1}^n \prod_{j=1}^n (x_i + y_j)}. \quad (1)$$

Similarly, a Cauchy matrix over a finite field is defined as follows: let  $X$  and  $Y$  be two sets of elements in a finite field. Among them,  $X = \{x_1, x_2, \dots, x_p\}$  and  $Y = \{y_1, y_2, \dots, y_q\}$ . If for  $\forall i \in \{1, 2, \dots, p\}, \forall j \in \{1, 2, \dots, q\}$ , the following is satisfied:

- (1)  $x_i + y_j \neq 0$
- (2)  $\forall i, j \in \{1, 2, \dots, p\}, i \neq j, x_i \neq x_j$
- (3)  $\forall i, j \in \{1, 2, \dots, q\}, i \neq j, y_i \neq y_j$

Then, the following matrix is called a Cauchy matrix over a finite field:

$$G = \begin{bmatrix} \frac{1}{x_1 + y_1} & \frac{1}{x_1 + y_2} & \dots & \frac{1}{x_1 + y_q} \\ \frac{1}{x_2 + y_1} & \frac{1}{x_2 + y_2} & \dots & \frac{1}{x_2 + y_q} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{x_p + y_1} & \frac{1}{x_p + y_2} & \dots & \frac{1}{x_p + y_q} \end{bmatrix}. \quad (2)$$

It is straightforward to obtain the following theorem from the construction rules of the Cauchy matrix:

**Theorem 1.** *When  $G$  is a Cauchy matrix, any square submatrix  $G_l$  of  $G$  is nonsingular, where  $l$  indicates the number of rows and columns of the submatrix  $G_l$  ( $1 \leq l \leq \min(p, q)$ ); then,*

$$\det(G_l) = \frac{\prod_{1 \leq i < j \leq l} (x_j - x_i)(y_j - y_i)}{\prod_{i=1}^l \prod_{j=1}^l (x_i + y_j)} \neq 0. \quad (3)$$

In other words, every submatrix of the Cauchy matrix is invertible.

**2.2. The Arithmetic Operation in Finite Fields.** Finite field is a field with a finite number of elements, for example,  $GF(2^\omega)$  represents a finite field containing  $2^\omega$  elements. Before describing the arithmetic operation in finite fields, we briefly introduce the concept of the primitive polynomial.

The primitive polynomial is essentially a polynomial that cannot be factored. When a finite field determines its primitive polynomials, the arithmetic operations in that field are also determined. In general, the primitive polynomial of a field can be obtained by looking up the table, and the primitive polynomial of a field is not unique. Take the finite field  $GF(2^3)$  as an example, there is more than one primitive polynomial over  $GF(2^3)$ , and the most common primitive polynomial is  $q(z) = z^3 + z + 1$ . Table 1 shows some of the primitive polynomials [25] present in  $GF(2^\omega)$ .

The addition and subtraction operation [22] of finite fields are the XOR operation in polynomial calculation. The rule for adding and subtracting is to XOR coefficients of the same order in two polynomials, and there is no difference between the two operations, such as  $(z^2 + z) + (z + 1) = z^2 + 1$ ,  $(z^2 + z) - (z + 1) = z^2 + 1$ . At present, the multiplication and division operations [23] of finite fields usually count on the look-up tables. Each field has positive and negative tables, which are denoted as *gf log* and *gfi log*, respectively, on the  $GF(2^\omega)$  field. Taking  $GF(2^3)$  as an example, its table *gf log* and *gfi log* are generated as shown in Table 2 [16]:

If the multiplication and division operations are performed on the  $GF(2^3)$  field, as shown in Table 2, the multiplication operation is as follows:

$$\begin{aligned} 2 \times 3 &= \text{gf log} [\text{gf log} [2] + \text{gf log} [3]] \\ &= \text{gf log} [1 + 3] = \text{gf log} [4] = 6, \end{aligned} \quad (4)$$

and the division operation is as follows:

$$\begin{aligned} 1 \div 5 &= \text{gfi log} [\text{gf log} [1] - \text{gf log} [5]] \\ &= \text{gfi log} [0 - 6] = \text{gfi log} [1] = 2. \end{aligned} \quad (5)$$

**2.3. Transformation from Field  $GF(2^\omega)$  to Field  $GF(2)[z]/q(z)$ .** Field  $GF(2^\omega)$  is constructed by finding a primitive polynomial of  $\omega$  degrees on  $GF(2)$  and then enumerating elements (in polynomial form) by using the generating element  $z$ . The addition in this field is performed using polynomial addition, and multiplication is performed using polynomial multiplication and modulo the result with respect to  $q(z)$ , such field  $GF(2^\omega)$  can be written as  $GF(2^\omega) = GF(2)[z]/q(z)$  [25], which can also be said that field  $GF(2^\omega)$  and field  $GF(2^\omega) = GF(2)[z]/q(z)$  are isomorphic [26].

TABLE 1: Table of common primitive polynomials.

Finite field	Primitive polynomial
$GF(2^2)$	$z^2 + z + 1$
$GF(2^3)$	$z^3 + z + 1$
$GF(2^4)$	$z^4 + z + 1$
$GF(2^8)$	$z^8 + z^4 + z^3 + z^2 + 1$
$GF(2^{16})$	$z^{16} + z^{12} + z^3 + z + 1$

TABLE 2:  $gf$  log and  $gfi$  log for  $GF(2^3)$ .

$i$	$gf$ log $[i]$	$gfi$ log $[i]$
0	-	1
1	0	2
2	1	4
3	3	3
4	2	6
5	6	7
6	4	5
7	5	-

The conversion rule for field  $GF(2^\omega)$  to field  $GF(2)[z]/q(z)$  [25] is the conversion of numerical form to polynomial form. Taking  $GF(2^\omega)$  as an example, the implementation steps are as follows:

*Step 1.* Initialize the set as  $\{0, 1, z\}$ .

*Step 2.* Multiply the last element of the set by  $z$ , such as  $z$ , and modulo the result with respect to  $q(z)$  if the resulting degree is greater than or equal to  $\omega$ .

*Step 3.* Continue Step 2 until there are  $2^\omega$  elements in the set, at which point the last element is multiplied by  $z$  and modulo  $q(z)$ , resulting in a value of 1.

To better understand the above steps, let us elaborate on a simple example:

*Example 1.* Suppose  $\omega = 2$ ; then, the original polynomial is  $q(z) = z^2 + z + 1$ ; to construct  $GF(2^2) = GF(4)$ , we initialize the set as  $\{0, 1, z\}$ , so the next element is  $z^2$ ; since the degree of the element is 2, modulo it with respect to  $q(z) = z^2 + z + 1$ , which resulting in  $z + 1$ . Therefore, four elements are generated:  $\{0, 1, z, z + 1\}$ , and the corresponding numerical forms are  $\{0, 1, 2, 3\}$ , which are shown in Table 3. If we continue, we can get the following:

$$(z + 1)z \bmod q(z) = (z^2 + z) \bmod (z^2 + z + 1) = 1. \quad (6)$$

According to Step 3, we can end the enumeration.

**2.4. Mathematical Model.** We want to construct  $(n, k)$  code that possesses the  $(n, k)$  CP. This section adopts the method in reference [27]. We represent each packet as a polynomial

TABLE 3: Transformation of  $GF(4)$  from field  $GF(2^\omega)$  to field  $GF(2^\omega) = GF(2)[z]/q(z)$ .

Generating element	Polynomial form	Numerical form
0	0	0
$z^0$	1	1
$z^1$	$z$	2
$z^2$	$z + 1$	3

of  $GF(2^\omega)$ , where a number is denoted by several bits within this packet. Source packet  $s_i$  is represented with the polynomial form, as shown in Formula (7),  $i \in K \triangleq \{1, 2, \dots, k\}$ ,

$$s_i(z) \triangleq s_{i,1} + s_{i,2}z + s_{i,3}z^2 + \dots + s_{i,L}z^{L-1}, \quad (7)$$

where  $L$  indicates the length of the source packet and  $z$  indicates the right shift by one bit.

For  $i \in K$ , the  $i$ -th encoded packet can be expressed as  $c_i(z) = s_i(z)$ . Let  $m \triangleq n - k$  denote the number of parity packets. For  $i \in M \triangleq \{1, 2, \dots, m\}$ , the polynomial form of the  $i$ -th parity packet is as follows:

$$c_{k+i}(z) \triangleq \alpha_{i,1}(z)s_1(z) + \alpha_{i,2}(z)s_2(z) + \dots + \alpha_{i,k}(z)s_k(z), \quad (8)$$

where  $\alpha_{i,j}(z) \triangleq z^{T_{ij}}$ ,  $i \in M, j \in K$ .

Combined with the systematic packets and parity packets, the final coding expression is shown in the following formula:

$$c(z) = A(z)s(z), \quad (9)$$

where  $c(z) \triangleq (c_1(z), c_2(z), \dots, c_n(z))$ ,  $s(z) \triangleq (s_1(z), s_2(z), \dots, s_k(z))$ , and

$$A(z) \triangleq \begin{bmatrix} I_K(z) \\ T(z) \end{bmatrix}, \quad (10)$$

which is a matrix with dimension of  $n \times k$ ,  $I_k(z)$  is a  $k \times k$  identity matrix, and  $T(z)$  is a  $m \times k$  shift matrix,

$$T(z) \triangleq \begin{bmatrix} z^{T_{11}} & z^{T_{12}} & z^{T_{13}} & \dots & z^{T_{1k}} \\ z^{T_{21}} & z^{T_{22}} & z^{T_{23}} & \dots & z^{T_{2k}} \\ z^{T_{31}} & z^{T_{32}} & z^{T_{33}} & \dots & z^{T_{3k}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z^{T_{m1}} & z^{T_{m2}} & z^{T_{m3}} & \dots & z^{T_{mk}} \end{bmatrix}. \quad (11)$$

The exponent of the element in  $(z)$  is indicated by  $T$ , whose actual meaning is the shifted bits of packets.

### 3. Encoding Design

The encoding framework that one element takes part in the encoding process multiple times when constructing one encoded packet based on Cauchy matrix is mainly constructed in three steps, and the detailed rules are as follows:



	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$	$s_{1,5}$	$s_{1,6}$	$s_{1,7}$	$s_{1,8}$	$s_{1,9}$	$s_{1,10}$	$s_{1,11}$	$s_{1,12}$
$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$	$s_{1,5}$	$s_{1,6}$	$s_{1,7}$	$s_{1,8}$	$s_{1,9}$	$s_{1,10}$	$s_{1,11}$	$s_{1,12}$	

FIGURE 1: The form of elements taking part in encoding process multiple times

*Step 1.* Determine the size of the finite field and the dimension of the Cauchy matrix according to the relation of coding parameters  $(n, k)$ .

According to the relation of  $(n, k)$ , when  $k < n \leq 2^\omega$  and  $\omega$  is a positive integer, we can determine the size of the finite field to be  $GF(2^\omega)$ , where  $\omega = \lceil \log_2 n \rceil$  and the sign  $\lceil \cdot \rceil$  indicates an integer ceiling function. In finite field  $GF(2^\omega)$ , the dimension of Cauchy matrix over  $GF(2^\omega)$  is determined as  $m \times k$  according to the size of  $m$  and  $k$ . The specific construction process of the corresponding Cauchy matrix (i.e., the coding matrix) is to determine the element set of  $X$  and  $Y$  first. The elements of  $X$  could be any  $m$  elements in  $GF(2^\omega)$ ,  $X = \{x_1, x_2, \dots, x_m\}$ , which correspond to  $m$  parity packets. The elements of  $Y$  are any  $k$  elements in the  $2^\omega - m$  elements in  $GF(2^\omega)$  finite field,  $Y = \{y_1, y_2, \dots, y_k\}$ , which correspond to  $k$  parity packets. According to the construction rules of Cauchy matrix in the finite field, the following relationships should be met between the element sets of  $X$  and  $Y$ :

If for  $\forall i \in \{1, 2, \dots, m\}$ ,  $\forall j \in \{1, 2, \dots, k\}$ , the following is satisfied:

- (1)  $x_i + y_j \neq 0$
- (2)  $\forall i, j \in \{1, 2, \dots, m\}$ ,  $i \neq j, x_i \neq x_j$
- (3)  $\forall i, j \in \{1, 2, \dots, k\}$ ,  $i \neq j, y_i \neq y_j$

Then, the elements in  $X$  and  $Y$  are different, and a coding matrix  $G$  with dimension  $m \times k$  can be constructed as follows:

$$G = \begin{bmatrix} \frac{1}{x_1 + y_1} & \frac{1}{x_1 + y_2} & \dots & \frac{1}{x_1 + y_k} \\ \frac{1}{x_2 + y_1} & \frac{1}{x_2 + y_2} & \dots & \frac{1}{x_2 + y_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{x_m + y_1} & \frac{1}{x_m + y_2} & \dots & \frac{1}{x_m + y_k} \end{bmatrix}. \quad (12)$$

*Step 2.* Convert the numeric form of elements in the coding matrix to polynomial form.

Through the arithmetic operation in finite fields and the transformation from field  $GF(2^\omega)$  to field  $GF(2)[z]/q(z)$  in Section 2.3, the matrix  $G$  is transformed into the polynomial form. Each element of  $G$  can be uniquely represented by a polynomial, that is,  $G \Rightarrow G(z)$ .

In particular, the exponent of  $z$  of each polynomial represents the size of the bit-shifting. For example,  $z^2 + z$  represents that a systematic package is shifted to the right by 2

and 1 bits, respectively, and then added over. For example, the shift value of the source packet  $s_1$  of length  $L = 12$  is  $z^2 + z$ , as shown in Figure 1.

*Step 3.* Determine the shift matrix combined with the systematic code.

Combined with the systematic code, a coding matrix

$$A(z) \triangleq \begin{bmatrix} I_k(z) \\ G(z) \end{bmatrix} \quad (13)$$

with dimension  $n \times k$  is obtained through vertical connection.

In order to better understand the design rules that one element takes part in the encoding process multiple times, we will walk through the coding steps with a concrete example.

*Example 2.* If  $(n, k) = (8, 5)$ , then the system has  $k$  systematic packages and  $m$  parity packages, where  $k = 5$  and  $m = n - k = 3$ .

*Step 1.* We have  $k < n \leq 2^\omega \Rightarrow 5 < 8 \leq 2^3$ ,  $\omega = \lceil \log_2 8 \rceil = 3$ , so we can determine that the size of the finite field is  $GF(2^3)$ . For the finite field  $GF(2^3)$ , the commonly used polynomial  $q(z) = z^3 + z + 1$  is chosen as the primitive polynomial. In this case,  $X$  and  $Y$  take 3 and 5 elements, respectively, and the elements of  $X$  and  $Y$  are different. Taking  $X = \{5, 6, 7\}$  and  $Y = \{0, 1, 2, 3, 4\}$  as an example, the coding matrix is as follows:

$$G = \begin{bmatrix} \frac{1}{5+0} & \frac{1}{5+1} & \frac{1}{5+2} & \frac{1}{5+3} & \frac{1}{5+4} \\ \frac{1}{6+0} & \frac{1}{6+1} & \frac{1}{6+2} & \frac{1}{6+3} & \frac{1}{6+4} \\ \frac{1}{7+0} & \frac{1}{7+1} & \frac{1}{7+2} & \frac{1}{7+3} & \frac{1}{7+4} \end{bmatrix}. \quad (14)$$

*Step 2.* Through the arithmetic operations in the finite field, the addition operation is converted to the XOR operation, for example,  $5 \oplus 1 \Rightarrow 101 \oplus 1 = 100 = 4$ . Multiplication and division can be calculated by looking up tables. For example, Table 2 lists the positive and negative tables of  $GF(2^3)$ . Based on the above, we can convert the matrix  $G$  into the following form:

$$G = \begin{bmatrix} 2 & 7 & 4 & 3 & 1 \\ 3 & 4 & 7 & 2 & 5 \\ 4 & 3 & 2 & 7 & 6 \end{bmatrix}. \quad (15)$$

Through the transformation from field  $GF(2^\omega)$  to field  $GF(2)[z]/q(z)$  in Section 2.3, the elements of coding matrix  $G$  are expressed by polynomials one by one to obtain the coding matrix  $G(z)$ :

$$G(z) = \begin{bmatrix} z & z^2 + z + 1 & z^2 & z + 1 & 1 \\ z + 1 & z^2 & z^2 + z + 1 & z & z^2 + 1 \\ z^2 & z + 1 & z & z^2 + z + 1 & z^2 + z \end{bmatrix}. \quad (16)$$

*Step 3.* By vertically concatenating 5 systematic codes, we can obtain the coding matrix  $A(z)$ :

$$A(z) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ z & z^2 + z + 1 & z^2 & z + 1 & 1 \\ z + 1 & z^2 & z^2 + z + 1 & z & z^2 + 1 \\ z^2 & z + 1 & z & z^2 + z + 1 & z^2 + z \end{bmatrix}. \quad (17)$$

## 4. Properties of the Code Based on Cauchy Matrix

*4.1. CP Property.* Before we prove the CP property of the code, we will introduce the properties and lemmas mentioned in the proof.

**Isomorphism property:** the mathematical idea of isomorphism is to establish a one-to-one mapping of two sets that have the same properties associated with operations. For example, assuming that the sets  $B$  and  $\bar{B}$  of algebraic operations are isomorphic, if one set  $B$  has a property that is only relevant to the algebraic operations of this set, then the other set  $\bar{B}$  has exactly similar properties [28].

**Lemma 2.** Any square submatrix of  $G(z)$  is invertible.

*Proof.* According to Section 3, matrix  $G$  is a Cauchy matrix, and the transformation from  $G$  to its polynomial form  $G(z)$  is equivalent to the transformation from field  $GF(2^\omega)$  to field  $GF(2)[z]/q(z)$ , as shown in Section 2.3. In addition, since field  $GF(2^\omega)$  and field  $GF(2)[z]/q(z)$  are isomorphic [17] and the isomorphism property indicates that reversibility in a field will remain reversibility in the isomorphism field, the reversibility of  $G$  in field  $GF(2^\omega)$  will be mapped to that of  $G(z)$  in field  $GF(2)[z]/q(z)$ . Theorem 1 says that any submatrix of  $G$  is invertible, so any submatrix of  $G(z)$  is also invertible.  $\square$

**CP property:** Any  $k$  out of  $n$  encoded packets are able to recover the information of the original  $k$  packets.

*Proof.* First, we use a mathematical model to solve the above coding and decoding problems. The Cauchy matrix  $G$  and the coding matrix  $A$  are constructed from the previous coding design in Section 3, where

$$A \triangleq \begin{bmatrix} I_k \\ G \end{bmatrix}, \quad (18)$$

and  $I_k$  represents the matrix

$$I_k = \begin{bmatrix} 0 & -\infty & \cdots & -\infty \\ -\infty & 0 & \cdots & -\infty \\ \vdots & \vdots & \ddots & \vdots \\ -\infty & -\infty & \cdots & 0 \end{bmatrix} \quad (19)$$

with the dimension of  $k \times k$ . For example,

$$I_3 = \begin{bmatrix} 0 & -\infty & -\infty \\ -\infty & 0 & -\infty \\ -\infty & -\infty & 0 \end{bmatrix}, \quad (20)$$

where element 0 means that the source packet involved in encoding shifts 0 bit, which means nonshift, and element  $-\infty$  means that the source packet does not participate in encoding.

According to the mathematical model of Section 2.4, in the finite field  $GF(2^\omega)$ , the matrix above is represented by polynomial form of  $z$ , where  $z$  is the radix (assuming that the modulus of  $z$  is greater than 1), and every element of matrix  $A$  is raised to a proper power and then mod the original polynomial of the finite field. This process is actually a transformation from field  $GF(2^\omega)$  to field  $GF(2)[z]/q(z)$ . For example, the identity matrix  $I_k$  to  $I_k(z)$  applies the above transformation to

$$I_k = \begin{bmatrix} 0 & -\infty & \cdots & -\infty \\ -\infty & 0 & \cdots & -\infty \\ \vdots & \vdots & \ddots & \vdots \\ -\infty & -\infty & \cdots & 0 \end{bmatrix} \Rightarrow I_k(z) \quad (21)$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

where element 1 indicates that the source packet participating in the encoding is not shifted, and element 0 indicates that the source packet is not participating in the encoding. The transformation process of matrix  $G$  to  $G(z)$  is shown

in Section 2.3. At this point, the polynomial form of the coding matrix  $A$  is

$$A(z) \triangleq \begin{bmatrix} I_k(z) \\ G(z) \end{bmatrix}, \quad (22)$$

which is a matrix with a size of  $n \times k$ .

Based on the polynomial form of  $z$  above and the mathematical model in Section 2.4, the polynomial  $c(z)$  represents the encoded packet and the polynomial  $s(z)$  represents the systematic packet. The coding relationship can be expressed as  $c(z) = A(z)s(z)$ . Take any  $k$  packets from  $n$  coded packets, that is, extract  $k$  lines at the same position from encoded packet  $c(z)$  and form  $c_k(z)$ , corresponding to take  $k$  lines at the same position from  $A(z)$  and form  $k \times k$  matrix  $A_k(z)$ . The encoding relationship can be expressed by the expression  $c_k(z) = A_k(z)s(z)$ .

Based on the above model, satisfying CP property is equivalent to the invertibility of  $A_k(z)$ . We claim that  $A_k(z)$  is invertible due to the following reasons:

In

$$A(z) \triangleq \begin{bmatrix} I_k(z) \\ G(z) \end{bmatrix}, \quad (23)$$

take any  $k$  rows from  $A(z)$  and form  $k \times k$  matrix  $A_k(z)$ . In this case,  $A_k(z)$  to be decoded can be formed in two ways: First,  $A_k(z)$  does not contain any rows of  $I_k(z)$ ; from Lemma 2 above, we know that every square submatrix of  $G(z)$  is invertible, so  $A_k(z)$  is invertible. Second,  $A_k(z)$  is composed of  $\alpha$  rows in  $I_k(z)$  and  $\beta$  rows in  $G(z)$ , where  $I_\alpha(z)$  represents  $\alpha$  rows in  $I_k(z)$  and  $G_\beta(z)$  represents  $\beta$  rows in  $G(z)$ , and  $\alpha + \beta = k$ . Since the matrix  $I_\alpha(z)$  is a known systematic packet, substituting it into  $G_\beta(z)$  is equivalent to deleting  $\alpha$  columns from it. The deleted  $G_\beta(z)$  is equivalent to extracting the  $\beta \times \beta$  submatrix from  $G(z)$ , where  $k - \alpha = \beta$ , as shown in the following Example 3. Similarly, Lemma 2 shows that any square submatrix of  $G(z)$  is invertible, so  $A_k(z)$  is also invertible.

To sum up,  $A(z)$  is invertible, so this coding framework can meet the CP property.  $\square$

*Example 3.* Following Example 2, if  $(n, k) = (8, 5)$ , the system has 5 systematic packets,  $m$  parity packets, where  $m = n - k = 3$ . The matrix  $G(z)$  has been constructed as

$$G(z) = \begin{bmatrix} z & z^2 + z + 1 & z^2 & z + 1 & 1 \\ z + 1 & z^2 & z^2 + z + 1 & z & z^2 + 1 \\ z^2 & z + 1 & z & z^2 + z + 1 & z^2 + z \end{bmatrix}. \quad (24)$$

TABLE 4: Success/failure for different  $(n, k)$  parameters.

$(n, k)$	Success	Failure	Total
(5, 3)	8	2	10
(8, 5)	46	9	56
(12, 8)	405	90	495
(15, 10)	2408	595	3003

Combining the identity matrix  $I_5(z)$ ,

$$A(z) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ z & z^2 + z + 1 & z^2 & z + 1 & 1 \\ z + 1 & z^2 & z^2 + z + 1 & z & z^2 + 1 \\ z^2 & z + 1 & z & z^2 + z + 1 & z^2 + z \end{bmatrix}. \quad (25)$$

$(z)$  is composed of some parts of  $I_5(z)$  and  $G(z)$ , let

$$A_5(z) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ z & z^2 + z + 1 & z^2 & z + 1 & 1 \\ z + 1 & z^2 & z^2 + z + 1 & z & z^2 + 1 \end{bmatrix}, \quad (26)$$

and then, it can be seen from the matrix  $A_5(z)$  that the known systematic packets are  $s_1, s_2, s_3$ , so  $A_5(z)$  can be simplified as

$$A_5(z)' = \begin{bmatrix} z + 1 & 1 \\ z & z^2 + 1 \end{bmatrix}, \quad (27)$$

which is equivalent to the  $2 \times 2$  submatrix of  $G(z)$ . According to Lemma 2, any submatrix of  $G(z)$  is invertible, so  $A_5(z)'$  is invertible.

**4.2. Zigzag Decodability (ZD) Property.** The ZD property of one element which takes part in the encoding process multiple times when constructing one encoded packet can be proved by experiment that it cannot be fully zigzag decodable. Starting from the experiment, we set several groups of parameters to obtain the probability of zigzag decodability of one element which takes part in the encoding process

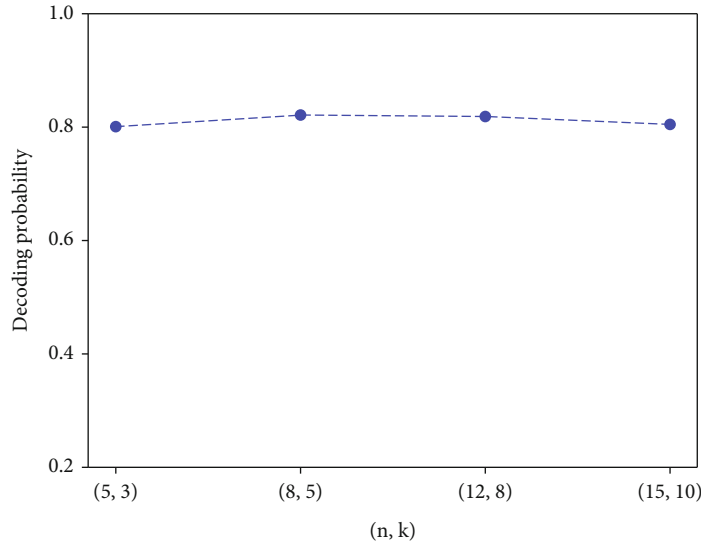


FIGURE 2: Decoding probability for different  $(n, k)$  parameters.

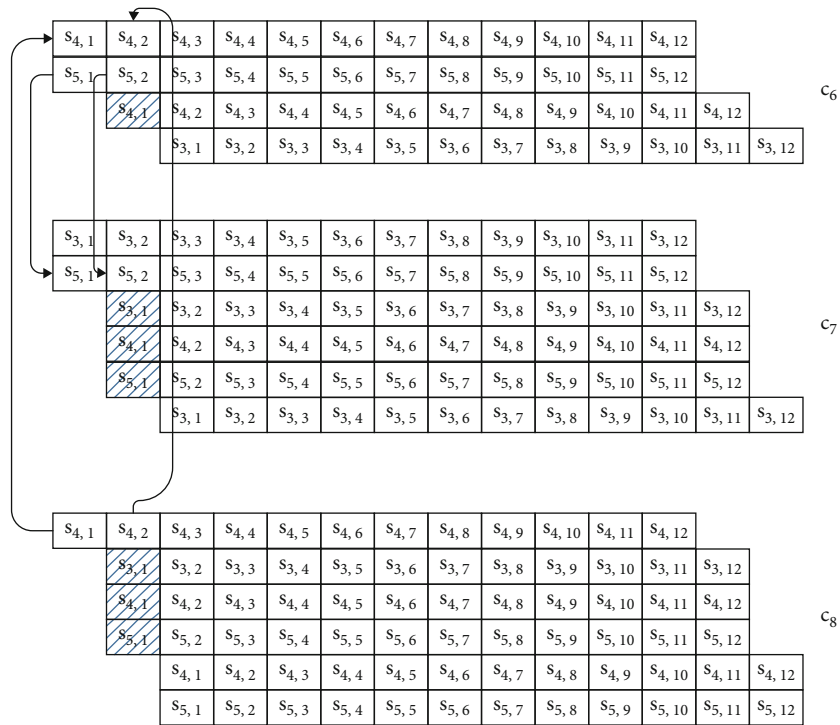


FIGURE 3: An example of able to perform ZD.

multiple times. Through experimental simulation, we set several sets of parameters  $(n, k)$  as  $(5, 3)$ ,  $(8, 5)$ ,  $(12, 8)$ , and  $(15, 10)$ , respectively, and the conditions of zigzag decodable and not are shown in Table 4.

It can be seen from Figure 2 that the encoding framework that one element takes part in the encoding process multiple times when constructing one encoded packet based on the Cauchy matrix has about 80% probability of ZD decoding, which means that there is about 20% probability that zigzag decoding will not be possible. The following

two examples illustrate the case that the encoding framework is able or unable to perform zigzag decoding.

*Example 4.* Following Example 2 where  $(n, k) = (8, 5)$ , we assume that there are encoded packets  $c_1, c_2, c_6, c_7, c_8$ , where  $c_1 = s_1$  and  $c_2 = s_2$ . Based on the above, the remaining problem is to use the encoded packets  $c_6, c_7, c_8$  to decode the three source packets  $s_3, s_4, s_5$ . The corresponding coding matrix is as follows:

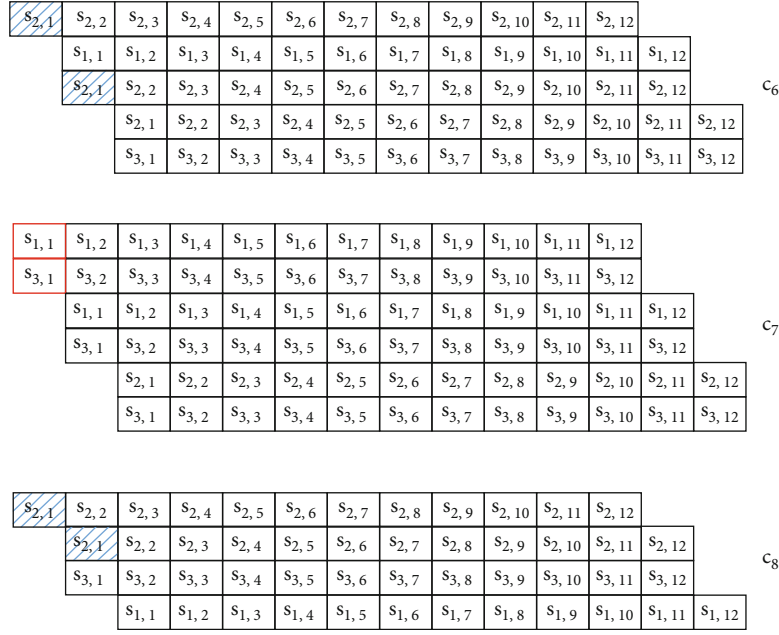


FIGURE 4: An example of failure in ZD.

$$G = \begin{bmatrix} z^2 & z+1 & 1 \\ z^2+z+1 & z & z^2+1 \\ z & z^2+z+1 & z^2+z \end{bmatrix}. \quad (28)$$

Assuming that the length of the encoded packets is  $L = 12$ , we can obtain the decoding diagram of Figure 3 through shift-and-addition encoding by row. In this condition, the coding framework can be zigzagged:  $s_{4,1}$  can be obtained directly from the first exposed bit of  $c_8$ . Substitute  $s_{4,1}$  into the first bit of  $c_6$  to obtain  $s_{5,1}$  through shift-and-addition, and then, substitute  $s_{5,1}$  into  $c_7$  to obtain  $s_{3,1}$ . In this manner, the first bit of each source packet has been obtained. The decoding of the second bit of each source packet is similar to the decoding of the first bit. Substitute  $s_{3,1}$ ,  $s_{4,1}$ ,  $s_{5,1}$  into  $c_6$ ,  $c_7$ ,  $c_8$ , respectively, so that  $s_{4,2}$  can be obtained. Substitute  $s_{4,2}$  into the second bit of  $c_6$ , and  $s_{5,2}$  can be obtained by XOR operation. Finally, substitute  $s_{5,2}$  into  $c_7$ , and then,  $s_{3,2}$  can be obtained.

*Example 5.* Following Example 2 where  $(n, k) = (8, 5)$ , we assume that there are encoded packets  $c_4, c_5, c_6, c_7, c_8$ , where  $c_4 = s_4$  and  $c_5 = s_5$ . Therefore, the remaining problem is to use the encoded packets  $c_6, c_7, c_8$  to decode the three source packets  $s_1, s_2, s_3$  to form the encoding matrix,

$$G = \begin{bmatrix} z & z^2+z+1 & z^2 \\ z+1 & z^2 & z^2+z+1 \\ z^2 & z+1 & z \end{bmatrix}. \quad (29)$$

Assuming the length of the encoded packets is  $L = 12$ , through shift-and-addition encoding by row, we cannot per-

form zigzag decoding in this case. From the first bit of each coded packet, only  $s_{2,1}$  is exposed bit. When it is substituted into the second bit of  $c_6, c_8$ , respectively, no new exposed bit can be obtained. The whole decoding process is locked, so the zigzag decoding cannot proceed, as shown in Figure 4.

## 5. Performance Analysis

In the case of one element takes part in the encoding only once in shift-and-addition encoding, the overhead of several existing CP-ZD codes is the square of  $k$  or  $n$ . As can be seen from Section 3, the overhead of the encoding framework that one element taking part in the encoding process multiple times when constructing one encoded packet is related to the size of the finite field; if the size of the finite field is  $GF(2^\omega)$ , then  $\omega = \lceil \log_2 n \rceil$ . Therefore, the maximum overhead of this encoding framework is determined by the number  $n$  of encoded packets, and the overhead is  $\text{OH}_{\text{cauchy}} = \lceil \log_2 n \rceil - 1$ . Therefore, the overhead OH has a logarithmic relationship with the number  $n$  of encoded packets, which has a huge advantage over the existing zigzag codes. However, due to the existence of multiple encodings, a source packet may be encoded once or more, which is more complex than the case of single encodings, leading to the possibility of decoding failure during zigzag decoding, as shown in Example 5. As can be seen from Figure 2, the ZD decoding rate of the encoding framework that one element takes part in the encoding process multiple times when constructing one encoded packet based on Cauchy matrix in shift-and-addition is about 80%.

In general, compared with the encoding framework of one element taking part in the encoding only once in shift-and-addition, the encoding framework of elements' multi-participation based on the Cauchy matrix has a good

constraint on the overhead and can reduce the overhead from the existing square level to the logarithmic level. However, it has some losses in ZD properties and cannot guarantee ZD decoding.

## 6. Conclusions

Aiming at the problem of high overhead in CP-ZD codes, in this paper, we design a coding framework based on the idea of elements taking part in encoding multiple times in constructing an encoded packet based on Cauchy matrix and shift-and-addition. It is proved here that the framework has CP properties, but not completely with ZD properties. Experimental results show that the ZD decoding rate of this code is about 80%. However, the overhead is  $OH_{\text{cauchy}} = \lceil \log_2 n \rceil - 1$ , which is reduced from the existing square level to the logarithmic level. The coding framework confirms the advantage of the element participating in the encoding for multiple times and lays a foundation for future research.

## 7. Future Works

The new coding framework proposed in this paper satisfies the CP property, but not the ZD property. At present, there is no mature encoding framework with one element taking part in encoding process multiple times. Aiming at the idea of elements taking part in encoding multiple times, it is obviously of prospective academic and application value to design a feasible ZD decoding framework. What is more, it is also necessary to study the closed form expression of CP-ZD code, which can be used to describe the necessary and sufficient conditions of CP-ZD code that elements taking part in encoding process multiple times.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the National Key Research and Development Program under Grant 2019YFB1803305, the Natural Science Foundation of China (62071304), the Natural Science Foundation of Guangdong Province (2020A1515010381), the Basic Research Foundation of Shenzhen City (20200826152915001), the Guangdong Basic and Applied Basic Research Foundation (2022A1515011219), the Natural Science Foundation of Shenzhen City (JCYJ20190808120415286), and the Natural Science Foundation of Shenzhen University (00002501).

## References

- [1] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," *Mobile networks & Applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [2] J. Dean and L. A. Barroso, "The tail at scale," *Communications of the ACM*, vol. 56, no. 2, pp. 74–80, 2013.
- [3] P. Soto and J. Li, "Straggler-free coding for concurrent matrix multiplications," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 3017–3021, Los Angeles, CA, USA, June 2020.
- [4] K. V. Rashmi, N. B. Shah, D. Gu, H. Kuang, D. Borthakur, and K. Ramchandran, "A solution to the network challenges of data recovery in erasure-coded distributed storage systems: a study on the Facebook warehouse cluster," in *5th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 13)*, San Jose, CA, June 2013.
- [5] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos et al., "XOR-ing elephants," *Proceedings of the VLDB Endowment*, vol. 6, no. 5, pp. 325–336, 2013.
- [6] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," *IEEE Journal on Selected Areas in Communications*, vol. 99, no. 3, pp. 476–489, 2011.
- [7] D. S. Papailiopoulos, J. Luo, A. G. Dimakis, C. Huang, and J. Li, "Simple regenerating codes: network coding for cloud storage," in *Proceedings - IEEE INFOCOM*, pp. 2801–2805, Orlando, FL, USA, March 2011.
- [8] V. R. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and C. Suh, "Asymptotic interference alignment for optimal repair of MDS codes in distributed storage," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2974–2987, 2013.
- [9] P. K. Akulakrishna, J. Lakshmi, and S. Nandy, "Efficient storage of big-data for real-time GPS applications," in *IEEE Fourth International Conference on Big Data & Cloud Computing*, pp. 1–8, Sydney, NSW, Australia, December 2014.
- [10] Y. Aikebaier, T. E. Yang, and M. Takizawa, "Energy-efficient computation models for distributed systems," in *International Conference on Network-Based Information Systems*, pp. 423–431, Indianapolis, IN, USA, August 2009.
- [11] S. Li, M. A. Maddah-Ali, and A. S. Avestimehr, "Coding for distributed fog computing," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 34–40, 2017.
- [12] A. Severinson, A. G. I Amat, and E. Rosnes, "Block-diagonal coding for distributed computing with straggling servers," in *2017 IEEE Information Theory Workshop (ITW)*, pp. 464–468, Kaohsiung, Taiwan, November 2017.
- [13] A. B. Das, L. Tang, and A. Ramamoorthy, "C3LES: codes for coded computation that leverage stragglers," in *2018 IEEE Information Theory Workshop (ITW)*, pp. 1–5, Guangzhou, China, November 2018.
- [14] T. Baharav, K. Lee, O. Ocal, and K. Ramchandran, "Straggler-proofing massive-scale distributed matrix multiplication with D-dimensional product codes," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1993–1997, Vail, CO, USA, June 2018.
- [15] M. Dai, Z. Zheng, S. Zhang, H. Wang, and X. Lin, "SAZD: a low computational load coded distributed computing framework for IoT systems," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3640–3649, 2020.
- [16] L. Chen, R. Zhao, K. He, Z. Zhao, and L. Fan, "Intelligent ubiquitous computing for future UAV-enabled MEC network systems," *Cluster Computing*, vol. 25, no. 4, pp. 1–11, 2022.
- [17] M. Dai, A. Xu, Q. Huang, Z. Zhang, and X. Lin, "Vertical federated DNN training," *Physical Communication*, vol. 49, p. 101465, 2021.

- [18] J. Li, S. Dang, Y. Yan, Y. Peng, S. al-Rubaye, and A. Tsourdos, "Generalized quadrature spatial modulation and its application to vehicular networks with NOMA," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4030–4039, 2021.
- [19] X. Pei, Y. Chen, M. Wen, H. Yu, E. Panayirci, and H. V. Poor, "Next-generation multiple access based on NOMA with power level modulation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1072–1083, 2022.
- [20] M. Wen, Q. Li, K. J. Kim et al., "Private 5G networks: concepts, architectures, and research landscape," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 7–25, 2022.
- [21] S. Gollakota and D. Katabi, "Zigzag decoding: combating hidden terminals in wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 159–170, 2008.
- [22] L. Zhang, W. Zhou, J. Xia et al., "DQN-based mobile edge computing for smart internet of vehicle," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, 16 pages, 2022.
- [23] J. Lu, L. Chen, J. Xia et al., "Analytical offloading design for mobile edge computing-based smart internet of vehicle," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, 19 pages, 2022.
- [24] H. Hou and Y. S. Han, "Cauchy MDS array codes with efficient decoding method," 2016, <http://arxiv.org/abs/1611.09968>.
- [25] J. S. Plank, "A tutorial on Reed–Solomon coding for fault-tolerance in raid-like systems," *Software: Practice and Experience*, vol. 27, no. 9, pp. 995–1012, 1997.
- [26] J. Bloemer, M. Kalfane, R. Karp, M. Karpinski, M. Luby, and D. Zuckerman, *An XOR-based erasure-resilient coding scheme*, Technical Report at ICSI, 1999.
- [27] M. Dai, C. W. Sung, H. Wang, X. Gong, and Z. Lu, "A new zigzag-decodable code with efficient repair in wireless distributed storage," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1218–1230, 2017.
- [28] N. Jacobson, *Basic Algebra I*, Freeman & Co, 1985.

## Research Article

# Multiband RoF System with Good Transmission Performance Based on Additional SCS

Anliang Liu  and Hongzhi Li 

*School of Information Science and Technology, Dalian Maritime University, Dalian, China*

Correspondence should be addressed to Anliang Liu; [alliu@dmlu.edu.cn](mailto:alliu@dmlu.edu.cn)

Received 16 June 2022; Accepted 8 August 2022; Published 22 August 2022

Academic Editor: Jun Li

Copyright © 2022 Anliang Liu and Hongzhi Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A wavelength division multiplexing radio-over-fiber (WDM-RoF) system that can support multiband radio signals based on an additional subcentral station (SCS) is proposed and demonstrated in this paper. The dispersion characteristics of the multiband radio signals over fiber transmission are explored, and we exploit an improved transmission structure consisting of the central station (CS), the SCS, and base stations (BSs) to optimize the antidispersion performance of the proposed RoF system. At the SCS, a dual-drive Mach-Zehnder modulator (MZM) is employed to realize a cost-effective multiband radio signal generation of WDM signals with great spectral flatness, and an optical cross-connect unit is utilized to achieve a flexible distribution of different band radio signals for remote BSs. Based on the additional SCS, we establish a verification WDM-RoF system, which can support reliable transmission of multiband radio signals including 13 GHz, 26 GHz, 39 GHz, and 52 GHz. The WDM signals with each data rate at 2.5 Gbps are successfully transmitted over a 40 km standard single-mode fiber.

## 1. Introduction

Affected by the COVID-19, network office and online education have become the main solutions for people's daily work and learning. The demand of mobile users for high-speed, large-capacity, and reliable communication networks has suddenly increased, and it is in line with the enhanced mobile broadband (eMBB) application scenario in the fifth-generation mobile communication technology (5G) communication systems, which means to provide users with the ultimate high-speed communication experience [1, 2]. Under the current limited spectrum resources, the 3GPP organization for the first time enabled a high-band frequency ranging from 24.25 GHz to 52.6 GHz to achieve a higher data rate transmission in the 5G system [3]. Various countries have carried out a series of studies on the high-band communication systems according to the distribution of their own spectrum resources, as shown in Table 1 specifically [4, 5]. The high-band communication systems are mainly concentrated in several radiofre-

quency (RF) bands such as 26 GHz, 28 GHz, and 39 GHz. With the increasing demand for data rate, more and more high-band carriers will be developed and utilized in 5G and even future sixth-generation mobile communication technology (6G) communication systems [6, 7]. Therefore, the collaborative communication technology between multiband radio signals is of great significance for large-capacity communication and effective signal coverage, which means that the research on reliable transmission and efficient generation of multiband radio signals is very essential for future mobile communication systems.

The transmission distance of high-band radio carriers is mainly limited by air attenuation, and radio-over-fiber (RoF) technology has become a preferred solution for long-distance and reliable transmission in 5G high-band communication systems [8]. In RoF systems, the generation of high-band radio carriers can be realized in the optical domain, which can effectively simplify the system structure and save implementation costs. The current implementation schemes are mainly based



TABLE 1: Deployments of 5G high bands.

Country	Frequency (GHz)
China	24.25~27.5; 37~43.5
USA	27.5 ~ 28.35; 37~40
Europe	24.25~27.5
Korea	26.5 ~ 29.5
Japan	27.5 ~ 28.28

on optical heterodyne, optical nonlinearity, and external modulation methods [9–13]. Among them, the external modulation technology based on the Mach-Zehnder modulator (MZM) has been widely investigated with its great flexibility, high bandwidth, and linearity [14]. Traditional optical carrier suppression modulation can be used to achieve a frequency-doubled electrical signal generation [15]. In reference [16], a W-band millimeter-wave (MMW) signal is generated by a frequency quadrupling technology with one single MZM and a wavelength selective switch (WSS). A 24 GHz microwave signal is obtained by a frequency octupling method based on a dual-polarization quadrature phase shift keying modulator [17]. Moreover, based on a 16-tupling frequency MMW generation scheme without filters, a 160 GHz electrical signal is obtained with two MZMs [18]. Other architectures with more MZMs are proposed to realize a high-band MMW signal generation [19–21]. However, with the increasing number of modulators, the system structure becomes more complex. However, the abovementioned structures mostly carried out the generation of a single RF carrier through a low-frequency local oscillator (LO). As far as we know, few studies focus on how to generate multiple RF carriers at different bands efficiently. In addition, compared with the traditional RoF systems, it becomes more difficult to overcome the fiber dispersion for a high-reliability transmission as the number of high-band signals increases. Therefore, the efficient generation, flexible allocation, and reliable transmission of multiband radio signals in the RoF system is essential for its application in future high-band communication systems.

In this paper, a flexible RoF system based on an additional subcentral station (SCS) that can support multiband radio signals for cooperative communication with great dispersion tolerance is proposed and demonstrated. The different numbers of high-quality optical sidebands is generated by setting appropriate parameters of a dual-arm MZM modulator. We build a verification system with one light source and analyze the influence of the fiber dispersion on the reliability of data transmission at different RF carriers. To achieve the generation and reliable transmission of a large number of multiband radio signals, we establish a wavelength division multiplexing (WDM) RoF system based on the SCS structure, which can transmit several RF signals including 13 GHz, 26 GHz, 39 GHz, and 52 GHz. A centralized multiband RF signal generation structure and an optical cross-connect (OXC) unit are utilized to realize cost-effective generation and flexible allocation of different RF signals. A reliable transmission of 3-channel WDM signals with each data rate at 2.5 Gbps over 40 km optical fiber has been demonstrated by the measured eye diagrams and bit error rate (BER) curves.

## 2. Principle

The principle of optical multiband MMWs generation based on the external MZM is to create new optical harmonics by setting a couple of appropriate parameters of the modulator. Generally, new order sidebands are generated at the central station (CS) with a low frequency LO. At the base stations (BSs), the modulated optical signal can be converted to electrical RF signals at different frequencies after the beat function of the photodetector (PD). The structure diagram of a dual-arm LiNbO<sub>3</sub>-MZM is shown in Figure 1.

The input optical carrier can be expressed as

$$E_{\text{in}}(t) = E_c e^{j\omega_c t}, \quad (1)$$

where  $E_c$  is the amplitude and  $\omega_c$  is the angular frequency of the input optical signal. The input signal is equally divided into the upper and lower channels through a Y-shaped waveguide. The phase of optical signals is controlled by the input RF signals and the DC bias voltages in each arm. The recombined signal at the output port can be expressed as

$$E_{\text{out}}(t) = \frac{1}{2} E_{\text{in}} \left[ e^{j(\pi/V_\pi)[V_1(t)+V_{\text{bias1}}]} + e^{j(\pi/V_\pi)[V_2(t)+V_{\text{bias2}}]} \right], \quad (2)$$

where  $V_\pi$  is the half-wave voltage of the MZM.  $V_1(t)$  and  $V_2(t)$  are the RF signals loaded on the upper and lower arms, respectively.  $V_{\text{bias1}}$  and  $V_{\text{bias2}}$  are the DC bias voltages severally. Normally, the MZM is set to work in push-pull mode, which means  $V_1(t) = -V_2(t)$ , and they can be described as

$$\begin{aligned} V_1(t) &= V_{\text{RF}} \cos(\omega_{\text{RF}} t), \\ V_2(t) &= V_{\text{RF}} \cos(\omega_{\text{RF}} t + \Delta\theta), \end{aligned} \quad (3)$$

where  $V_{\text{RF}}$  and  $\omega_{\text{RF}}$  are the amplitude and angular frequency term of the RF signal, respectively. Where  $\Delta\theta = (2k+1)\pi$ , it is the phase difference between the two modulated RF signals. Then, equation (2) can be simplified by the Bessel function expansion as

$$E_{\text{out}}(t) = \frac{1}{2} E_{\text{in}} e^{j(V_a/2)} \left\{ \sum_{n=-\infty}^{+\infty} J_n(m) e^{jn(\omega_{\text{RF}} t + (\Delta\theta + \pi)/2)} \cos\left(\frac{\Delta\varphi}{2} + n\frac{\Delta\theta}{2}\right) \right\}, \quad (4)$$

where  $m = \pi V_{\text{RF}}/V_\pi$  is the modulation index,  $V_a = \pi(V_{\text{bias1}} + V_{\text{bias2}})/V_\pi$ ,  $\Delta\varphi = \pi(V_{\text{bias1}} - V_{\text{bias2}})/V_\pi$ , and  $J_n(\bullet)$  is the first kind Bessel function of  $n$ th order. It can be seen that the intensity of different order optical sidebands can be controlled by the modulation depth  $m$ , the phase difference of the modulation signal, and the magnitude of the bias voltage. For example, when  $m = 1.84$ ,  $\Delta\varphi = 1.76V_\pi$ , and  $\Delta\theta = \pi$ , the amplitudes of different order sidebands  $|E_0| \approx |E_{\pm 1}| \approx |E_{\pm 2}|$ , which means a five-line comb with a great flatness can be acquired. By detecting the generated optical comb lines, the electrical RF signals at different bands can be obtained.

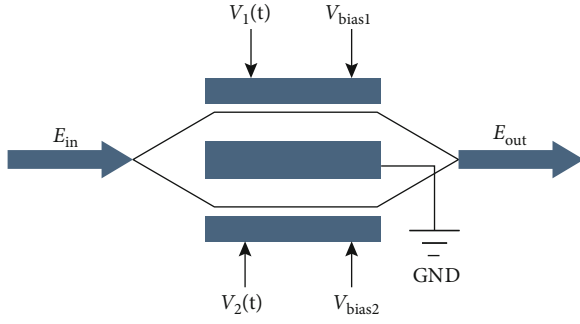


FIGURE 1: Typical structure of LiNbO<sub>3</sub>-MZM.

Therefore, with the above structure, only one LO has been utilized to achieve a centralized generation of multiple RF signals, which can effectively reduce the cost of multiband radio signal generation.

### 3. Simulation Results and Discussion

**3.1. Multiband RoF System Based on CS-BS Structure.** According to the principle of multiband radio signal generation with the MZM modulator mentioned in the previous section, we build a multiband RoF verification system based on the traditional CS-BS structure to analyze its transmission characteristics. The system architecture is shown in Figure 2.

At the CS, the laser diode (LD) with a wavelength of 1550 nm is injected to a dual-arm MZM after the polarization controller (PC). The optical carrier is modulated by a RF signal that consists of a 13 GHz LO and a 2.5 Gbps baseband data signal. The half-wave voltage  $V_\pi$  of the modulator is 4 V. Under the push-pull operation mode, to acquire a five optical comb-line, the bias voltage of MZM is set to  $V_{\text{bias}} = V_\pi = 4$  V and the amplitude of the mixed RF signal is set to 2.34 V to set the modulation index  $m$  to 1.84. The optical spectrum of the MZM output optical is shown in Figure 2(a). The frequency of the central optical carrier is 193.1 THz, and the comb line spacing is 13 GHz, which has a great flatness of 0.35 dB, and an unwanted-mode suppression ratio of 15 dB. The modulated signal is amplified by an erbium-doped fiber amplifier (EDFA) with a gain of 20 dB and then transmitted to the receiver over a standard single-mode fiber (SSMF).

At the BS, a PD is used to convert the optical signal to multiband radio signals. The spectrum of the detected electrical signal is shown in Figure 2(b). It can be seen that multiple signals at different frequencies including 13 GHz, 26 GHz, 39 GHz, and 52 GHz are achieved. And they can be selected and transmitted to mobile users by the RF antenna. In order to analyze the transmission characteristics of these RF signals over the fiber system, a tunable band-pass filter (BPF) and different LOs (including 13 GHz, 26 GHz, 39 GHz, and 52 GHz) are employed to downconvert each received RF signal. The original baseband data signal can be recovered from different frequency carriers after a low-pass filter (LPF), and the eye diagrams can be measured by an oscilloscope. By changing the fiber length, the bit error

rate curves for different RF signals are measured and as shown in Figure 3. Due to the influence of the fiber dispersion, the reliability of data signals with lower-frequency carriers is better than the data with higher-frequency carriers. Moreover, the original baseband data signal is modulated to each new generated optical carrier, the periodic fading effect has a much more serious impact on the multiband RoF system.

The power of the received RF signals at 13 GHz, 26 GHz, and 39 GHz is influenced by the transmission length. As the frequencies of RF signals increase, the power fading points appear more frequently. For example, when the transmission distance is 12.5 km, the data transmission performance on the 13 GHz carrier frequency decreases significantly, while the data transmission quality on other frequency carriers is better. Particularly, the 52 GHz RF carrier signal can be immune to the power fading effect because it comes from the  $\pm 2\text{nd}$ -order sidebands. With the increasing number of different frequency carriers, it becomes more difficult to select a suitable optical fiber transmission distance to avoid the power fading points and achieve reliable signals transmission. In order to suppress the influence of chromatic dispersion on system reliability, the amount of frequency spectrum entering the beat frequency of the photodetector should be decreased.

**3.2. Multiband WDM-RoF System Based on Additional SCS.** A multiband WDM-RoF system based on the CS-SCS-BS structure is proposed to improve the transmission performance of different RF carriers. The proposed system structure is shown in Figure 4.

Baseband data signals are utilized between the CS and SCS to alleviate the serious impact of fiber dispersion on link reliability. Multichannel downstream signals are transmitted to the remote SCS through a single optical fiber by wavelength division multiplexing and then centrally modulated by the MZM-based multiband RF signals generation structure mentioned above. The modulated optical signals consist of multiorder spectral components. An additional OXC unit is employed to realize a flexible separation of the downstream signals at different frequencies for their respective BSs. Since the SCS is close to the BS sides, the dispersion effect can be ignored. Thus, the additional SCS can greatly improve the dispersion resistance and flexibility of the proposed multiband WDM-RoF system.

A verification structure of the multiband WDM-RoF system is shown in Figure 5. At the CS,  $3 \times 2.5$  Gbps baseband data are modulated to the optical TXs (transmitters) with different central frequencies of  $f_1 = 1552.52$  nm,  $f_2 = 1551.72$  nm, and  $f_3 = 1550.92$  nm, respectively. The modulated optical signals are multiplexed by a WDM multiplexer and then amplified by an EDFA. The fiber length between the CS and SCS is set to 40 km. The downstream signals are centrally modulated by a 13 GHz LO to realize 5 optical lines generation for each optical carrier. The output optical spectra are shown in Figure 5(a). The flatness of the new generated optical lines is less than 0.5 dB, and the unwanted-mode suppression ratio is more than 15 dB for each optical carrier.

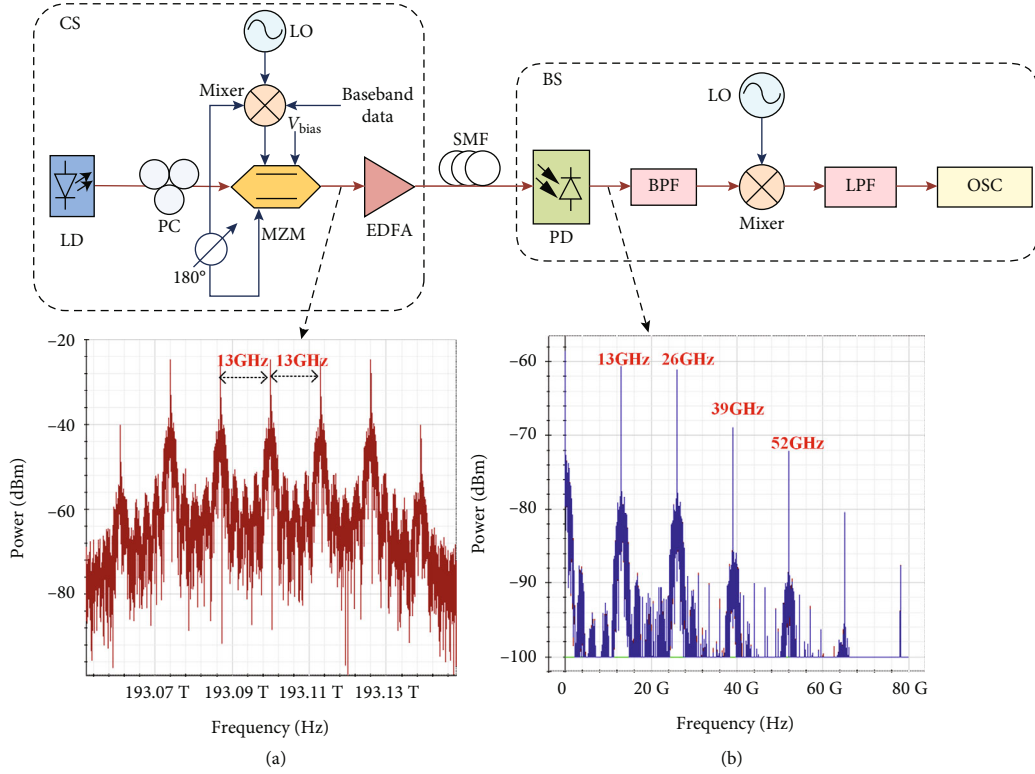


FIGURE 2: Schematic setup of multiband RoF system. (a) Output optical spectra after MZM. (b) Detected RF signal spectra after PD.

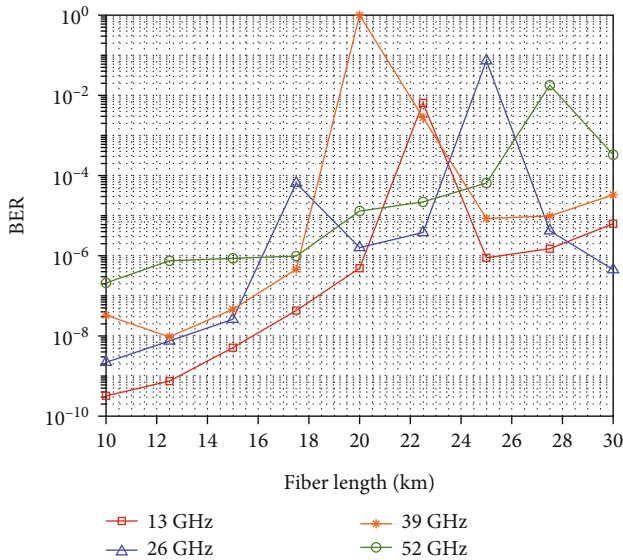


FIGURE 3: BER curves versus fiber lengths for different RF signals.

An OXC unit consisting of a WSS with a minimum slice of 12.5GHz and two optical couplers (OCs) is employed to implement an allocation for different RF signals. For example, if a 26 GHz RF signal needs to be transmitted to the BS1 through the optical carrier at  $f_1$ , the WSS slice number of output port 1 should be set to 141~142 as a band pass

filter which central frequency is 193.0875 THz and the bandwidth is from 193.075 THz to 193.1 THz and the slice number of output port 2 is set to 143~144 which central frequency is 193.0875 THz and the bandwidth is from 193.1 THz to 193.125 THz. The new generated  $\pm 1$ st-order sidebands will be filtered out and then recombined by an OC. The output spectra after WSS is shown in Figure 5(b). At the BS1, a PD is used to convert the optical signal to the 26 GHz RF signal. In order to analyze the reliability of system transmission, a LO is employed to recover the original baseband data and measured by a BERT (bit error rate tester).

By changing the parameters of the WSS at the SCS, several specified RF signals can be sent to BS1. The measured BER curves for different RF signals are shown in Figure 6. It can be found that the BER of 13 GHz, 26 GHz, 39 GHz, and 52 GHz signals can reach  $3.8 \times 10^{-3}$  when the received signal power is about  $-25$  dBm. In particular, the BER of 52 GHz signal can be as low as  $10^{-6}$  due to the width of the WSS slice and the data rate of the baseband signal. The RF signals over different optical carriers can also be distributed to each remote BS. The BER curves of 26 GHz RF signal over different wavelengths are shown in Figure 7. It can be seen that the optical receiver sensitivity is approximately  $-25$  dBm with the BER  $< 3.8 \times 10^{-3}$ . On the basis of the proposed multiband RF signals generator and the optical wavelength switch unit at the SCS, more RF signals at different frequencies can be obtained and delivered to the required BSs. Furthermore, the multiband radio signal generation structure

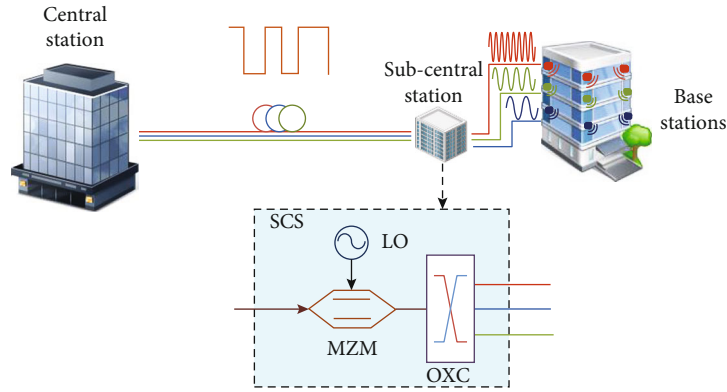


FIGURE 4: Architecture of the proposed multiband WDM-RoF system.

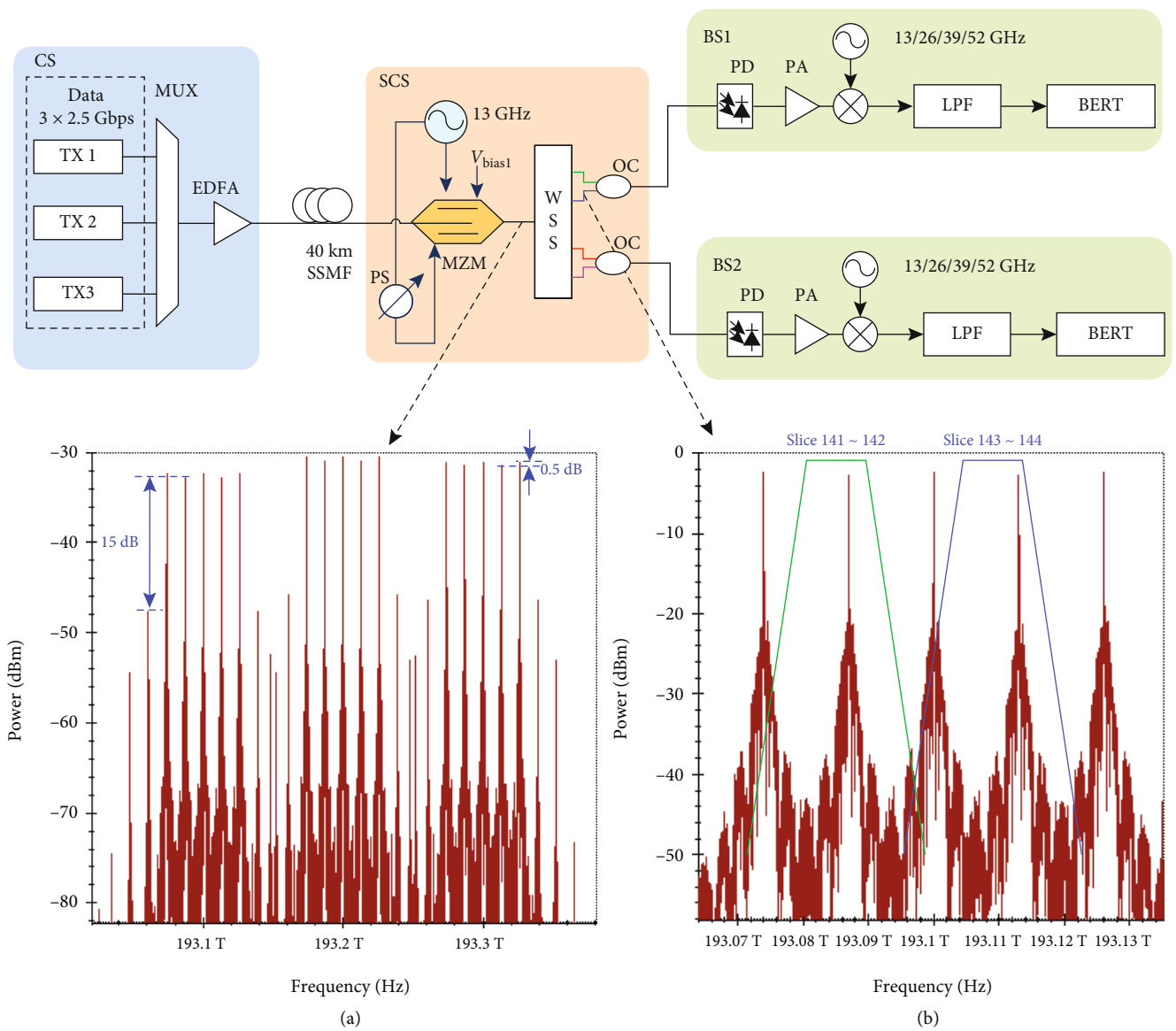


FIGURE 5: Mutiband WDM-RoF verification system based on the additional SCS. (a) Output optical spectra of the MZM. (b) Spectra after WSS at the SCS.

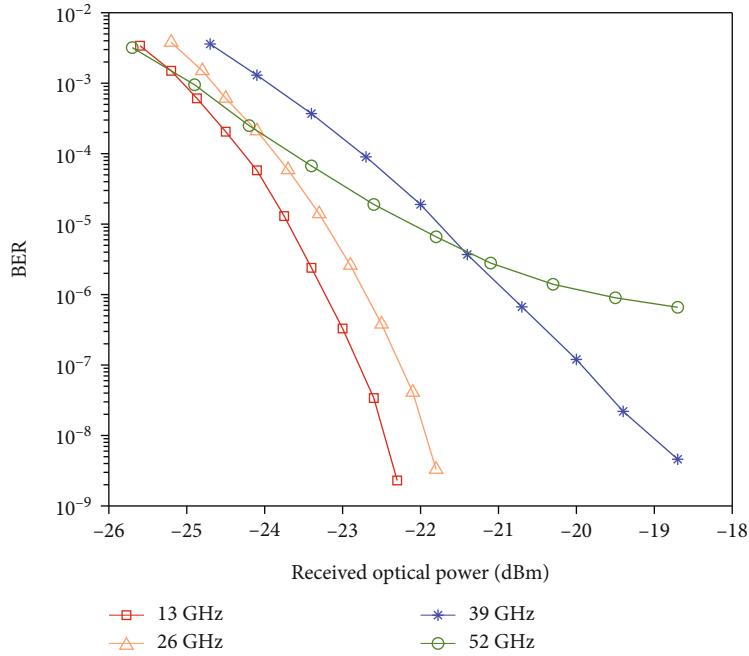


FIGURE 6: BER curves for multiband radio signals.

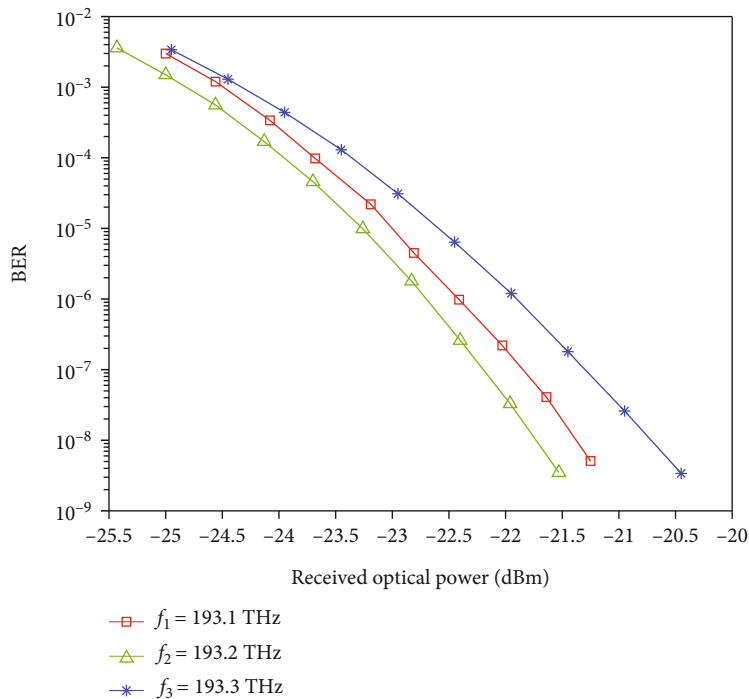


FIGURE 7: BER curves of 26 GHz RF signal over different optical carriers.

can be further expanded if the phase noise of the adjacent optical carriers be controlled effectively.

#### 4. Conclusions

In this paper, a WDM-RoF system with multiband radio signals based on the CS-SCS-BS structure is proposed with a

great dispersion tolerance and flexibility. On the basis of the external MZM, a centralized multiband RF signals generation structure with great flatness for the WDM signal is realized at the SCS, which can not only improve the capacity of the RoF system but also effectively eliminate the dispersion influence on the reliability of the multiband communication system. Moreover, with the use of the WSS and OCs

unit, the system can achieve a flexible allocation of different RF signals to different BSs. Finally, an experimental verification RoF system with three channels WDM signal at 2.5 Gbps which can support multiband radio signals including 13 GHz, 26 GHz, 39 GHz, and 52 GHz over a 40 km SSMF is demonstrated and its communication reliability is achieved. The proposed multiband WDM-RoF system is of great significance in future cooperative and high-band communication systems.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Conflicts of Interest

The authors declare that they have no competing interests.

### Acknowledgments

This work is supported in part by the China Postdoctoral Science Foundation under Grant 2019M651095 and the Fundamental Research Funds for the Central Universities under Grant 3132019348.

### References

- [1] P. T. Dat, A. Kanno, N. Yamamoto, and T. Kawanishi, "Seamless convergence of fiber and wireless systems for 5G and beyond networks," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 592–605, 2019.
- [2] R. Li, Z. Zhao, X. Zhou et al., "Intelligent 5G: when cellular networks meet artificial intelligence," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175–183, 2017.
- [3] J. Lee, E. Tejedor, K. Ranta-aho et al., "Spectrum for 5G: global status, challenges, and enabling technologies," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 12–18, 2018.
- [4] M. J. Marcus, "ITU WRC-19 Spectrum Policy Results," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 4–5, 2019.
- [5] Huawei, *5G Spectrum—Public Policy Position*, White paper, 2016.
- [6] L. Zhang, M. Xiao, G. Wu, M. Alam, Y. C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5G networks," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 44–51, 2017.
- [7] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, pp. 134–142, 2019.
- [8] J. Yao, "Microwave photonics," *Journal of Lightwave Technology*, vol. 27, no. 3, pp. 314–335, 2009.
- [9] H. J. Khashi, V. Sharma, and S. Sergeev, "Phase-stable millimeter-wave generation using switchable dual-wavelength fiber laser," *Optics and Lasers in Engineering*, vol. 137, article 106390, 2021.
- [10] U. Gliese, T. N. Nielsen, S. Nørskov, and K. E. Stubkjaer, "Multifunctional fiber-optic microwave links based on remote heterodyne detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 5, pp. 458–468, 1998.
- [11] R. Pant, D. Marpaung, I. V. Kabakova, B. Morrison, C. G. Poulton, and B. J. Eggleton, "On-chip stimulated Brillouin scattering for microwave signal processing and generation," *Laser & Photonics Reviews*, vol. 8, no. 5, pp. 653–666, 2014.
- [12] W. Xu, X. Gao, M. Zhao, M. Xie, and S. Huang, "Full duplex radio over fiber system with frequency quadrupled millimeter-wave signal generation based on polarization multiplexing," *Optics and Laser Technology*, vol. 103, pp. 267–271, 2018.
- [13] J. Tang, J. Sun, L. Zhao, T. Chen, T. Huang, and Y. Zhou, "Tunable multiwavelength generation based on Brillouin-erbium comb fiber laser assisted by multiple four-wave mixing processes," *Optics express*, vol. 19, no. 15, pp. 14682–14689, 2011.
- [14] J. Ma, J. Yu, C. Yu, X. Zhang, J. Zeng, and L. Chen, "Fiber dispersion influence on transmission of the optical millimeter-waves generated using LN-MZM intensity modulation," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3244–3256, 2007.
- [15] J. Yu, Z. Jia, L. Yi, Y. Su, G. K. Chang, and T. Wang, "Optical millimeter-wave generation or up-conversion using external modulators," *IEEE Photonics Technology Letters*, vol. 18, pp. 265–267, 2006.
- [16] L. Chen, R. Deng, J. He, Q. Chen, Y. Liu, and C. Xiang, "W-band vector signal generation by photonic frequency quadrupling and balanced pre-coding," *IEEE Photonics Technology Letters*, vol. 28, no. 17, pp. 1807–1810, 2016.
- [17] Y. Gao, A. Wen, W. Jiang, D. Liang, W. Liu, and S. Xiang, "Photonic microwave generation with frequency octupling based on a DP-QPSK modulator," *IEEE Photonics Technology Letters*, vol. 27, no. 21, pp. 2260–2263, 2015.
- [18] D. Wang, X. Tang, L. Xi, X. Zhang, and Y. Fan, "A filterless scheme of generating frequency 16-tupling millimeter-wave based on only two MZMs," *Optics and Laser Technology*, vol. 116, pp. 7–12, 2019.
- [19] Y. Wang, L. Pei, J. Li, and Y. Li, "Full-duplex radio-over-fiber system with tunable millimeter-wave signal generation and wavelength reuse for upstream signal," *Applied Optics*, vol. 56, no. 17, pp. 4982–4989, 2017.
- [20] Y. Chen, A. Wen, J. Guo, L. Shang, and Y. Wang, "A novel optical mm-wave generation scheme based on three parallel Mach-Zehnder modulators," *Optics Communications*, vol. 284, no. 5, pp. 1159–1169, 2011.
- [21] H. Chen, T. Ning, J. Li et al., "Optical millimeter-wave generation with tunable multiplication factors and reduced power fluctuation by using cascaded modulators," *Optics and Laser Technology*, vol. 103, pp. 206–211, 2018.

## Research Article

# Highly Efficient MIMO-OFDM with Index Modulation

Zeng Hu <sup>1</sup>, Dehuan Wan <sup>2</sup>, Yankun Chen,<sup>3</sup> Jinpei Yan,<sup>4</sup> Yun Liu <sup>2</sup>, Pengfei Guo <sup>5</sup>,  
Liming Chen,<sup>6</sup> and Wen Zhou<sup>7</sup>

<sup>1</sup>College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

<sup>2</sup>Center for Data Science and Artificial Intelligence, Guangdong University of Finance, Guangzhou 510521, China

<sup>3</sup>South China Sea Marine Survey and Technology Center, State Oceanic Administration, Guangzhou 510300, China

<sup>4</sup>School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

<sup>5</sup>School of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

<sup>6</sup>Electric Power Research Institute of CSG, Guangzhou 510663, China

<sup>7</sup>School of Information Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Dehuan Wan; wan\_e@gduf.edu.cn

Received 9 May 2022; Accepted 29 June 2022; Published 22 July 2022

Academic Editor: Yingyang Chen

Copyright © 2022 Zeng Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Index modulation (IM) is a novel digital modulation technique, which inactivates some subcarriers in orthogonal frequency division multiplexing (OFDM) to exploit the indices of the subcarriers to transmit bits implicitly, and has potential to further improve the energy efficiency and error performance. For the multiple-input multiple-output- (MIMO-) aided IoT devices, a highly efficient and low-complexity IM-aided scheme is needed to reduce the computational complexity at the receiver sides. In this paper, we propose a novel highly efficient MIMO-OFDM with IM scheme by performing IM on each transmit antenna subgroup, which contains two transmit antennas, to achieve two transmit diversity order and significant reduction in computational complexity at the cost of a minor spectral efficiency. To reduce the demodulation complexity, a low-complexity sequential Monte Carlo (SMC) theory-based detector is proposed, which exploits the null space submatrix of the preprocessed channel response matrix by using QR decomposition, to calculate the most likely transmitted IM patterns before the detection of the modulated symbols. Computer simulation results and complexity analysis show that the proposed IM-aided scheme achieves better error performance with extremely low computational complexity under the same constellation and the proposed SMC detector has potential to achieve near optimal bit error rate performance with considerably low demodulation complexity.

## 1. Introduction

The next-generation wireless communication networks will enable various types of devices with different requirements in enhanced mobile broadband (eMBB), massive machine type of communication (mMTC), and ultrareliable and low-latency communication (uRLLC) scenarios [1, 2]. In beyond fifth-generation (B5G) wireless communication networks and the sixth-generation (6G) wireless communication networks, the ubiquitous infrastructure devices and various Internet of Things (IoT) in industry, home, hospital, smart agriculture, and other scenarios will dominate the terminal side in the next decade [3–5]. As a

typical smart devices in mMTC scenario, IoT in most applications are known for its low cost, battery dependent, low power consumption, and limited computing capability equipment, which put forward higher requirements in data transmission and computational complexity. To support machine type of communication (MTC), narrow band IoT (NB-IoT) is specified as the long-term evolution for machine-type (LTE-M) communication by the third-generation partnership project (3GPP) in the fourth-generation (4G) LTE. For the energy and computing capability sensitive devices, low power consumption wireless communication techniques and low-complexity transceiver algorithms are crucial for prolonging the

lifetime and enhancing the robustness of the mMTC networks [6–8].

Index modulation (IM) is a novel emerged promising technique which exploits the index domain of some building blocks of a communication system, such as antennas, time slots, frequency subcarriers, multiple symbol mode or constellation, and signal processing matrices, to convey information bits to meet the demands of low energy consumption and high data transmission [9–12]. To alleviate the interantenna synchronization (IAS) problem at the transmitter and reduce the signal processing complexity at the receiver, spatial modulation (SM), which exploits the index domain of the transmit antennas in a multiple-input multiple-output (MIMO) system, has been proposed to activate one antenna according to the input information bits to transmit the modulated symbol [12–14]. It is clearly that there are two ways to transmit the information bits in SM, which are known as the conventional modulated symbols and the index patterns of the transmit antennas. For the clarity of presentation, we define the information bits transmitted by the modulated symbols and the index patterns are modulation bits and index bits, respectively. To further increase the spectral efficiency (SE), a subgroup of transmit antennas are activated to transmit modulated symbols in generalized SM (GSM) [15, 16]. Then, a novel dual-hop SM-aided relay network is proposed to transmit its own information while forwarding the SM signal to the destination with both decode-and-forward and amplify-and-forward protocols [17]. To reduce the computational complexity in soft decision, the deterministic sequential Monte Carlo- (SMC-) based detectors are proposed for single-carrier (SC) GSM [18]. In [19], the  $K$ -best sphere decoding (SD) soft detection algorithm is proposed to effectively detect the active antennas and the modulated symbols by exploiting the null space of the GSM channel based on the QL decomposition.

In 4G, 5G, and wireless local area network (WLAN) communication standards, orthogonal frequency division multiplexing (OFDM) has become one of the most dominant multicarrier techniques due to its merits in combating with the frequency selective Rayleigh fading channel [20–23]. Owing to the superior of bit error rate (BER) performance and flexible design introduced by the IM technique, IM-aided schemes have attracted considerable attention over the past few years. By applying IM in the frequency subcarrier domain, OFDM with index modulation (OFDM-IM), which appears as a competitive candidate alternative to conventional OFDM, exhibits better bit error rate performance and higher energy efficiency than those of conventional OFDM [24–27]. In this paper, let us define the subcarrier activation patterns (SAPs) to denote the IM patterns of the active statuses of the subcarriers within one subblock after the signal processing of the IM. The achievable rate and mutual information of OFDM-IM are analyzed, which shows that OFDM-IM has potential to outperform conventional OFDM [28, 29]. An intercarrier interference (ICI) self-cancellation scheme is proposed to alleviate the effect in the presence of ICI in the underwater acoustic communication scenario [30]. Then, a novel enhanced coordinate interleaving OFDM-IM is proposed to improve the BER performance of the index bits [31]. By

introducing IM in the spread spectrum, a novel index-modulated OFDM spread spectrum is proposed to exploit the index domain of spreading codes [32]. Then, an IM-aided subcarrier mapping scheme is proposed for dual-hop OFDM-based relay networks with different relay protocols [33] and an OFDM-IM-based distributed cooperative system is analyzed in [34]. To further improve the SE, a dual mode scheme is proposed in which another distinguishable constellation is transmitted at the inactive subcarriers [35]. Moreover, by dividing the constellation into several distinguishable subconstellations, multiple mode- (MM-) aided OFDM-IM scheme is proposed to further improve the system SE and the BER performance [36–38]. By utilizing silent subcarriers and MM, a transmit diversity scheme for OFDM-IM is proposed to further improve the system error performance [39]. In [40], a novel cascade IM scheme is proposed by performing MM IM onto the activated subcarriers in OFDM-IM, which combines the advantages of both MM IM and conventional OFDM-IM and can further enhance the system error performance.

In OFDM-IM, the available subcarriers in each block are divided into several subblocks to reduce the modulation and demodulation complexity. The dependence of the active statuses of the subcarriers in each subblock make the demodulation of the optimal maximum likelihood (ML) detector should be performed in subblock-wise, which leads to extremely high computational complexity at the receiver side. By taking the advantage of the orthogonality of the subcarriers in OFDM, a subcarrier-wise ML detector, which calculates the most likely transmitted symbols in subcarrier-wise using the ML criterion and makes the decision of the estimate of the transmitted symbol vector in subblock-wise, is proposed to reduce computational complexity at the receiver [41]. To further reduce the demodulation complexity, a series of low-complexity detection algorithms, which have potential to achieve near-optimal BER performance, are proposed for OFDM-IM, such as minimum mean square error (MMSE) detector [24], subcarrier-wise-based tree search detectors [41], and log-likelihood ratio- (LLR-) based detectors [42]. With the aid of the low-complexity detection algorithms, OFDM-IM has potential to achieve the same order of the detection complexity as that of conventional OFDM.

MIMO is another key technique for the next-generation wireless networks, which can improve the BER performance and the data rate of the systems [43]. To harvest the advantages of both IM and MIMO techniques, MIMO-OFDM with index modulation (MIMO-OFDM-IM) is proposed to meet the requests of high data rate and reliable data transmission in B5G and 6G [44, 45]. Owing to the advantage of IM in BER performance, MIMO-OFDM-IM exhibits ability to surpass conventional MIMO-OFDM. Specifically, MIMO-OFDM-IM achieves superior BER performance due to the robustness of the index bits, which is theoretically analyzed in [46]. Compared with conventional MIMO-OFDM, MIMO-OFDM-IM can provide an interesting trade-off between SE, error performance, and performing complexity, which provides a more flexible transceiver design.

In MIMO-OFDM-IM, each antenna transmit independent OFDM-IM block, which means that independent



OFDM-IM process is implemented in each branch. Inheriting from OFDM-IM, the dependence of the active statuses of the subcarriers within each subblock makes the detection process becomes an intricate problem. For convenience of expression, we define antenna activation patterns (AAPs) to denote the active statuses of the transmit antennas of each subcarrier level after IM in each transmit antenna. By exploiting the orthogonality of the subcarriers within each subblock, the most likely transmitted symbol vectors under the condition of different AAPs are analyzed in [47], which shows that the optimal error performance can be obtained by performing subcarrier-wise ML detection in the first step and full search of SAPs under the constrain of the IM lookup table of OFDM-IM in each transmit antenna level in the second step. Clearly, each AAP corresponds to a most likely transmitted symbol vector in the subcarrier-wise detection at the receiver. To further alleviate the computational burden at the receiver, LLR-based detectors [44], ordered successive interference cancellation (OSIC-) based MMSE detector [46], and SMC theory-based detectors [47, 48] are proposed for the detection of MIMO-OFDM-IM. Although subcarrier-wise detection algorithms have reduced the demodulation complexity tremendously, the number of the AAPs increases exponentially with the number of the transmit antenna, which limits the applications of MIMO-OFDM-IM in the mMTC scenarios. Therefore, higher efficient IM-aided schemes and low-complexity detection algorithms are critical to the mMTC scenario.

To this end, we propose a novel highly efficient (HE-) MIMO-OFDM-IM scheme for the mMTC scenario, which has a fixed number of the activated transmit antennas at each subcarrier level, to improve the error performance and reduce the computational complexity of signal processing at both transceiver sides. In the proposed HE-MIMO-OFDM-IM scheme, the transmit antennas are divided into several subgroups, and IM is performed within each transmit antenna subgroup to achieve higher transmit diversity. With the fixed number of the activated transmit antennas for all the subcarriers at the antenna dimension, the AAPs, which are similar to those of the SM scheme, are obtained in HE-MIMO-OFDM-IM. Unlike the number of the activated transmit antennas ranges from zero to the number of the transmit antennas  $N_T$  in classical MIMO-OFDM-IM scheme, HE-MIMO-OFDM-IM can use the low-complexity detection algorithms for the SM schemes to further reduce the detection complexity of each subcarrier level, which means that the proposed scheme can achieve near same computational complexity level as SM and harvests the benefits of OFDM-IM. The main contributions of this paper are summarized as follows:

- (i) A novel HE-MIMO-OFDM-IM scheme is proposed to reduce the computational complexity at both the transmitter and receiver sides. By dividing the transmit antennas into several subgroups, higher order of transmit diversity of index patterns and lower demodulation complexity at the receiver can be realized with the aid of the novel IM structure at the transmitter
- (ii) A subcarrier-wise detection architecture, which exploits the AAPs in transmit antenna dimension

to detect the most likely transmitted symbol vectors subcarrier by a subcarrier, is proposed for HE-MIMO-OFDM-IM to easily lift the demodulation burden at the receiver. Then, the computational complexity in terms of the floating-point operations (FLOPs) is analyzed for different detectors, which shows that the proposed HE-MIMO-OFDM-IM can provide a significant reduction in demodulation complexity compared with conventional MIMO-OFDM-IM under the same MIMO configuration

- (iii) A low-complexity two-step SMC detector is proposed for HE-MIMO-OFDM-IM, which exploits the null space submatrix based on the QR decomposition of the preprocessed channel response matrix to calculate the most likely transmitted IM patterns in the first step and employs SMC theory to detect the most likely transmitted symbol vectors of each subcarrier based of the IM patterns obtained in the first step. The simulation results in term of the BER performance of HE-MIMO-OFDM-IM show that the proposed SMC detector has potential to achieve near-optimal BER performance with considerably low computational complexity

The remainder of this paper is organized as follows. The system model of HE-MIMO-OFDM-IM and the optimal ML detector are given in Section 2. Then, a low-complexity detectors is proposed in Section 3. Computer simulation results are given in Section 4. Finally, Section 5 concludes the paper.

*Notation:*  $\mathbf{X}$  denotes a matrix and  $\mathbf{x}$  denotes a column-vector.  $(\cdot)^T$  and  $(\cdot)^H$  denote transposition and Hermitian transposition of a matrix or a vector, respectively.  $\text{diag}\{\mathbf{x}\}$  returns a diagonal matrix whose diagonal elements are included in  $\mathbf{x}$ .  $x \sim \mathcal{CN}(0, \sigma_x^2)$  represents the distribution of a zero mean circularly symmetric complex Gaussian random variable  $x$  with variance  $\sigma_x^2$ .  $\lfloor \cdot \rfloor$  is the integer floor operation, and  $\emptyset$  denotes the empty set.  $\|\cdot\|$  stands for Frobenius norm.  $|\cdot|$  denotes the absolute value of a complex number.  $C(\cdot, \cdot)$  denotes the binomial coefficient.  $p(\cdot)$  denotes the probability of an event.  $\mathcal{S}$  denotes the complex symbol constellation of size  $M$ .  $\mathcal{O}(\cdot)$  denotes the order of detection complexity with respect to the constellation size.  $\text{FFT}\{\cdot\}$  denotes the fast Fourier transform (FFT) operator, and  $\text{IFFT}\{\cdot\}$  denotes the inverse FFT operator.

## 2. System Model

In this section, we present the transceiver structure and the signal model of HE-MIMO-OFDM-IM.

*2.1. Transmitter.* In this paper, we consider a wireless communication scenario in which the devices are equipped with  $N_T$  transmit antennas and  $N_R$  receive antennas. A basic block diagram of the MIMO communication model for the proposed HE-MIMO-OFDM-IM is shown in Figure 1. In HE-MIMO-OFDM-IM, the transmit antennas are divided into  $N_{TA}$  subgroups, each of which contains two antennas, where  $N_{TA} = N_T/2$ . In each subgroup, we can choose two

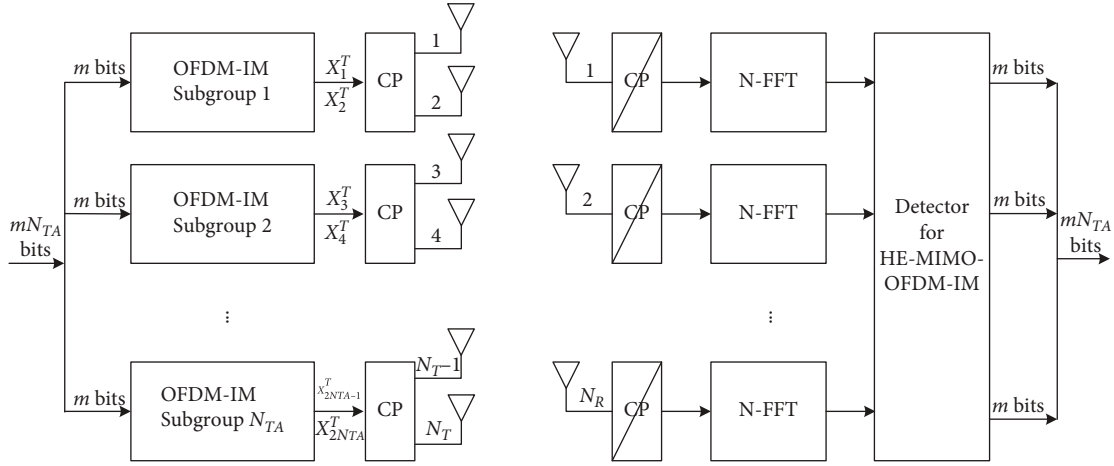


FIGURE 1: Transceiver structure of the HE-MIMO-OFDM-IM system.

adjacent transmit antennas or interleaved transmit antennas to form a subgroup, which are termed as localized grouping and interleaved grouping, respectively. Since modulation and demodulation for the localized grouping and interleaved grouping methods follow the identical procedures, we therefore take two adjacent transmit antennas to form a subgroup to illustrate the signal processes of the proposed HE-MIMO-OFDM-IM.

In HE-MIMO-OFDM-IM, independent OFDM-IM process is performed within each transmit antenna subgroup. For each transmit antenna subgroup, the available  $N$  subcarriers are equally partitioned into  $G$  subblocks to perform independent IM to reduce the modulation complexity, each of which contains  $N_S = N/G$  subcarriers. The basic block diagram of OFDM-IM for each transmit antenna subgroup is given in Figure 2. At each transmission block, a total number of  $mN_{TA}$  input information bits are equally separated into  $N_{TA}$  subgroups to fed into each transmit antenna subgroup, each of which contains  $m$  information bits. Then, the corresponding  $m$  input information bits are further divided into  $G$  subgroups for each subblock, each of which contains  $p = m/G$  information bits. The IM processes in all the subblocks are the same and independent of each other, so let us take the  $g$ th subblock of the  $n_{ta}$ th transmit antenna subgroup as an illustrative case to show the procedures, where  $g \in \{1, 2, \dots, G\}$ ,  $n_{ta} \in \{1, 2, \dots, N_{TA}\}$ . According to the principle of IM,  $K$  out of  $N_S$  subcarriers within the subblock are selected as the active subcarriers to transmit the modulated symbols, which indicates that there are total  $C(N_S, K)$  available SAPs for the IM of each subblock. In this paper, we define  $C_{N_S, K}$  to denote that there are  $N_S$  subcarriers in the subblock and  $K$  subcarriers are activated to transmit modulated symbols. For each subblock, the input  $p$  information bits are transmitted by two ways, the first way is  $p_1 = \log_2[C(N_S, K)]$  index bits transmitted implicitly by the SAPs, while the second way is  $p_2 = N_S \log_2 M$  modulation bits transmitted by  $N_S$  modulated symbols, where  $M$  is the order of the modulated symbol constellation.

In IM process, the index selector chooses one SAP according to the  $p_1$  information bits based on the SAPs lookup table or the combinatorial method. Since  $N_{SAP} =$

$2^{p_1} \leq C(N_S, K)$ , there are always some unused SAPs in IM, which are treated as illegal SAPs and should be avoided in the subcarrier-wise detectors. For the  $g$ th subblock of the  $n_{ta}$ th transmit antenna subgroup, the output of the SAPs selector for the first antenna in the subgroup is given by

$$\mathcal{F}_i^{(n_{ta}, g)} = \{n_1, n_2, \dots, n_K\}, \quad (1)$$

where  $i \in \{1, 2, \dots, N_{SAP}\}$  is the index of SAPs,  $n_k \in \{1, 2, \dots, N_S\}$ ,  $n_j \neq n_k$  if  $j \neq k$ , and the elements in  $\mathcal{F}_i^{(n_{ta}, g)}$  are sorted in ascending order. The SAP of the second antenna in the  $n_{ta}$ th transmit antenna subgroup is defined as the complementary set of the  $\mathcal{F}_i^{(n_{ta}, g)}$ , which can be expressed as

$$\bar{\mathcal{F}}_i^{(n_{ta}, g)} = \{\bar{n}_1, \bar{n}_2, \dots, \bar{n}_{N_S - K}\}, \quad (2)$$

where the elements in  $\bar{\mathcal{F}}_i^{(n_{ta}, g)}$  are sorted in ascending order,  $\bar{n}_k \in \{1, 2, \dots, N_S\}$ ,  $\bar{n}_j \neq \bar{n}_k$  if  $j \neq k$ ,  $\mathcal{F}_i^{(n_{ta}, g)} \cup \bar{\mathcal{F}}_i^{(n_{ta}, g)} = \{1, 2, \dots, N_S\}$ ,  $\mathcal{F}_i^{(n_{ta}, g)} \cap \bar{\mathcal{F}}_i^{(n_{ta}, g)} = \emptyset$ . Then, the  $N_S \times 1$  modulated symbol vector is generated according to the rest  $p_2$  information bits, where the output of the  $M$ -ary symbol modulator is given by

$$\mathbf{s}^{(n_{ta}, g)} = [s_1, s_2, \dots, s_K, s_{K+1}, \dots, s_{N_S}]^T, \quad (3)$$

where  $s_k \in \mathcal{S}$  is an element of  $M$ -ary quadrature amplitude modulation (QAM) or phase-shift keying (PSK) constellation,  $k = 1, 2, \dots, N_S$ . In  $\mathbf{s}^{(n_{ta}, g)}$ , the first  $K$  modulated symbols are used to generate the transmitted symbol vector of the first transmit antenna, and the rest are for the second transmit antenna. We assume that the modulated symbols are normalized to unit average power, i.e.  $E\{\mathbf{ss}^H\} = N_S$ . Then, the OFDM-IM subblock creator maps the modulated symbols to the activated subcarriers according to the obtained SAPs. The final  $N_S \times 1$  transmitted symbol

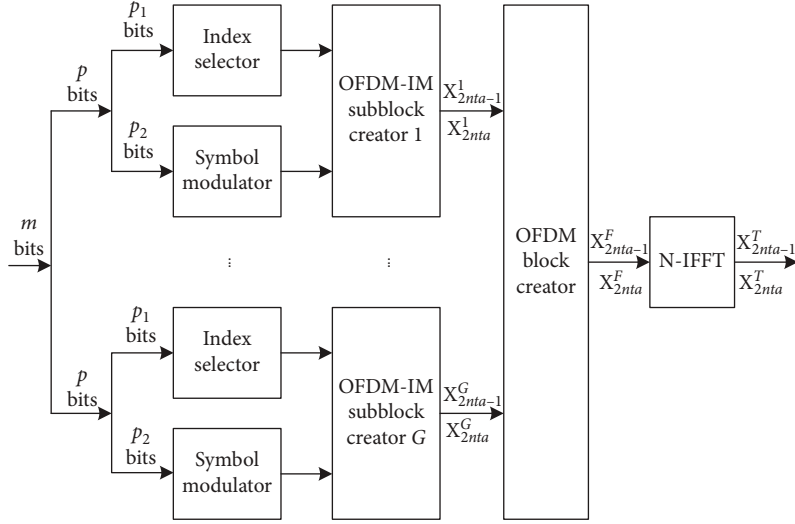


FIGURE 2: Block diagram of the OFDM-IM transmitter in each transmit antenna subgroup.

vector of the first transmit antenna, whose index is given by  $2n_{ta} - 1$ , can be expressed as

$$\mathbf{x}_{2n_{ta}-1}^g = [x_1^g, x_2^g, \dots, x_{N_S}^g]^T, \quad (4)$$

where  $x_k^g \in \{\mathcal{S}, 0\}$ ,  $k = 1, 2, \dots, N_S$ . Similarly, the  $N_S \times 1$  transmitted symbol vector of the second transmit antenna with index  $2n_{ta}$  is given by

$$\mathbf{x}_{2n_{ta}}^g = [x_1^g, x_2^g, \dots, x_{N_S}^g]^T. \quad (5)$$

After  $G$  subblocks corresponding to each transmit antenna are obtained, the transmitted symbol vector can be generated by concatenating these modulated symbol vectors in order. The transmitted symbol vectors of the  $n_{ta}$ th transmit antenna subgroup are given by

$$\mathbf{x}_t^F = [(\mathbf{x}_t^1)^T, (\mathbf{x}_t^2)^T, \dots, (\mathbf{x}_t^G)^T]^T, \quad (6)$$

where the index of transmit antenna is given by  $t = 2n_{ta} - 1, 2n_{ta}$ ,  $n_{ta} = 1, 2, \dots, N_{TA}$ . After this point, independent inverse fast Fourier transform (IFFT) for each transmit antenna is performed to generate the time domain transmitted symbol vector, which is given by

$$\mathbf{x}_t^T = \text{IFFT}(\mathbf{x}_t^F) = [x_1^T, x_2^T, \dots, x_N^T]^T, \quad (7)$$

where  $t = 2n_{ta} - 1, 2n_{ta}$  and  $n_{ta} = 1, 2, \dots, N_{TA}$ . A cyclic prefix (CP) of  $N_{CP}$  samples, which is longer than the number of the channel taps, is appended to the beginning of the time domain transmitted symbol vector in each transmit antenna. Then, identical procedures as conventional OFDM such as parallel to serial and digital-to-analog are performed to get the high-frequency broadband signals, which are sent simultaneously from all the trans-

mit antennas to the destination. We assume that the frequency selective Rayleigh fading channel remains constant during one block and changes blast to blast. For the clarity of presentation, an example of IM process is given as follows.

*Example 1.* In HE-MIMO-OFDM-IM, two transmit antennas form a subgroup as a unit for IM to maintain fixed number of the activated transmit antennas for each subcarrier. We assume the parameters in HE-MIMO-OFDM-IM are given by  $N_T = N_R = 4$ ,  $N_{TA} = 2$ ,  $N_S = 4$ ,  $K = 2$ , and  $M = 16$ . Table 1 gives an example of SAP lookup table and the corresponding complementary set. Without loss of generality, we take the modulation process of the  $g$ th subblock as an illustrative case. In this case, a total of  $p_1 = 2$  index bits and  $p_2 = 16$  modulation bits can be transmitted per subblock. For convenience of expression, let us define  $p_1^{(n_{ta})}$  to denote the index bits of the  $n_{ta}$ th transmit antenna subgroup,  $n_{ta} = 1, 2, \dots, N_{TA}$ . Assuming the input index bits for two transmit antenna subgroups of the  $g$ th subblock are given by  $p_1^{(1)} = [1, 0]$  and  $p_1^{(2)} = [0, 0]$ . The SAP and the complementary set of the first transmit antenna subgroup can be easily obtained according to lookup table, which are given by  $\{2, 3\}$  and  $\{1, 4\}$ , respectively. Similarly, the SAP and the complementary set of the second transmit antenna subgroup are given by  $\{1, 3\}$  and  $\{2, 4\}$ , respectively. Finally, the transmit symbol matrix of the  $g$ th subblock is given by

$$\mathbf{X}^g = \begin{bmatrix} 0 & s & s & 0 \\ s & 0 & 0 & s \\ s & 0 & s & 0 \\ 0 & s & 0 & s \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^g \\ \mathbf{x}_2^g \\ \mathbf{x}_3^g \\ \mathbf{x}_4^g \end{bmatrix}, \quad (8)$$

$$\mathbf{X}^g = [\bar{\mathbf{x}}_1^g \bar{\mathbf{x}}_2^g \bar{\mathbf{x}}_3^g \bar{\mathbf{x}}_4^g], \quad (9)$$

TABLE 1: Reference SAPs lookup table with parameters C 4, 2.

Bits	SAPs	Complementary set
[0, 0]	{1, 3}	{2, 4}
[0, 1]	{1, 4}	{2, 3}
[1, 0]	{2, 3}	{1, 4}
[1]	{2, 4}	{1, 3}

where  $\mathbf{x}_t^g$  with  $t = 1, 2, \dots, 4$  in (8) denote the obtained transmitted symbol vectors in (4) and (5), and  $\bar{\mathbf{x}}_n^g$  denotes the transmitted symbol vector corresponding to the  $n$ th subcarriers in the  $g$ th subblock, which is the data vector contains the simultaneously transmitted symbols from all the transmit antennas, and is the  $n$ th column vector in  $\mathbf{X}^g$ ,  $n = 1, 2, \dots, 4$ . According to IM method, the transmitted symbol vectors in (8) with odd indices of transmit antennas have  $K$  activated subcarriers, while the remaining with even indices have  $N_S - K$  activated subcarriers. Furthermore, the transmitted symbol vectors  $\mathbf{x}_t^g$  in (8) are the final output of IM in each transmit antenna level at the transmitter, while  $\bar{\mathbf{x}}_n^g$  in (9) forms the basic signal demodulation model in subcarrier-wise detectors.

Similar to the SM systems, the AAPs are defined to denote the active statuses of  $\bar{\mathbf{x}}_n^g$  in (9). To illustrate the AAPs more clarity, an example of AAPs with parameters  $N_T = 4$  and  $N_{TA} = 2$  is given in Table 2, and the number of the AAPs is given by  $N_{AAP} = 2^{N_{TA}}$ . Let us define space-frequency activation patterns (SFAPs) set  $\mathcal{X}$  to denote the active statuses of the transmitted symbol matrix  $\mathbf{X}^g$ , which is the combinations of the SAPs of each transmit antenna subgroups. Since the independent IM of each transmit antenna subgroups, there are a total of  $N_{SFAP} = (N_{SAP})^{N_{TA}}$  SFAPs in the set  $\mathcal{X}$ . The SFAP of the transmitted symbol matrix in (8) is given by

$$\mathcal{X}^g = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad (10)$$

where 1 denotes the corresponding subcarrier is activated to transmit modulated symbol, while 0 denotes the inactive subcarrier.

**2.2. Receiver.** At the receiver, the time domain received signals are transformed to frequency signals by FFT operation in each branch after removing the CP samples. According to the signal model of MIMO, the received signals of the receiver antenna  $r$  ( $1 \leq r \leq N_R$ ) corresponding to the  $g$ th ( $1 \leq g \leq G$ ) subblock are given by

$$\mathbf{y}_r^g = \sum_{t=1}^{N_T} \gamma \text{diag}(\mathbf{h}_{r,t}^g) \mathbf{x}_t^g + \mathbf{w}_r^g, \quad (11)$$

where  $\mathbf{y}_r^g = [y_{r,1}^g, y_{r,2}^g, \dots, y_{r,N_S}^g]^T$  denotes the  $N_S \times 1$  received signal vector,  $\gamma = \sqrt{N_S/K}$  is the power reallocation factor at

TABLE 2: Reference AAPs for HE-MIMO-OFDM-IM with parameters  $N_T = 4$  and  $N_{TA} = 2$ .

AAPs	$\mathcal{A}(1)$	$\mathcal{A}(2)$	$\mathcal{A}(3)$	$\mathcal{A}(4)$
Active state	$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

the transmitter,  $\mathbf{h}_r^g = [h_{r,t,1}^g, h_{r,t,2}^g, \dots, h_{r,t,N_S}^g]^T$  denotes the  $N_S \times 1$  channel frequency response between receive antenna  $r$  and transmit antenna  $t$ , and  $\mathbf{w}_r^g = [w_{r,1}^g, w_{r,2}^g, \dots, w_{r,N_S}^g]^T$  denotes the  $N_S \times 1$  frequency additive white Gaussian noise (AWGN) vector with zero mean and  $N_0$  variance. We define the signal to noise ratios (SNR) of the system as  $r_{\text{SNR}} = E_b/N_0$ , where  $E_b$  is the average transmitted energy per bit. Owing to the orthogonality between different subcarriers in OFDM system, the calculation can be performed in subcarrier-wise within each subblock. According to the transmitted symbol vector model given in (9) and received signal model in (11), the received signal vector for the  $n$ th ( $1 \leq n \leq N_S$ ) subcarrier within the  $g$ th subblock can be expressed as

$$\begin{bmatrix} y_{1,n}^g \\ y_{2,n}^g \\ \vdots \\ y_{N_R,n}^g \end{bmatrix} = \gamma \begin{bmatrix} h_{1,1,n}^g & h_{1,2,n}^g & \cdots & h_{1,N_T,n}^g \\ h_{2,1,n}^g & h_{2,2,n}^g & \cdots & h_{2,N_T,n}^g \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1,n}^g & h_{N_R,2,n}^g & \cdots & h_{N_R,N_T,n}^g \end{bmatrix} \times \begin{bmatrix} x_{1,n}^g \\ x_{2,n}^g \\ \vdots \\ x_{N_T,n}^g \end{bmatrix} + \begin{bmatrix} w_{1,n}^g \\ w_{2,n}^g \\ \vdots \\ w_{N_T,n}^g \end{bmatrix}, \quad (12)$$

$$\bar{\mathbf{y}}_n^g = \gamma \bar{\mathbf{H}}_n^g \bar{\mathbf{x}}_n^g + \bar{\mathbf{w}}_n^g$$

where  $\bar{\mathbf{y}}_n^g$  is the  $N_R \times 1$  received signal vector from all the receive antennas at the  $n$ th subcarrier,  $\bar{\mathbf{H}}_n^g$  denotes the corresponding  $N_R \times N_T$  channel matrix between the all the receive antennas and transmit antennas at the  $n$ th subcarrier, and  $\bar{\mathbf{w}}_n^g$  denotes the  $N_R \times 1$  AWGN vector. Due to the IM, only  $N_{TA}$  transmit antennas are activated to transmit nonzero modulated symbols, which can not be determined only based on the observation of current subcarrier. Therefore, the subcarrier-wise detection algorithms need to calculate the most likely transmitted symbol vectors corresponding to all the AAPs for each subcarrier within the subblock. Then, a joint decision is made based on the obtained most likely transmitted symbol vectors of  $N_S$  subcarriers. For the  $n$ th subcarrier, the subcarrier-wise ML detector calculates the most likely transmitted symbol vectors under the condition of each AAP

$$\hat{\bar{\mathbf{x}}}_{n,\mathcal{A}(a)}^g = \arg \min_{\bar{\mathbf{x}} \in \mathcal{A}(a)} \|\bar{\mathbf{y}}_n^g - \gamma \bar{\mathbf{H}}_n^g \bar{\mathbf{x}}\|^2, \quad (13)$$

where  $a = 1, 2, \dots, N_{AAP}$ ,  $\bar{\mathbf{x}} \in \mathcal{A}(a)$  denote all the possible transmitted symbol vectors whose AAP is  $\mathcal{A}(a)$ , and  $\hat{\bar{\mathbf{x}}}_{n,\mathcal{A}(a)}^g$  denotes the obtained most likely transmitted symbol

vector corresponding to the AAP  $\mathcal{A}(a)$ . The metrics of the obtained most likely transmitted symbol vectors of the  $n$ th subcarrier can be expressed as

$$p\left(\widehat{\mathbf{x}}_{n,\mathcal{A}(a)}^g\right) = \left\| \bar{\mathbf{y}}_n^g - \gamma \bar{\mathbf{H}}_n^g \widehat{\mathbf{x}}_{n,\mathcal{A}(a)}^g \right\|^2. \quad (14)$$

After all the most likely transmitted symbol vectors of each subcarrier in the  $g$ th subblock are obtained, an exhaustive search of all the SFAPs is performed to obtain the final estimate of the transmitted symbol matrix, which can be calculated as

$$\widehat{\mathbf{X}}^g = \arg \min_{\mathcal{A}_n \in \mathcal{X}} \sum_{n=1}^{N_S} p\left(\widehat{\mathbf{x}}_n^g \in \mathcal{A}_n\right), \quad (15)$$

where  $\mathcal{A}_n$  denotes the AAP of the  $n$ th subcarrier which is the  $n$ th column of a given  $\mathcal{X}$ , and  $\mathcal{A}_n \in \{\mathcal{A}(1), \mathcal{A}(2), \dots, \mathcal{A}(N_{\text{AAP}})\}$ . Although the subcarrier-wise ML detector can achieve the optimal error performance, both the detection complexity in the first step and the second step increase exponentially with the order of constellation and SAPs, which are given by  $\mathcal{O}(M^{N_{\text{TA}}})$  and  $\mathcal{O}((N_{\text{SAP}})^{N_{\text{TA}}})$ , respectively. The demodulation complexity in terms of the FLOPs to get the estimate of the transmitted symbol matrix per subcarrier is given by

$$(3N_R N_T + 2N_R - 1)M^{N_{\text{TA}}} N_{\text{AAP}} + (N_S - 1)(N_{\text{SAP}})^{N_{\text{TA}}}. \quad (16)$$

### 3. Low-Complexity Detector

The computational complexity of the optimal ML detector increases exponentially with the order of the constellation and the number of the transmit antenna subgroups, which give rise to prohibitive computational complexity in the practical applications in the mMTC scenario. In this section, a novel two-step SMC detector is proposed to reduce detection complexity. In the first step, a null space algorithm is proposed to calculate the metrics of the AAPs of each subcarrier, which can be used to calculate the most likely transmitted SFAPs of the subblock. Then, the most likely transmitted symbol vectors of each subcarrier can be calculated according to the obtained SFAPs. Before the detection, we first calculate the noise whitening matrix of the  $n$ th ( $1 \leq n \leq N_S$ ) subcarrier, which is given by

$$\Omega_n^g = \left( (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{H}}_n^g \right)^{1/2}. \quad (17)$$

The output after the signal processing of noise whitening is given by

$$\begin{aligned} \bar{\mathbf{y}}_{n,NW}^g &= (\Omega_n^g)^{-1} (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{y}}_n^g \\ &= \gamma \left( (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{H}}_n^g \right)^{-1/2} (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{H}}_n^g \bar{\mathbf{x}}_n^g + (\Omega_n^g)^{-1} (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{w}}_n^g \\ &= \gamma \left( (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{H}}_n^g \right)^{1/2} \bar{\mathbf{x}}_n^g + (\Omega_n^g)^{-1} (\bar{\mathbf{H}}_n^g)^H \bar{\mathbf{w}}_n^g = \gamma \Omega_n^g \bar{\mathbf{x}}_n^g + \bar{\mathbf{w}}_{n,NW}^g, \end{aligned} \quad (18)$$

where  $\bar{\mathbf{y}}_{n,NW}^g$  denotes the  $N_T \times 1$  received signal vector after the signal processing and  $\bar{\mathbf{w}}_{n,NW}^g$  denotes the processed  $N_T \times 1$  AWGN sample vector. Since IM is performed in each branch, there are some inactive antennas in each subcarrier, which indicates that there are some 0 symbols in  $\bar{\mathbf{x}}_n^g$ . Assume the AAP of the  $n$ th subcarrier is  $\mathcal{A}_n$ , the formula in (18) can be rewritten as

$$\bar{\mathbf{y}}_{n,NW}^g = \gamma \Omega_{n,\mathcal{A}_n}^g \bar{\mathbf{x}}_{n,\mathcal{A}_n}^g + \bar{\mathbf{w}}_{n,NW}^g, \quad (19)$$

where  $\Omega_{n,\mathcal{A}_n}^g$  denotes the submatrix with dimension  $N_T \times N_{\text{TA}}$  that deletes the columns corresponding to the inactive subcarriers in  $\mathcal{A}_n$  and  $\bar{\mathbf{x}}_{n,\mathcal{A}_n}^g$  denotes the  $N_{\text{TA}} \times 1$  modulated symbol vector by deleting the zero symbols in  $\bar{\mathbf{x}}_n^g$  according to its AAP  $\mathcal{A}_n$ . By performing the QR decomposition of the matrix  $\Omega_{n,\mathcal{A}_n}^g$ , the result can be expressed as

$$\Omega_{n,\mathcal{A}_n}^g = \left[ \mathbf{Q}_{\alpha,n,\mathcal{A}_n}^g, \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right] \begin{bmatrix} \left( \mathbf{R}_{\alpha,n,\mathcal{A}_n}^g \right)_{N_{\text{TA}} \times N_{\text{TA}}} \\ \mathbf{0}_{(N_T - N_{\text{TA}}) \times N_{\text{TA}}} \end{bmatrix} = \mathbf{Q}_{\alpha,n,\mathcal{A}_n}^g \mathbf{R}_{\alpha,n,\mathcal{A}_n}^g, \quad (20)$$

where  $\mathbf{Q}_{\alpha,n,\mathcal{A}_n}^g$  is an  $N_T \times N_{\text{TA}}$  unitary matrix,  $\mathbf{Q}_{\beta,n,\mathcal{A}_n}^g$  is the null space matrix with dimension  $N_T \times (N_T - N_{\text{TA}})$ , and  $\mathbf{R}_{\alpha,n,\mathcal{A}_n}^g$  is an  $N_{\text{TA}} \times N_{\text{TA}}$  upper triangular matrix. From (20), we have

$$\left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H \Omega_{n,\mathcal{A}_n}^g = \mathbf{0}, \quad (21)$$

which implies that  $\left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H$  is orthogonal to the matrix  $\Omega_{n,\mathcal{A}_n}^g$ . By left multiply  $\left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H$  to the received signal vector  $\bar{\mathbf{y}}_{n,NW}^g$  given in (18), the output can be expressed as

$$\bar{\mathbf{y}}_{n,C}^g = \left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H \bar{\mathbf{y}}_{n,NW}^g = \gamma \left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H \Omega_{n,\mathcal{A}_n}^g \bar{\mathbf{x}}_{n,\mathcal{A}_n}^g + \left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H \bar{\mathbf{w}}_{n,NW}^g. \quad (22)$$

If  $\mathbf{Q}_{\beta,n,\mathcal{A}_n}^g$  is the correct null space matrix, the result of (22) can be expressed as  $\bar{\mathbf{y}}_{n,C}^g = \left( \mathbf{Q}_{\beta,n,\mathcal{A}_n}^g \right)^H \bar{\mathbf{w}}_{n,NW}^g$ , which follows zero mean Gaussian distribution; otherwise, the transmitted symbol vectors are nonorthogonal to the null space and the output will be nonzero mean variables. Therefore,  $\|\bar{\mathbf{y}}_{n,C}^g\|^2$  can be used as the metrics for the detection of the possible AAPs of each subcarrier and the SFAPs of each subblock. Let us define  $\varphi_n^g(\mathcal{A}(a))$  to denote the metric of the  $n$ th subcarrier corresponding to the AAP  $\mathcal{A}(a)$ , which is given by

$$\varphi_n^g(\mathcal{A}(a)) = \left\| \left( \mathbf{Q}_{\beta,n,\mathcal{A}(a)}^g \right)^H \bar{\mathbf{y}}_{n,NW}^g \right\|^2, \quad (23)$$

where  $\mathbf{Q}_{\beta,n,\mathcal{A}(a)}^g$  denotes the null space of the  $n$ th subcarrier obtained from the submatrix  $\Omega_{n,\mathcal{A}(a)}^g$  which is generated by

deleting the columns corresponding to the inactive subcarriers according to  $\mathcal{A}(a)$ ,  $a = 1, 2, \dots, N_{\text{AAP}}$ .

After the calculation of  $N_S$  subcarriers within the  $g$ th subblock, the estimate of SFAP can be obtained by using a SD-like algorithm to find the most likely transmitted SFAPs of the subblock in the subcarrier to subcarrier manner based on the metrics obtained in (23). Considering there are some illegal SAPs in IM in each branch, which will lead to catastrophic errors both in the demodulation of IM bits and the location of the modulated symbols as well as their order. Therefore, we need to check the validity of the SAP of each transmit antenna subgroup during the searching process. The SD-like algorithm can be performed in several steps as follows.

Step 1: let us define  $\hat{\mathcal{X}}_{n,k}^g$  to denote the survival partial SFAPs reach the  $n$ th ( $1 \leq n \leq N_S - 1$ ) subcarrier level, where the  $k$ th survival partial SFAPs can be expressed as

$$\hat{\mathcal{X}}_{n,k}^g = [\mathcal{A}_{1,k}^g, \mathcal{A}_{2,k}^g, \dots, \mathcal{A}_{n,k}^g], \quad (24)$$

where  $\mathcal{A}_{n,k}^g$  denotes the AAP at the  $n$ th subcarrier in the  $k$ th survival partial SFAPs  $\hat{\mathcal{X}}_{n,k}^g$ ,  $k = 1, 2, \dots, \theta_A$  denotes the index of the survival partial SFAPs and  $\theta_A$  is the maximum number of the reserved survival partial SFAPs at each subcarrier level. The corresponding cumulative metrics of the partial SFAPs can be calculated as

$$\varphi(\hat{\mathcal{X}}_{n,k}^g) = \sum_{i=1}^n \varphi_i^g(\mathcal{A}_{i,k}^g), \quad (25)$$

where  $\varphi_i^g(\mathcal{A}_{i,k}^g)$  denotes the metric at  $i$ th subcarrier corresponding to the AAP in the  $k$ th survival partial SFAP  $\hat{\mathcal{X}}_{n,k}^g$ . Then, the partial SFAPs of the  $(n+1)$ th subcarrier can be directly calculated based on the obtained calculation results in (24) and (25).

Step 2: the entire partial SFAPs at the  $(n+1)$ th subcarrier level based on the obtained results in (24) and (25) can be expressed as

$$\hat{\mathcal{X}}_{n+1,l}^g = [\hat{\mathcal{X}}_{n,k}^g, \mathcal{A}(a)] = [\mathcal{A}_{1,k}^g, \mathcal{A}_{2,k}^g, \dots, \mathcal{A}_{n,k}^g, \mathcal{A}(a)], \quad (26)$$

where  $l = 1, 2, \dots, \theta_A N_{\text{AAP}}$  is the index of entire partial SFAPs at the  $(n+1)$ th subcarrier level,  $k = 1, 2, \dots, \theta$ ,  $a = 1, 2, \dots, N_{\text{AAP}}$ . The cumulative metrics of the partial SFAPs in (26) can be obtained by updating the metrics in (25), which can be calculated as

$$\varphi(\hat{\mathcal{X}}_{n+1,l}^g) = \varphi(\hat{\mathcal{X}}_{n,k}^g) + \varphi_{n+1}^g(\mathcal{A}(a)). \quad (27)$$

After this point,  $\theta_A$  partial SFAPs with better metrics are kept as the survival partial SFAPs at the  $(n+1)$ th subcarrier level for the calculation of the next subcarrier level. In this process, we need to check the validity of the SFAPs of each transmit antenna subgroup, which can be actualized with the aid of an illegal SAP table at each subcarrier level. An example of the illegal SAPs table of each subcarrier level cor-

responding to the SAPs given in Table 1 is given in Table 3. The obtained survival SFAPs and the corresponding metrics are given by

$$\hat{\mathcal{X}}_{n+1,k}^g = [\mathcal{A}_{1,k}^g, \mathcal{A}_{2,k}^g, \dots, \mathcal{A}_{n,k}^g, \mathcal{A}_{n+1,k}^g], \quad (28)$$

$$\varphi(\hat{\mathcal{X}}_{n+1,k}^g) = \sum_{i=1}^{n+1} \varphi_i^g(\mathcal{A}_{i,k}^g), \quad (29)$$

respectively, where the index  $k$  is given by  $k = 1, 2, \dots, \theta_A$ . Note that although the formulae expresses in (28) and (29) have the same form as that in (24) and (25), the elements in them may different even with same index  $k$  because some partial SFAPs may be eliminated due to the metrics in (27). Finally, let us set  $n_{\text{current}} = n+1$  and check whether the search process reach the last subcarrier level. If  $n_{\text{current}} = N_S$ , go to Step 3 for the following demodulation procedures; otherwise set  $n = n_{\text{current}}$  and repeat Step 2.

Step 3: when it arrives the last subcarrier level,  $\theta_F$  SFAPs with better metrics are selected as the most likely transmitted SFAPs for the calculation of the modulated symbols, which are given by

$$\hat{\mathcal{X}}_{N_S,c}^g = [\mathcal{A}_{1,c}^g, \mathcal{A}_{2,c}^g, \dots, \mathcal{A}_{N_S,c}^g], \quad (30)$$

where  $c = 1, 2, \dots, \theta_F$ . For convenience of expression, let us define  $\mathcal{X}_n^g = \{\mathcal{A}_{n,1}^g, \mathcal{A}_{n,2}^g, \dots, \mathcal{A}_{n,\theta_F}^g\}$  to denote the set of the AAPs of the  $n$ th subcarrier in  $\hat{\mathcal{X}}_{N_S,c}^g$ ,  $n = 1, 2, \dots, N_S$ .

According to the obtained AAP set of the  $n$ th subcarrier, the most likely transmitted symbol vectors can be calculated based on the SMC theory. Note that the AAPs of each subcarrier may identical in different SFAP estimates, the most likely transmitted symbol vectors corresponding to these AAPs are only needed to be calculated once. For a given AAP  $\mathcal{A}_{n,c}^g$ , the noising submatrix  $\mathbf{\Omega}_{n,\mathcal{A}_{n,c}^g}^g$  which only keeps the columns corresponding to the activated transmit antennas according to the AAP  $\mathcal{A}_{n,c}^g$ . Then, QR decomposition given in (20) is perform to get the demodulation signal model for SMC calculation. Consequently, we have

$$\bar{\mathbf{y}}_{n,c}^g = \left( \mathbf{Q}_{\alpha,n,\mathcal{A}_{n,c}^g}^g \right)^H \bar{\mathbf{y}}_{n,NW}^g = \gamma \mathbf{R}_{\alpha,n,\mathcal{A}_{n,c}^g}^g \bar{\mathbf{x}}_{n,c}^g + \left( \mathbf{Q}_{\alpha,n,\mathcal{A}_{n,c}^g}^g \right)^H \bar{\mathbf{w}}_{n,NW}^g, \quad (31)$$

where  $\bar{\mathbf{x}}_{n,c}^g$  denotes the nonzero modulated symbol vector of the transmitted symbol vector  $\bar{\mathbf{x}}_n^g$  corresponding to the AAP  $\mathcal{A}_{n,c}^g$ . Formula (31) can be rewritten in matrix wise

$$\begin{bmatrix} y_{1,n} \\ y_{2,n} \\ \vdots \\ y_{K,n} \end{bmatrix} = \gamma \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,K} \\ 0 & r_{2,2} & \cdots & r_{2,K} \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{K,K} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_K \end{bmatrix} + \begin{bmatrix} \bar{w}_1 \\ \bar{w}_2 \\ \vdots \\ \bar{w}_K \end{bmatrix}. \quad (32)$$

TABLE 3: Illegal paths table for OFDM-IM with parameters C 4, 2.

Subcarrier level	Illegal paths
Subcarrier 1	$\emptyset$
Subcarrier 2	[0 0], [1 1]
Subcarrier 3	$\emptyset$
Subcarrier 4	[0 1 0 0], [0 1 1 1] [1 0 0 0], [1 0 1 1]

According to the SMC theory, the detection process of the most likely transmitted symbol vectors associated with  $\mathcal{A}_{n,c}^g$  can be performed in antenna to antenna manner from the  $K$ th activated transmit antenna to the first activated transmit antenna. Based on the sequential distribution and the sequential structure derived in [48], the *a posteriori* probability at the  $k$ th activated transmit antenna level can be calculated as

$$\begin{aligned} p(\{\bar{\mathbf{s}}_{n,c}^g\}_k | \{\bar{\mathbf{y}}_{n,c}^g\}_k) &\propto p(\{\bar{\mathbf{y}}_{n,c}^g\}_k | \{\bar{\mathbf{s}}_{n,c}^g\}_k) \\ &\propto \prod_{t=K}^k p(\mathcal{Y}_{n,t}^g | \{\bar{\mathbf{s}}_{n,c}^g\}_t) \propto p(\{\bar{\mathbf{y}}_{n,c}^g\}_{k+1} | \{\bar{\mathbf{s}}_{n,c}^g\}_{k+1}) p(\mathcal{Y}_{n,k}^g | \{\bar{\mathbf{s}}_{n,c}^g\}_k), \end{aligned} \quad (33)$$

where  $\{\bar{\mathbf{s}}_{n,c}^g\}_k \triangleq [s_k, s_{k+1}, \dots, s_K]^T$  is the partial symbol vector when the calculation reaches the  $k$ th activated transmit antenna level and  $\{\bar{\mathbf{y}}_{n,c}^g\}_k \triangleq [y_k, y_{k+1}, \dots, y_K]^T$  denotes the partial receive signal vector reaches the  $k$ th activated transmit antenna level. For the calculation of each activated transmit antenna level, we define the  $(K - k + 1) \times 1$  nonzero vector  $\mathbf{r}_k = [r_{k,k}, r_{k,k+1}, \dots, r_{k,K}]$  to denote the vector which is constructed by  $K - k + 1$  nonzero elements in the  $k$ th row of the matrix  $\mathbf{R}_{\alpha,n,\mathcal{A}_{n,c}^g}^g$ . Therefore, the SMC algorithm draws particles from each activated transmit antenna level and keeps  $\theta_S$  particles according to the probabilities in (33) for the calculation of the next antenna level. After the calculation of the  $k$ th activated transmit antenna level,  $\theta_S$  reserved particles can be expressed as  $\{\bar{\mathbf{s}}_{n,c}^g\}_k^{(\theta)}$  with their important weights  $p(\{\bar{\mathbf{y}}_{n,c}^g\}_k | \{\bar{\mathbf{s}}_{n,c}^g\}_k^{(\theta)})$ , where  $\theta = 1, 2, \dots, \theta_S$ . In the SMC algorithm, the particles at the  $(k - 1)$ th activated transmit antenna level can be expressed as

$$\{\bar{\mathbf{s}}_{n,c}^g\}_{k-1}^{(l)} = \left[ s_{k-1}, \left( \{\bar{\mathbf{s}}_{n,c}^g\}_k^{(\theta)} \right)^T \right]^T, \quad (34)$$

where  $l = 1, 2, \dots, M\theta_S$ . Note that all the particles at the  $(k - 1)$ th activated transmit antenna level are updated from  $\theta_S$  particles at the  $k$ th activated transmit antenna level, which indicates that the important weights can be updated from that of the  $k$ th activated transmit antenna level. The updating of the important weights at the next antenna level, which is performed based of the important weights of the obtained  $\theta_S$  particles, can be calculated as

$$p(\{\bar{\mathbf{y}}_{n,c}^g\}_{k-1} | \{\bar{\mathbf{s}}_{n,c}^g\}_{k-1}^{(l)}) \propto p(\{\bar{\mathbf{y}}_{n,c}^g\}_k | \{\bar{\mathbf{s}}_{n,c}^g\}_k^{(\theta)}) p(\mathcal{Y}_{n,k-1}^g | \{\bar{\mathbf{s}}_{n,c}^g\}_{k-1}^{(l)}). \quad (35)$$

When it reaches the first activated transmit antenna level, the final  $\theta_S$  particles can be expressed as  $\{\bar{\mathbf{s}}_{n,c}^g\}_1^{(\theta)}$  with  $\theta = 1, 2, \dots, \theta_S$ . The complete  $\theta_S$  symbol vectors can be generated directly by mapping the obtained particles  $\{\bar{\mathbf{s}}_{n,c}^g\}_1^{(\theta)}$  to the non-zero indices according to the AAP  $\mathcal{A}_{n,c}^g$ , which are given by  $\{\bar{\mathbf{x}}_{n,c}^g\}^{(\theta)}$ . Then, the most likely transmitted symbol vector of the  $n$ th subcarrier corresponding to the AAP  $\mathcal{A}_{n,c}^g$  can be calculated based on the signal model given in (18)

$$\hat{\bar{\mathbf{x}}}_{n,\mathcal{A}_{n,c}^g}^g = \arg \min_{\bar{\mathbf{x}} \in \{\bar{\mathbf{x}}_{n,c}^g\}^{(\theta)}} \left\| \bar{\mathbf{y}}_{n,NW}^g - \gamma \mathbf{\Omega}_n^g \bar{\mathbf{x}} \right\|^2, \quad (36)$$

and the corresponding decision metric is given by

$$p(\hat{\bar{\mathbf{x}}}_{n,\mathcal{A}_{n,c}^g}^g) = \left\| \bar{\mathbf{y}}_{n,NW}^g - \gamma \mathbf{\Omega}_n^g \hat{\bar{\mathbf{x}}}_{n,\mathcal{A}_{n,c}^g}^g \right\|^2. \quad (37)$$

The final SFAP estimate of the  $g$ th subblock can be calculated based on the metrics of most likely transmitted symbol vectors of the AAPs in  $\hat{\mathcal{X}}_{N_S,c}^g$

$$\hat{\mathcal{X}}^g = \arg \min_{\hat{\mathcal{X}}_{N_S,c}^g} \sum_{n=1}^{N_S} p(\hat{\bar{\mathbf{x}}}_{n,\mathcal{A}_{n,c}^g}^g), \quad (38)$$

where  $c = 1, 2, \dots, \theta_F$ . After this point, the obtained final SFAP can be expressed as

$$\hat{\mathcal{X}}^g = \left[ \mathcal{A}_1^g, \mathcal{A}_2^g, \dots, \mathcal{A}_{N_S}^g \right]. \quad (39)$$

Finally, the estimate of the transmitted symbol matrix can be obtained by concatenating the most likely symbol vectors of each subcarrier in (36) according to the final SFAP estimate  $\hat{\mathcal{X}}^g$ , which is given by

$$\hat{\mathbf{X}}^g = \left[ \hat{\bar{\mathbf{x}}}_{1,\mathcal{A}_1^g}^g, \hat{\bar{\mathbf{x}}}_{2,\mathcal{A}_2^g}^g, \dots, \hat{\bar{\mathbf{x}}}_{N_S,\mathcal{A}_{N_S}^g}^g \right]. \quad (40)$$

The computational complexity in terms of the FLOPs of the SMC detector can be obtained according to the above formulae. Since the repetitions of AAPs in the different obtained SFAPs are difficult to count, we omit it in the following computational complexity analysis. The computational complexity in terms of the FLOPs to get the final estimate of the transmitted symbol matrix per subcarrier is less than

$$\begin{aligned} &N_{\text{AAP}}(2N_{TA}N_T - N_{TA} + \theta_A + 1) \\ &+ N_T(2N_TN_R + N_R - 1) + \theta_F(2\theta_S N_T(N_T + 1)) \\ &+ \theta_F(2N_T(N_{TA} - 1) + \theta_S M((N_{TA})^2 + 3N_{TA})). \end{aligned} \quad (41)$$

From the above analysis, the computational complexity of a given system is affected by the parameters  $\theta_F$  and  $\theta_S$ , which

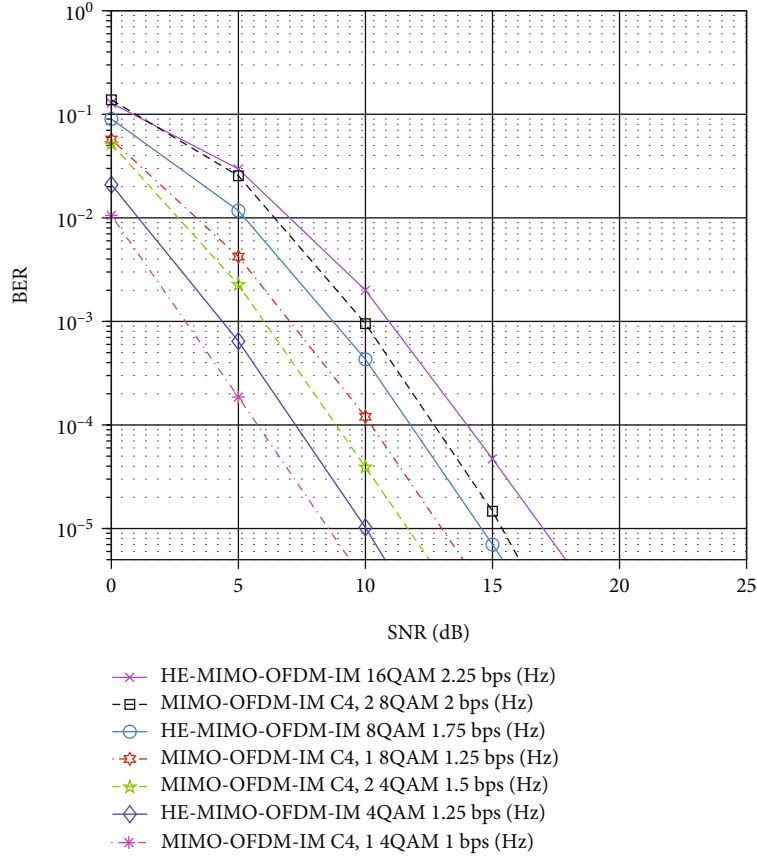


FIGURE 3: Error performance comparison of conventional MIMO-OFDM-IM and HE-MIMO-OFDM-IM with optimal ML detectors.

indicates that an interesting trade-off between the system error performance and demodulation complexity can be realized by adjusting these parameters.

#### 4. Simulation Results

In this section, we perform Monte Carlo simulations to verify the error performance of HE-MIMO-OFDM-IM with different parameters. The MIMO configuration in simulations is given by  $N_T = 4$  and  $N_R = 4$ . The length of the OFDM bloc is given by  $N = 128$ , and the length of CP is  $N_{CP} = 16$ . In HE-MIMO-OFDM-IM, the parameters of IM in each subblock are given by C4, 2. In the computer simulations, uncoded transmission is considered over frequency-selective Rayleigh fading MIMO channels, whose maximum delay spread is less than 12 sampling periods. It is assumed that the receiver estimates the perfect channel state information, which is unknown at the transmitter. Here, we use MMSE-LLR to denote the proposed low-complexity MMSE-LLR detector for MIMO-OFDM-IM in [44] and HSMC detector to denote the SMC detector proposed in [47], which can also be employed in the proposed HE-MIMO-OFDM-IM scheme.

Figure 3 compares the BER performance between conventional MIMO-OFDM-IM and HE-MIMO-OFDM-IM with constellations 4QAM, 8QAM, and 16QAM by employing the optimal ML detector. Since the computation complexity of the ML detector of MIMO-OFDM-IM with

constellation 16QAM is too high, the corresponding curve is not provided in Figure 3. It is observed in Figure 3 that HE-MIMO-OFDM-IM outperforms conventional MIMO-OFDM-IM with same constellation and IM parameters due to the transmit diversity achieved by the HE-MIMO-OFDM-IM scheme. According to the principle of IM,  $p_1$  index bits occupies only a small percentage in the  $p$  total bits because small size of subblock is needed to maintain low complexity in both modulation and demodulation. In particular, there are total  $p = 10$  bits per subblock in MIMO-OFDM-IM when 16QAM employed and the proposed scheme suffers only one bits loss per subblock, the percentage of loss decreases as the order of constellation increases because the loss of index bits do not change with the order of the constellation. By reducing the number of the activated subcarriers in the subblock, i.e.,  $K = 1$ , the proposed HE-MIMO-OFDM-IM outperforms conventional MIMO-OFDM-IM with the same SE. Note that, the computation complexity in the calculation of the most likely transmitted symbol vectors for each subcarrier does not change with the number of the activated subcarriers  $K$  in conventional MIMO-OFDM-IM. Combining with the demodulation complexity comparison in Figure 4, the proposed HE-MIMO-OFDM-IM achieves extremely low detection complexity when higher-order constellations are employed. Therefore, low detection complexity and gradually diminished SE loss in higher-order constellations applications



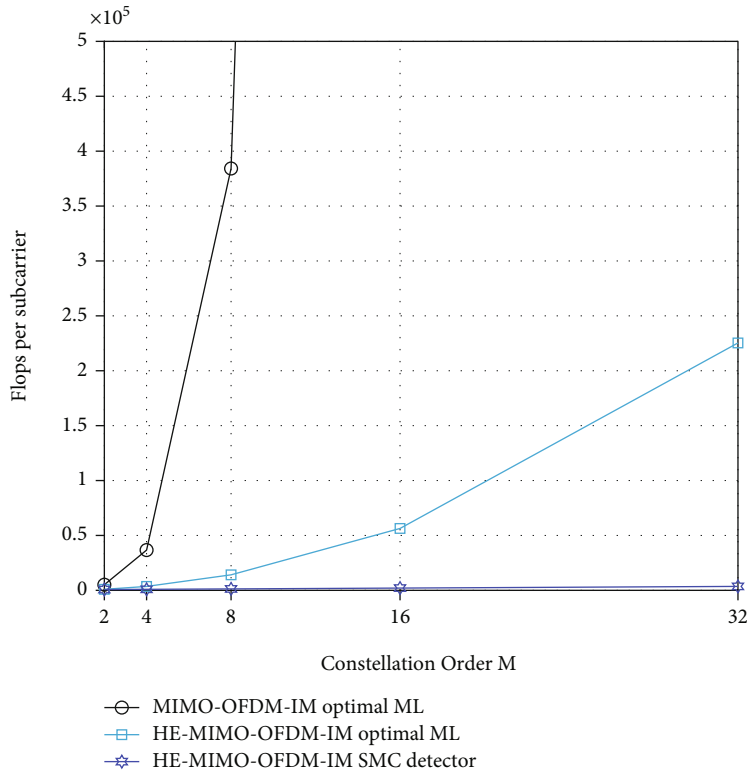


FIGURE 4: Flops comparison of different detectors for conventional MIMO-OFDM-IM and HE-MIMO-OFDM-IM.

make the proposed HE-MIMO-OFDM-IM becomes a competitive alternative to conventional MIMO-OFDM-OFDM.

Figure 4 compares the computational complexity in terms of the flops for the optimal ML detectors for both MIMO-OFDM-IM and HE-MIMO-OFDM-IM, the SMC detector with parameter  $\theta_A = \theta_F = 4$  and  $\theta_S = 5$  for HE-MIMO-OFDM-IM. In comparisons,  $4 \times 4$  MIMO configuration is considered. The OFDM-IM parameters in comparison are given by  $N_S = 4$  and  $K = 2$ . Due to there are some identical AAPs in  $\theta_F$  final SFAPs at the same subcarrier level, the computational complexity of the SMC detector is obtained by the computer simulations. As shown in Figure 4, the computational complexity of the optimal ML detector for HE-MIMO-OFDM-IM is less than that of MIMO-OFDM-IM, which validates that the proposed scheme has potential to meet the request of low-complexity implementation scenarios. This can be explained that the proposed HE-MIMO-OFDM-IM has only  $2^{N_{TA}}$  AAPs which have  $N_{TA}$  nonzero symbols, while MIMO-OFDM-IM has  $2^{N_T}$  AAPs in which the nonzero symbols range from 0 to  $N_T$ . The SMC detector can further reduce the demodulation complexity by exploiting the null space to calculate the most likely transmitted SFAPs and calculates the most likely transmitted symbol vectors by using the SMC algorithm in the second step. Furthermore, other low-complexity detection algorithms can be used to substitute the SMC algorithm after the null space calculations in the second step, which can further reduce the demodulation complexity for high-order MIMO configurations.

Figure 5 compares the BER performance of different detectors for HE-MIMO-OFDM-IM with 4QAM and 16QAM constellations. The number of the survival partial SFAPs of each subcarrier level in the proposed SMC detector is given by  $\theta_A = \theta_F = 4$  and the number of the particles in the SCM algorithm is given by  $\theta_S = 5$ . The number of the particles in the HSMC detector for 4QAM and 16QAM constellations are given by  $\theta_S = 6$  and  $\theta_S = 13$ , respectively, and the survival paths in the SD-like algorithm to get the estimate of the submatrix is given by  $\theta_{\text{path}} = 8$ . Because there are some illegal AAPs, a validity check of the AAPs in the calculation of the most likely transmitted symbol vectors is needed in the HSMC detector. The MMSE-LLR detector suffers from a significant error performance degradation compared with the optimal ML detector and the proposed SMC detector due to only one receive diversity order achieved. As shown in Figure 5, all the SMC detectors achieve near-optimal BER performance with the reduced computational complexity, which indicates that the SMC detectors achieve the diversity order  $N_R$  as that of the optimal ML detector. In particularly, the proposed SMC detector has potential to achieve almost the same BER performance as that of the optimal ML detector in the medium to high SNRs, which demonstrate that the null space method can effectively acquire almost the same SFAP estimate as that of the optimal ML detector if enough partial SFAPs are kept at each subcarrier level in the calculations. This phenomenon indicates that two-step detection algorithm, in which the first step uses the null space method to get the most

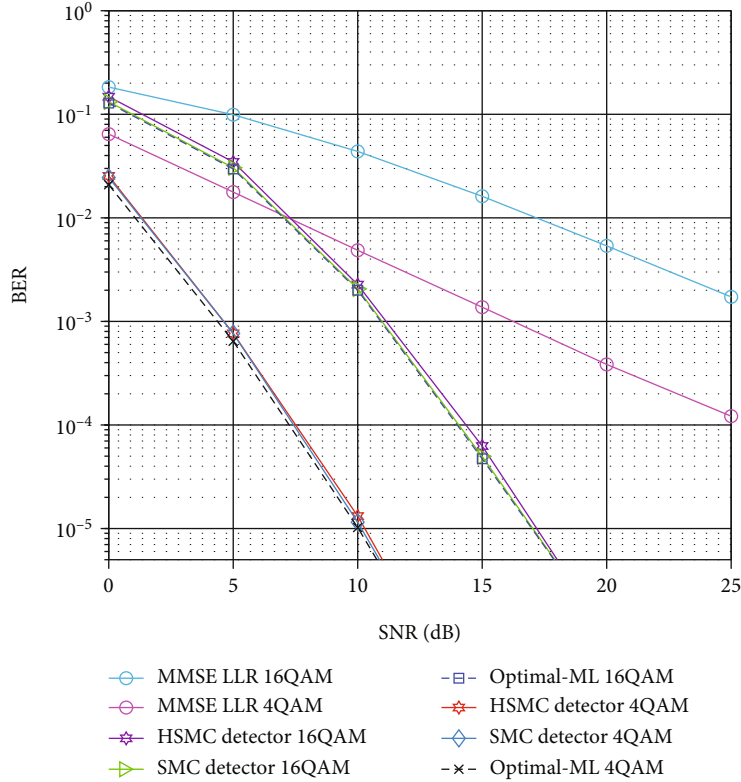


FIGURE 5: Error performance comparison of different detectors for HE-MIMO-OFDM-IM.

likely transmitted SFAPs and the second step detects the modulated symbol vectors based the results in the first step, can be a valuable method for the IM-aided MIMO systems. Furthermore, the SMC detector can achieve a trade-off between error performance, system SE, and demodulation complexity by adjusting the corresponding parameters to meet the requests of different scenarios.

## 5. Conclusion

In this paper, we have proposed a novel HE-MIMO-OFDM-IM scheme which performs IM in each transmit antenna subgroup to achieve two transmit diversity order of IM patterns and achieves extremely low computational complexity at the receiver side. Then, the subcarrier-wise ML detector, which is performed in two steps, is proposed for HE-MIMO-OFDM-IM. Due to the fixed number of the activated transmit antennas at each subcarrier level, the null space submatrix of the QR decomposition of the noise whitening preprocessed channel response matrix can be used to calculate the most likely transmitted IM patterns of each subblock. Then, a novel SMC detector is proposed to exploit the null space submatrix to calculate the most likely transmitted SFAPs first and then calculates the most likely transmitted symbol vectors of each subcarrier based on the SMC theory according to the obtained most likely transmitted SFAPs. The computer simulation results have shown that the proposed HE-MIMO-OFDM-IM achieves better BER performance under same constellation at cost of a minor

SE and achieves extremely low demodulation complexity compared with that of conventional MIMO-OFDM-IM. Consequently, the advantages in error performance and computational complexity makes the proposed HE-MIMO-OFDM-IM more favorable to the mMTC scenarios.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work was funded in part by the National Natural Science Foundation of China (grant numbers 62071504, 61971149, and 61971198), in part by the Joint Fund of National Natural Science Foundation of China and Zhejiang Province (grant number U1809211), in part by the Guangdong Basic and Applied Basic Research Foundation (grant number 2021A1515011657 and 2020A1515110958), in part by the Special Foundation in Key Fields for Colleges and Universities of Guangdong Province (New Generation of Information Technology) (grant number 2020ZDZX3025), and in part by the Shaanxi Provincial Natural Science Foundation (grant number 2021JQ-486), in part by the Key-Area Research and Development Program of Guangdong

Province (grant number 2020B1111010002). The authors would also like to acknowledge the support of the School of Electronic and Information Engineering, South China University of Technology.


## References

- [1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. S. Oland, and F. Tufvesson, "6G wireless systems: vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.
- [2] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.
- [3] E. Memisoglu, E. Basar, and H. Arslan, "Fading-aligned OFDM with index modulation for mMTC services," *Physical Communication*, vol. 35, p. 100680, 2019.
- [4] Q. Liu, S. Sun, H. Wang, and S. Zhang, "6G green IoT network: joint design of intelligent reflective surface and ambient backscatter communication," *Wireless Communications and Mobile Computing*, vol. 2021, 2031 pages, 2021.
- [5] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.
- [6] M. Wen, S. Lin, K. J. Kim, and F. Ji, "Cyclic delay diversity with index modulation for green internet of things," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 600–610, 2021.
- [7] R. S. Biljana, T. Kire, and D. Danco, "Internet of things framework for home care systems," *Wireless Communications and Mobile Computing*, vol. 2017, 10 pages, 2017.
- [8] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, "Towards enabling critical mMTC: a review of URLLC within mMTC," *IEEE Access*, vol. 8, pp. 131796–131813, 2020.
- [9] M. Wen, X. Cheng, and L. Yang, *Index Modulation for 5G Wireless Communications*, Springer International Publishing AG, Cham, Switzerland, 2017.
- [10] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.
- [11] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: a survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 315–348, 2019.
- [12] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 6–26, 2015.
- [13] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.
- [14] M. Wen, B. Zheng, K. J. Kim et al., "A survey on spatial modulation in emerging wireless systems: research progresses and applications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 1949–1972, 2019.
- [15] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1605–1615, 2012.
- [16] H. Qing, H. Yu, Y. Liu, and M. Wen, "Enhanced spatial modulation with generalized antenna selection in MISO channels," *IET Communications*, vol. 15, no. 16, pp. 2046–2053, 2021.
- [17] Q. Li, M. Wen, M. D. Renzo, H. V. Poor, S. Mumtaz, and F. Chen, "Dual-hop spatial modulation with a relay transmitting its own information," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4449–4463, 2020.
- [18] S. Lin, B. Zheng, F. Chen, F. Ji, and H. Yu, "Soft demodulators based on deterministic SMC for single-carrier GSM in broadband channels," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 1973–1985, 2019.
- [19] B. Zheng, M. Wen, F. Chen, N. Huang, F. Ji, and H. Yu, "The k-best sphere decoding for soft detection of generalized spatial modulation," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4803–4816, 2017.
- [20] B. Farhang-Boroujeny and H. Moradi, "OFDM inspired waveforms for 5G," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2474–2492, 2016.
- [21] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: channel estimation and reflection optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, 2020.
- [22] Y. Liu, F. Ji, M. Wen, D. Wan, and B. Zheng, "Vector OFDM with index modulation," *IEEE Access*, vol. 5, pp. 20135–20144, 2017.
- [23] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: channel estimation and training design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8315–8329, 2020.
- [24] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, 2013.
- [25] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4786–4801, 2017.
- [26] Q. Li, M. Wen, B. Clerckx, S. Mumtaz, A. Al-Dulaimi, and R. Q. Hu, "Subcarrier index modulation for future wireless networks: principles, applications, and challenges," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 118–125, 2020.
- [27] M. Wen, Q. Li, and X. Cheng, *Index Modulation for OFDM Communications Systems*, Springer International Publishing AG, Cham, Switzerland, 2021.
- [28] N. Ishikawa, S. Sugiura, and L. Hanzo, "Subcarrier-index modulation aided OFDM - will it work?," *IEEE Access*, vol. 4, pp. 2580–2593, 2016.
- [29] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the achievable rate of OFDM with index modulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1919–1932, 2016.
- [30] M. Wen, X. Cheng, L. Yang, Y. Li, X. Cheng, and F. Ji, "Index modulated OFDM for underwater acoustic communications," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 132–137, 2016.
- [31] Y. Liu, F. Ji, H. Yu, F. Chen, D. Wan, and B. Zheng, "Enhanced coordinate interleaved OFDM with index modulation," *IEEE Access*, vol. 5, pp. 27504–27513, 2017.
- [32] Q. Li, M. Wen, E. Basar, and F. Chen, "Index modulated OFDM spread spectrum," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2360–2374, 2018.

- [33] M. Wen, X. Chen, Q. Li, E. Basar, Y.-C. Wu, and W. Zhang, "Index modulation aided subcarrier mapping for dual-hop OFDM relaying," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6012–6024, 2019.
- [34] H. Qing, H. Yu, Y. Liu, W. Duan, M. Wen, and F. Ji, "Distributed cooperative OFDM-IM system," *China Communications*, vol. 17, no. 9, pp. 167–176, 2020.
- [35] T. Mao, Z. Wang, Q. Wang, S. Chen, and L. Hanzo, "Dual-mode index modulation aided OFDM," *IEEE Access*, vol. 5, pp. 50–60, 2017.
- [36] M. Wen, E. Basar, Q. Li, B. Zheng, and M. Zhang, "Multiple-mode orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3892–3906, 2017.
- [37] Q. Li, M. Wen, E. Basar, H. V. Poor, B. Zheng, and F. Chen, "Diversity enhancing multiple-mode OFDM with index modulation," *IEEE Transactions on Communications*, vol. 66, no. 8, pp. 3653–3666, 2018.
- [38] M. Wen, Q. Li, E. Basar, and W. Zhang, "Generalized multiple-mode OFDM with index modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6531–6543, 2018.
- [39] J. Zheng and R. Chen, "Achieving transmit diversity in OFDM-IM by utilizing multiple signal constellations," *IEEE Access*, vol. 5, pp. 8978–8988, 2017.
- [40] Z. Hu, J. Yang, P. Guo, and Q. Li, "Orthogonal frequency division multiplexing with cascade index modulation," *IET Communications*, vol. 16, no. 10, pp. 1057–1070, 2022.
- [41] Z. Hu, F. Chen, Y. Liu, S. Liu, H. Yu, and F. Ji, "Low-complexity detection for multiple-mode OFDM with index modulation," *Physical Communication*, vol. 34, pp. 38–47, 2019.
- [42] B. Zheng, F. Chen, M. Wen, F. Ji, H. Yu, and Y. Liu, "Low-complexity ML detector and performance analysis for OFDM with in-phase/quadrature index modulation," *IEEE Communications Letters*, vol. 19, no. 11, pp. 1893–1896, 2015.
- [43] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.
- [44] E. Basar, "Multiple-input multiple-output OFDM with index modulation," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2259–2263, 2015.
- [45] T. Datta, H. S. Eshwaraiah, and A. Chockalingam, "Generalized space-and-frequency index modulation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 4911–4924, 2016.
- [46] E. Basar, "On multiple-input multiple-output OFDM with index modulation for next generation wireless networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3868–3878, 2016.
- [47] Z. Hu, S. Lin, B. Zheng, F. Chen, Q. Wang, and Y. Wei, "Low complexity subcarrier-wise detection for MIMO-OFDM with index modulation," *IEEE Access*, vol. 5, pp. 23822–23832, 2017.
- [48] B. Zheng, M. Wen, E. Basar, and F. Chen, "Multiple-input multiple-output OFDM with index modulation: low-complexity detector design," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2758–2772, 2017.

## Research Article

# Chebyshev-Gauss Approximation Analysis for Mobile Edge Computing-Aided IoT Networks

Fusheng Zhu,<sup>1</sup> Liming Chen,<sup>2</sup> Wen Zhou,<sup>3</sup> Dan Deng,<sup>4</sup> Yajuan Tang,<sup>5</sup> Jun Liu ,<sup>6</sup> Yuwei Zhang,<sup>6</sup> Tao Cui,<sup>6</sup> Lin Zhang,<sup>6</sup> Jing Wang,<sup>6</sup> and Sun Li<sup>7</sup>

<sup>1</sup>Guangdong New Generation Communication and Network Innovative Institute (GDCNi), Guangzhou, China

<sup>2</sup>Electric Power Research Institute of CSG, Guangzhou, China

<sup>3</sup>Nanjing Forestry University, Nanjing, China

<sup>4</sup>University of Science and Technology of China, China

<sup>5</sup>Shantou University, Shantou, China

<sup>6</sup>Tsinghua University, Beijing, China

<sup>7</sup>Xi'an Jiaotong University, China

Correspondence should be addressed to Jun Liu; [junliu.thu@ieee.org](mailto:junliu.thu@ieee.org)

Received 5 May 2022; Revised 25 May 2022; Accepted 30 May 2022; Published 18 July 2022

Academic Editor: Jun Li

Copyright © 2022 Fusheng Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, mobile edge computing (MEC) has been widely applied into Internet of Things (IoT) networks, which has attracted a lot of attention from researchers. A critical challenge in the MEC-aided IoT networks is that the performance analysis is often complicated, where it is quite difficult for us to obtain some analytical or closed-form solution to the performance analysis, such as outage probability and bit error rate. This has been the bottleneck of the development of MEC-aided IoT networks. To address this challenge, we deeply investigate the Chebyshev-Gauss approximation method and derive the analytical solution to implement this powerful and useful approximation. We then give several examples to show the effectiveness of the Chebyshev-Gauss approximation in the performance analysis for the MEC-aided IoT systems. The results in this work can serve as an important reference and reveal some important inherent mechanisms for the MEC-aided IoT networks.

## 1. Introduction

Recently, a lot of wireless nodes cooperate together, to communicate and compute collaboratively, which form the Internet of Things (IoT) networks [1, 2]. In such a system, a lot of wireless nodes access the system spectrum, by using orthogonal or nonorthogonal multiple access schemes [3, 4]. These nodes can communicate and compute in a collaborative way, when facing some intensive calculating tasks. Besides the communication and calculation, the privacy protection also becomes a key research topic in the study of IoT networks [5, 6], where some privacy protection methods from the physical layer to the application layer should be incorporated into the system, in order to enhance the data communication privacy and

data calculation privacy, especially for some sensitive data such as medical data and financial data.

Some novel techniques have been proposed by researchers to promote the development of IoT networks, among which mobile edge computing (MEC) is a key technology [7, 8]. In the MEC-aided IoT networks, some edge nodes have some powerful ability to help calculate the intensive tasks from other nodes, which will be helpful in leading to a smaller delay and power consumption (PoCo). In this area, a lot of studies have been performed to utilize the communication resources as well as calculating resources in the MEC-aided IoT networks, through some conventional optimization methods such as convex optimization or some intelligent algorithms such as deep reinforcement learning (DRL) algorithms, in order to reduce the system delay and

PoCo [9]. This can help make the MEC-aided IoT networks fit the various applications [10–12].

Because of restricted regularity sources, cochannel interference has actually been unavoidable in the wireless communication systems. Cochannel interference has restricted the system efficiency seriously, as well as being the traffic jam of the wireless systems. The effect of cochannel interference on the system efficiency of wireless communications was thoroughly examined in the literary works. Some authors examined the dual-hop communicating systems in the existence of cochannel interference, as well as the system data rate and outage possibility. For the secure communicate systems in cochannel interference, the system efficiency might be examined through obtaining the capacity as well as asymptotic privacy outage possibility, whereby the impact of interfering energy on the system efficiency might be exposed.

A critical challenge in the MEC-aided IoT networks is that the performance analysis is often complicated, where it is quite difficult to obtain some analytical or closed-form solution to the performance analysis, such as outage probability and bit error rate. This has been the bottleneck of the development of MEC-aided IoT networks. To address this challenge, we deeply investigate the Chebyshev-Gauss approximation method and derive the analytical solution to implement this powerful and useful approximation. We then give several examples to show the effectiveness of the Chebyshev-Gauss approximation in the performance analysis for the MEC-aided IoT systems. The results in this work can serve as an important reference and reveal some important inherent mechanisms for the MEC-aided IoT networks.

## 2. Chebyshev-Gauss Quadrature

In numerical analysis, numerical integration is the method and theory of calculating the value of definite integration. We can use the Leibniz integral rule to calculate the definite integral through the original function. However, it is regularly difficult to calculate the original value of the function. There are few functions that can be expressed by elementary functions, and the integration of most integrable functions cannot be expressed by elementary functions or even analytical expressions. Therefore, in many cases, we can only use numerical integration to calculate the approximate value of the function.

At present, there are many algorithms for calculating definite integral. For instance, these algorithms mainly include the following:

- (i) Rectangle rule
- (ii) Trapezoidal rule
- (iii) Romberg's method
- (iv) Gauss quadrature

Among the above algorithms, the Gaussian quadrature rule with  $n$ -point is a quadrature rule constructed to yield an exact result for polynomials of degree  $2n - 1$  or less by

a suitable choice of the nodes  $x_i$  and weights  $w_i$  for  $i = 1, \dots, n$ . Specifically, the Gauss quadrature includes three different forms:

- (1) Chebyshev-Gauss quadrature
- (2) Gauss-Hermite quadrature
- (3) Gauss-Jacobi quadrature

Now, we mainly discuss the Chebyshev-Gauss quadrature. Chebyshev-Gauss quadrature is an extension of Gaussian quadrature method, which is used to approximate the following two types of integral value. For the first kind, we can have

$$\int_{-1}^{+1} \frac{f(x)}{\sqrt{1-x^2}} dx \approx \sum_{i=1}^n w_i f(x_i), \quad (1)$$

where  $x_i = \cos((2i-1)\pi/2n)$ , the weight  $w_i = \pi/n$ , and the approximation error decreases with a larger number of item  $n$ .

On the contrary, for the second kind, we have

$$\int_{-1}^{+1} \sqrt{1-x^2} g(x) dx \approx \sum_{i=1}^n w_i g(x_i), \quad (2)$$

where  $x_i = \cos(i\pi/(n+1))$ , the weight  $w_i = (\pi/(n+1)) \sin^2(i\pi/(n+1))$ , and the approximation error decreases with a larger number of item  $n$ .

For a random function  $f(x)$ , its integral  $\int_a^b f(x) dx$  can be approximated as

$$\int_a^b f(x) dx \approx \sum_{i=1}^n w_i f(x_i), \quad (3)$$

where  $w_i$  is the weight coefficient, and the approximation error decreases with a larger number of item  $n$ . Note that the accuracy can be improved by increasing the number of  $x_i$  or finding the right  $x_i$ . We can rewrite  $f(x)$  as

$$f(x) = \rho(x)g(x). \quad (4)$$

Therefore, (3) can be approximated as

$$\int_a^b \rho(x)g(x) dx \approx \sum_{i=1}^n w_i g(x_i), \quad (5)$$

where  $\rho(x)$  is the weight function and the approximation error decreases with a larger number of item  $n$ . We can use the Chebyshev-Gauss quadrature to approximate (5). The Chebyshev polynomials of the first kind are obtained

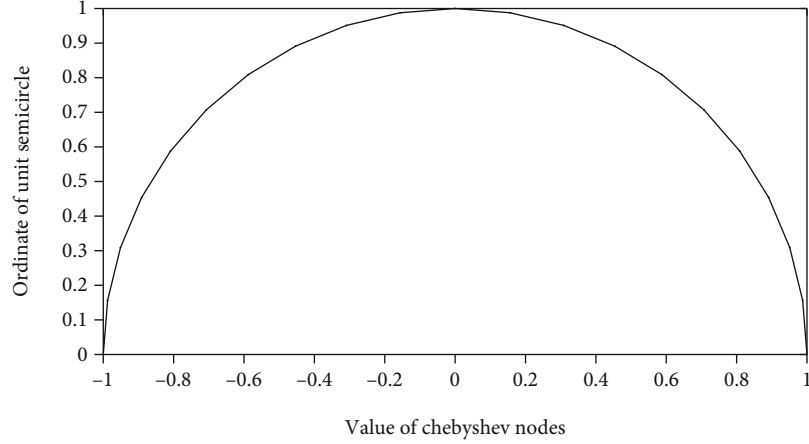


FIGURE 1: Chebyshev node on the unit semicircle.

from the following recurrence relation:

$$\begin{aligned}
 \cos 2\theta &= 2 \cos^2\theta - 1, \\
 \cos 3\theta &= 4 \cos^2\theta - 3 \cos \theta, \\
 \cos n\theta &= T_n(\cos \theta), \\
 T_n(x) &= \cos (n \cdot \arccos x).
 \end{aligned} \tag{6}$$

Let  $T_n(x) = 0$ , and the root of the equation can be obtained as

$$\begin{aligned}
 T_n(x) &= \cos (n \cdot \arccos x) = 0, \\
 n \cdot \arccos x &= \frac{\pi}{2} + k\pi = \frac{(2k+1)\pi}{2}, \\
 \arccos x &= \frac{(2k+1)\pi}{2n}, \\
 x_k &= \cos \left( \frac{(2k+1)\pi}{2n} \right),
 \end{aligned} \tag{7}$$

where  $k=0, 1, \dots, n-1$ , and  $x_k$  is the Chebyshev node, namely, the root of the Chebyshev polynomials of the first kind.

As shown in Figure 1, the Chebyshev node is equivalent to the  $x$ -axis coordinates of  $N$  equally spaced points on the unit semicircle. The Chebyshev-Gauss quadrature can obtain a relatively approximate solution only when the function  $f(x)$  can be approximated by polynomials in the interval  $[-1, 1]$ . We use an affine transformation for nodes over an arbitrary interval  $[a, b]$  as

$$x_k = \frac{1}{2}(a+b) + \frac{1}{2}(b-a) \cos \left( \frac{(2k-1)\pi}{2n} \right), \quad k = 1, 2, \dots, n. \tag{8}$$

For (3), let

$$x = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)y, \tag{9}$$

and thus, we have

$$\int_a^b f(x) dx = \int_{-1}^1 \frac{1}{2}(b-a) \cdot f \left( \frac{1}{2}(a+b) + \frac{1}{2}(b-a)y \right) dy. \tag{10}$$

Let  $y = \cos \theta$ , where  $\theta = ((2k+1)\pi/2n)$ ,  $k=0, 1, \dots, n-1$ , and then, we have

$$\begin{aligned}
 \int_{-1}^1 \frac{1}{2}(b-a) \cdot f \left( \frac{1}{2}(a+b) + \frac{1}{2}(b-a)y \right) dy \\
 = - \int_{\pi}^0 \frac{1}{2}(b-a) \sin \theta \cdot f \left( \frac{1}{2}(a+b) + \frac{1}{2}(b-a) \cos \theta \right) d\theta.
 \end{aligned} \tag{11}$$

We can further write

$$\int_a^b f(x) dx = \int_0^{\pi} \frac{1}{2}(b-a) \sin \theta \cdot f \left( \frac{1}{2}(a+b) + \frac{1}{2}(b-a) \cos \theta \right) d\theta. \tag{12}$$

According to the definition of definite integral, we can write  $\int_a^b f(x) dx$  as

$$\begin{aligned}
 \int_a^b f(x) dx &= \sum_{k=0}^{n-1} \frac{1}{2}(b-a) \sin \left( \frac{(2k+1)\pi}{2n} \right) \frac{\pi}{n} \\
 &\cdot f \left( \frac{1}{2}(a+b) + \frac{1}{2}(b-a) \cos \left( \frac{(2k+1)\pi}{2n} \right) \right).
 \end{aligned} \tag{13}$$

By comparing with (5), we can obtain

$$x_i = \frac{1}{2}(a+b) + \frac{1}{2}(b-a) \cos \left( \frac{(2k+1)\pi}{2n} \right), \tag{14}$$

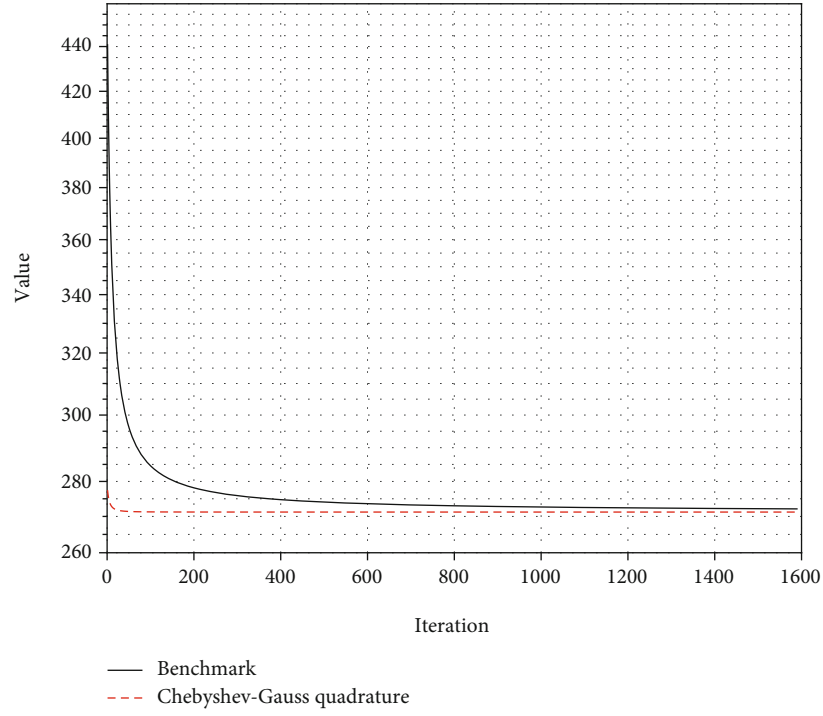


FIGURE 2: Comparison of the convergence effect of  $I_1$ .

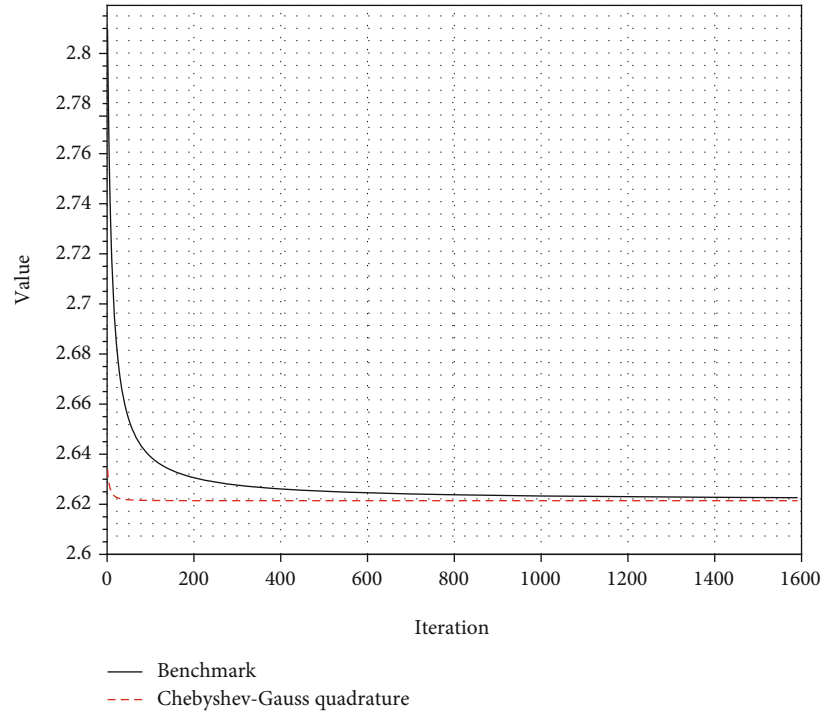


FIGURE 3: Comparison of the convergence effect of  $I_2$ .

where  $k = 1, 2, \dots, n - 1$ , and

$$w_i = \frac{1}{2} (b - a) \sin \left( \frac{(2k + 1)\pi}{2n} \right) \frac{\pi}{n}, \quad (15)$$

### 3. Numerical and Simulation Results

In this part, we present some numerical and simulation results to verify the convergence effect of the conventional rectangle rule and the Chebyshev-Gauss quadrature. We compare the convergence effect of the following three



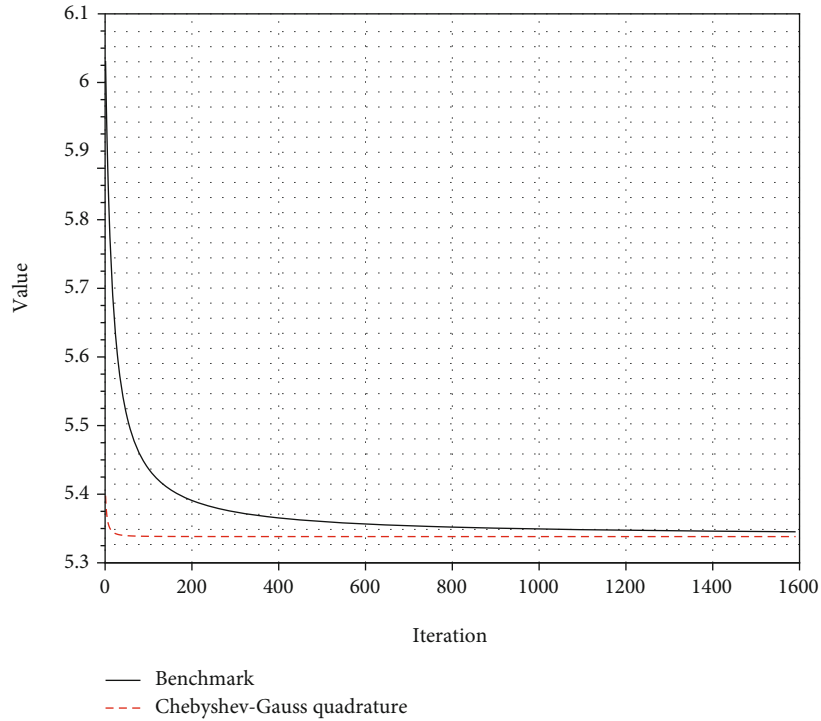


FIGURE 4: Comparison of the convergence effects of  $I_3$ .

functions, in Figures 2–4 and Tables 1–3.

$$\begin{aligned}
 I_1 &= \int_1^3 e^{x^2-x^{1/2}} dx, \\
 I_2 &= \int_1^3 \log(e^x - x^2) dx, \\
 I_3 &= \int_1^3 \frac{x^{e^x-x}}{x^3} dx.
 \end{aligned} \tag{16}$$

Figure 2 shows the convergence effect of  $I_1$ . We can find that the Chebyshev-Gauss quadrature and the benchmark method (namely the rectangle rule) both reach the convergent state after 1600 iterations, and the ultimate convergent values are 271.2296 and 272.1258, respectively. Similarly, as shown in Table 1, the Chebyshev-Gauss quadrature method approaches convergence with the convergent value of 271.2469, after about 200 iterations, which is much faster than the benchmark method which approaches the convergent state, after about 1200 iterations.

In Figure 3, the Chebyshev-Gauss quadrature approximation is performed on  $I_2$  to demonstrate the advantage of faster convergence than the benchmark method. From Table 2, we can also find that the Chebyshev-Gauss quadrature method approaches the convergent state with the convergent value of 2.6207, after about 200 iterations, while the benchmark method approaches the convergent state, after about 1200 iterations. The associated ultimate convergent values are 2.6207 and 2.6218 after 1600 iterations, respectively.

TABLE 1: Approximation of  $I_1$  versus the number of iterations.

Iteration	Benchmark method	Chebyshev-Gauss
1	441.1021	277.4677
200	278.1464	271.2469
400	247.7484	271.2329
600	273.5892	271.2310
800	273.0044	271.2303
1000	272.6519	271.2300
1200	272.4162	271.2298
1400	272.2475	271.2297
1600	272.1258	271.2296

TABLE 2: Approximation of  $I_2$  versus the number of iterations.

Iteration	Benchmark method	Chebyshev-Gauss
1	2.8100	2.6329
200	2.6296	2.6207
400	2.6252	2.6207
600	2.6237	2.6207
800	2.6230	2.6207
1000	2.6225	2.6207
1200	2.6222	2.6207
1400	2.6220	2.6207
1600	2.6218	2.6207

Figure 4 illustrates the convergence of the function  $I_3$ . It can be seen that the Chebyshev-Gauss quadrature method approaches the convergent state with the convergent value

TABLE 3: Approximation of  $I_3$  versus the number of iterations.

Iteration	Benchmark method	Chebyshev-Gauss
1	6.026	5.3915
200	5.38493	5.3354
400	5.3608	5.3353
600	5.3525	5.3353
800	5.348	5.3353
1000	5.3457	5.3353
1200	5.3440	5.3353
1400	5.3428	5.3353
1600	5.3419	5.3353

of 5.3354 after about 200 iterations, while the benchmark method approaches the convergent state with the convergence value of 5.3428 after about 1400 iterations, as shown in Table 3. This also shows the advantage that the Chebyshev-Gauss quadrature method converges much faster than the benchmark method. The associated final convergent values are 5.3353 and 2.6218 after 1000 iterations, respectively.

#### 4. Conclusions

In the MEC-aided IoT networks, a critical challenge was that the performance analysis was often complicated, where it was quite difficult to obtain some analytical or closed-form solution to the performance analysis, such as outage probability and bit error rate. To address this challenge, we deeply investigated the Chebyshev-Gauss approximation method and derived the analytical solution to implement this powerful and useful approximation. We then gave several examples to show the effectiveness of the Chebyshev-Gauss approximation in the performance analysis for the MEC-aided IoT systems. The results in this work could serve as an important reference and reveal some important inherent mechanisms for the MEC-aided IoT networks.

#### Data Availability

The data of this paper can be obtained through email to the authors.

#### Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this work.

#### Acknowledgments

The work in this paper was supported by the Key-Area Research and Development Program of Guangdong Province (No. 2019B090904014) and the National Natural Science Foundation of China (No. 62871349).

#### References

- [1] S. Arzykulov, A. Celik, G. Nauryzbayev, and A. M. Eltawil, "UAV-Assisted cooperative & cognitive NOMA: deployment, clustering, and resource allocation," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 263–281, 2022.
- [2] Q. Tao, J. Wang, and C. Zhong, "Performance analysis of intelligent reflecting surface aided communication systems," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2464–2468, 2020.
- [3] M. Liu, B. Li, Y. Chen et al., "Location parameter estimation of moving aerial target in space-air-ground-integrated networks-based IoV," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5696–5707, 2022.
- [4] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3393–3406, 2018.
- [5] X. Hu, C. Zhong, Y. Zhang, X. Chen, and Z. Zhang, "Location information aided multiple intelligent reflecting surface systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7948–7962, 2020.
- [6] M. Liu, C. Liu, M. Li, Y. Chen, S. Zheng, and N. Zhao, "Intelligent passive detection of aerial target in space-air-ground integrated networks," *China Communications*, vol. 19, no. 1, pp. 52–63, 2022.
- [7] W. Zhou, L. Chen, S. Tang et al., "Offloading strategy with PSO for mobile edge computing based on cache mechanism," *Cluster Computing*, vol. 2021, no. 1, pp. 1–11, 2021.
- [8] J. Zhao, X. Sun, Q. Li, and X. Ma, "Edge caching and computation management for real-time internet of vehicles: an online and distributed approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2183–2197, 2021.
- [9] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.
- [10] Z. Zhu, S. Wan, P. Fan, and K. B. Letaief, "Federated multi-agent actor-critic learning for age sensitive mobile-edge computing," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1053–1067, 2022.
- [11] S. Premkumar and A. N. Sigappi, "ASIS edge computing model to determine the communication protocols for IoT based irrigation," *Journal of Mobile Multimedia*, vol. 18, no. 3, 2022.
- [12] F. Alqahtani, M. Al-Maitah, and O. A. Elshakankiry, "A proactive caching and offloading technique using machine learning for mobile edge computing users," *Computer Communications*, vol. 181, pp. 224–235, 2022.