IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 3, 2022, accepted March 19, 2022, date of publication March 28, 2022, date of current version April 6, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3162875

On the Energy Efficiency Maximization of NOMA-Aided Downlink Networks With Dynamic User Pairing

KHA-HUNG NGUYEN[1], HIEU V. NGUYEN[1,2], (Member, IEEE), MAI T. P. LE[2], LUCA SANGUINETTI[3], (Senior Member, IEEE), AND OH-SOON SHIN[1], (Member, IEEE)

[1]Department of ICMC Convergence Technology, School of Electronic Engineering, Soongsil University, Seoul 06978, South Korea
[2]Faculty of Electronics and Telecommunication Engineering, The University of Danang - University of Science and Technology, Da Nang 5000, Vietnam
[3]Dipartimento di Ingegneria dell' Informazione, University of Pisa, 56122 Pisa, Italy

Corresponding author: Oh-Soon Shin (osshin@ssu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2019R1A2C1084834 and Grant 2017R1A5A1015596, in part by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab Project (Departments of Excellence), and in part by the Ministry of Education and Training under Project B2020-DNA-06.

ABSTRACT This study investigates a combined system comprising non-orthogonal multiple access (NOMA) and beamforming in a downlink network. To fully exploit the advantages of NOMA, user (UE) pairing and beamforming design are jointly optimized via a generalized model for UE association, subject to energy efficiency maximization. Owing to the combination of binary variables and nonconvex constraints, the resulting optimization problem belongs to the class of mixed-integer nonconvex programming. An innovative algorithm, integrating the inner-approximation and Dinkelbach methods, is proposed herein to address a nonconvex fractional function. By introducing a pairing matrix and relaxing the binary variables into continuous ones, our approach is capable of reaching an optimal solution, where two arbitrary UEs are optimally paired regardless of geographical or spatial constraints. For practical scenarios, we further propose a robust design to manage the effect of channel estimation errors under settings involving channel uncertainty. Numerical results show that our proposed designs, even with the presence of the imperfect channel state information at the base station, significantly outperform the conventional beamforming and existing pairing schemes.

INDEX TERMS Beamforming, non-orthogonal multiple access (NOMA), convex optimization, user pairing, energy efficiency, robust design.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has garnered significant attention from researchers for potential use in future wireless networks, i.e., the sixth generation (6G) wireless networks, owing to its capability of achieving superior system throughput and its compatibility with other systems [1], [2]. Unlike conventional orthogonal multiple access (OMA), NOMA allows users (UEs) to share the same time-frequency resource for data transmission. To remove the inter-user interference, NOMA exploits either power, code or spatial domain, corresponding to the three typical approaches of NOMA: power-domain NOMA [2], [3], code-domain NOMA [4], [5], and spatial-domain NOMA (i.e., multiple-input multiple-output (MIMO), massive MIMO [6]). Herein, we focus on power-domain NOMA,[1] wherein UE signals can be superimposed with different power allocation coefficients. In particular, higher power is allocated to UEs with poorer channel gains to perform reliable detection at the receiver. Meanwhile, UEs with better channel gains can suppress the signals of other UEs with poorer channel gains through successive interference cancellation (SIC) techniques before detecting their own signals [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Ding Xu.

[1]Power-domain NOMA is shortly referred to as NOMA in this paper.

A. RELATED WORKS AND MOTIVATION

To fully exploit the advantages of NOMA, the UEs must be grouped/paired efficiently. Therefore, the issue of clustering/pairing UEs has been widely investigated in recent NOMA-related studies. The majority of papers investigated the UE pairing strategy, where each pair consists of exactly two UEs, based on different criteria and under different contexts [8]–[15]. For example, the effect of UE pairing was investigated for a cognitive radio-inspired NOMA (CR-NOMA) system in [8], [9]. Each pair, which includes a primary UE with a poor channel condition and a cognitive UE with a strong channel condition, may share the same spectrum, but different pairs are designed to transmit onto orthogonal channels. In particular, UEs are paired based on the matching theory to improve the system throughput, while satisfying the minimum UE rate requirement [9]. The authors of [10] studied a centralized radio access network (C-RAN) utilizing NOMA in multiple sub-carriers (MSC), where UE pairing and SIC ordering algorithms were proposed to decrease the co-channel interference. For the NOMA-based massive MIMO scenario, the work [11] proposed a joint algorithm considering both UE pairing and pair scheduling for data transmission, wherein the UE pairing is conducted by grouping UEs with similar channel conditions based on their channel norms. The UE pairing strategy for a NOMA broadcast wireless network was considered in [12], [13], in which UE pairs are assigned to different sub-channels. The approach of [12] aims to pair two UEs with the best and worst channel gains in the unpaired UE set until the set is empty, whereas the approach of [13] pairs the two unpaired UEs with the worst channel gains consecutively. In addition, a channel gain based pairing scheme was proposed for a multiple-input single-output (MISO) MSC NOMA system in [14], wherein two UEs with the best and middle channel gains are paired to employ NOMA technique. On the other hand, the distance between UEs and the base station (BS) was used as a criterion of the UE pairing method in [15], [16], where each pair consists of two UEs in one of two disjoint zones, i.e., inner and outer zones.

It is worth to mention that the works in [17]–[21] allocate more than two UEs per NOMA group/cluster instead of making use of UE pairing. In particular, UE clustering algorithms are used to further exploit NOMA benefits for a full-duplex system in [17], for a NOMA-aided massive MIMO systems in [19], [20], and for a NOMA-aided mmWave system in [21]. However, the use of NOMA with a group comprising more than two UEs is typically more challenging to implement than UE pairing algorithms. This is because the virtual partition of cell zones in UE clustering relies on specific practical conditions, such as the cell size, channel conditions, and received power threshold [18], [22].

The beamforming design issue in a NOMA-aided MISO network has received limited attention in the literature [10], [23]–[25]. Specifically, the minorization-maximization algorithm was used for the SE optimization problem in a MISO-NOMA system in [24]. On the other hand, the worst-case achievable sum rate of all users was maximized with a robust beamforming design in [25]. However, the NOMA protocol in those works mainly rely on the SIC ordering without taking into account UE grouping. This motivates us to formulate a generalized multiuser MISO-NOMA framework that combines beamforming and UE pairing.

Since the energy efficiency (EE) is considered as a core criterion for future green cellular networks, the EE performance of NOMA-aided systems have recently attracted attention from the wireless community, e.g., [10], [17], [23], [26], [27]. One of the earliest studies pertaining to EE maximization for a NOMA system was presented in [26], where a suboptimal power allocation scheme was proposed with the assumption of statistical channel state information (CSI) at the transmitter. The authors of [27] investigated a power allocation scheme for EE optimization in a DL NOMA broadcast network. As mentioned above, EE maximization was investigated in [17] via joint sub-channel assignment and power allocation, but for a DL NOMA-aided network using a conventional single-input single-output framework rather than an MU-MISO framework. Besides, a multi-objective optimization problem of both the SE and EE was studied for a downlink (DL) NOMA system in [23]. However, the NOMA protocol in this paper is applied to all UEs without adopting any clustering method, i.e., each UE utilizes SIC to remove the interference caused by the signals of all weaker UEs. This approach is thus difficult to implement due to high complexity and expensive computational cost.

For a NOMA-aided C-RAN network, although the EE maximization problem was investigated via joint sub-carrier assignment and user pairing, as in [10], the QoS requirement for each UE is still left untreated. Meanwhile, a hybrid zero-forcing (ZF) beamforming design was proposed for an MU-MISO system in [28]. Notice that, the ZF beamformer proposed in [28] could only remove the residual interference completely, given that sufficient degree of freedom (DoF) for the ZF scheme is provided, i.e., the number of BS antennas must be at least larger than the number of UE clusters.

To overcome the above DoF limitations of the ZF beamforming and the channel gain restrictions of the greedy-based pairing schemes (as mentioned in previous works, e.g., [11]–[14], [28]), we aim to develop a novel approach that not only provides an optimal solution for UE pairing with a hybrid design of NOMA and beamforming, but achieves a balanced tradeoff between EE performance and complexity. Although there is a plethora of related works on the integration of NOMA and beamforming techniques, we stress that our proposed UE pairing strategy is different from the previous schemes in that it works in more dynamic fashion, where two arbitrary UEs could be optimally paired regardless of geographical or spatial constraints. In addition, by characterizing the UE paring variables with an auxiliary matrix, we reach a general framework for all UE pairing methods. The contributions of this paper are fully described in the following.

B. CONTRIBUTIONS

In this study, we investigate a combined beamforming and UE pairing scheme for a DL NOMA MU-MISO network. Instead of using channel-gain difference for UE pairing as the conventional UE pairing strategy, we propose a novel method to fully exploit the association among UEs for the efficient use of NOMA. The jointly optimized beamforming design for a NOMA-based system results in a mixed-integer nonconvex fractional problem, which cannot be solved using previous approaches. Hence, a combination of relaxation and inner-approximation (IA) methods overlaying the Dinkelbach framework is proposed to provide at least a local optimal solution. Furthermore, the proposed method can be conveniently extended to a robust design if imperfect CSI is available.

Overall, main contributions of this paper are listed as follows:

- First, we develop a novel hybrid NOMA-beamforming design, wherein the NOMA beamforming and conventional beamforming methods are simultaneously conceived to exploit the advantages of both technologies. When the NOMA beamforming performance deteriorates, the proposed design automatically switches to the conventional counterpart. Furthermore, the proposed design offers an optimal approach to utilize the entire spectral resource instead of assigning UEs to subchannels.
- Second, we implement the UE pairing strategy in a dynamic UE selection manner, where two arbitrary UEs can be optimally paired in a group, rather than greedy UE pairing based on channel conditions, i.e., channel correlation or geometrical distance, as in [17], [18], [22], [28].
- Third, we feature the binary variables for UE pairing using an upper triangle matrix. This not only reduces the number of binary variables and constraints, but may also yield a generalized framework that is applicable to all UE pairing methods. However, the EE maximization with NOMA-aided beamforming designs under a dynamic UE pairing can be classified as a mixed-integer nonconvex fractional problem, which is highly nontrivial to solve directly. Hence, we transform the problem into a more tractable form by relaxing the binary variables into continuous ones. To address the nonconvexity of the relaxed problem, we jointly employ the IA and Dinkelbach transformation to attain successive convex programming. In the end, a low-complexity iterative algorithm is devised to obtain at least a local optimal point for the original problem.
- Fourth, to achieve a robust design, we further extend the proposed algorithm for EE maximization under channel uncertainty, i.e., with imperfect CSI. In this regard, we exploit a statistical model for estimation error as an additional condition in the original problem formulation.

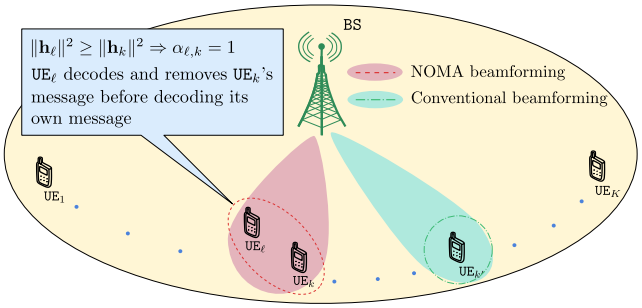


FIGURE 1. An illustration of the hybrid NOMA and conventional beamforming.

- Finally, numerical results are compared with existing schemes in terms of EE performance and complexity to verify the effectiveness of our proposed design. Furthermore, we demonstrate that the proposed method provides a fast convergence speed while offering robustness under an imperfect CSI assumption.

C. PAPER ORGANIZATION AND NOTATION

The remainder of this paper is organized as follows. Sections II and III describe the system model and problem formulation of EE maximization, respectively. In Section IV, we propose a low-complexity iterative algorithm for EE maximization using the IA method and Dinkelbach transformation. Section V presents a robust design under channel uncertainty. Numerical results are provided in Section VI, and conclusions are presented in Section VII.

Notation: Uppercase and lowercase letters in bold represent matrices and vectors, respectively. $\mathbf{x}^H$ denotes the Hermitian transpose of $\mathbf{x}$. $\mathbb{C}$ and $\mathbb{R}$ are the spaces of all complex and real numbers, respectively. $\|\cdot\|$ is the Euclidean norm. $|x|$ and $|\mathbf{X}|$ denote the absolute value of a number x and the determinant value of a matrix $\mathbf{X}$, respectively. $\mathbb{E}\{\cdot\}$ represents the expectation and $\Re\{.\}$ is the real part of a complex number.

II. SYSTEM MODEL

A. DATA TRANSMISSION MODEL

We herein consider a DL NOMA network, where a BS equipped with N antennas is located in the center of a cell and serves K single-antenna DL UEs uniformly distributed in the cell, as shown in Fig. 1. The k-th UE, denoted by UE_k, $k \in \mathcal{K} \triangleq \{1, 2, \ldots, K\}$, is associated with a channel gain vector $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ from the BS. For a NOMA pair, the SIC technique is applied at the stronger UE, i.e., UE with the stronger channel gain, who decodes and then subtracts the message of the weaker UE from the received signal before decoding its own message. To conveniently employ the NOMA technique, we first sort the channel gains of UEs in descending sequence, then assign the indices for UEs based on this order such that

the following condition is satisfied[2]:

$$\|\mathbf{h}_\ell\|^2 \geq \|\mathbf{h}_k\|^2, \quad \forall k, \ell \in \mathcal{K}, \ell \leq k. \tag{1}$$

The signal received at UE_k can be expressed as

$$y_k = \sum_{\forall \ell \in \mathcal{K}} \mathbf{h}_k^H \mathbf{w}_\ell x_\ell + n_k, \tag{2}$$

where $\mathbf{w}_k \in \mathbb{C}^{N\times 1}$ is the beamforming vector from the BS to UE_k. x_k is the data signal to be transmitted to UE_k, where $\mathbb{E}\{|x_k|^2\} = 1$ is assumed. $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ represents the additive white Gaussian noise. We begin by assuming that perfect CSI is available at the BS; subsequently, a robust design under channel uncertainty is presented based on the proposed solution.

For decoding operations, the UEs are dynamically paired prior to the messages sent from the BS. We introduce a new binary variable $\alpha_{\ell,k} \in \{0, 1\}$, with $\forall \ell, k \in \mathcal{K}$, to perform UE pairing, described as

$$\alpha_{\ell,k} = \begin{cases} 1, & \text{if } \mathrm{UE}_\ell \text{ and } \mathrm{UE}_k \text{ are paired with SIC being} \\ & \text{executed at } \mathrm{UE}_\ell, \text{ i.e., } \mathrm{UE}_\ell \text{ removes the} \\ & \text{message intended to } \mathrm{UE}_k \text{ before} \\ & \text{decoding its own message,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

For notational convenience, we define a matrix $\boldsymbol{\alpha} \triangleq [\alpha_{\ell,k}]_{\ell,k\in\mathcal{K}} \in \{0, 1\}^{K\times K}$. It is clear that when $\alpha_{\ell,k} = 1$, the other users are unpaired with UE_ℓ and UE_k. This implies that all other elements on two columns ℓ and k and two rows ℓ and k must be equal to zero.

By setting $\mathbf{w} = \left[\mathbf{w}_1^T \mathbf{w}_2^T \cdots \mathbf{w}_K^T\right] \in \mathbb{C}^{NK\times 1}$, the signal-to-interference-plus-noise (SINR) for decoding the UE_k message can be expressed as

$$\mathrm{SINR}_k(\mathbf{w}, \boldsymbol{\alpha}) = \min\left\{\mathrm{SINR}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}), \min_{\ell\in\mathcal{K}\setminus\{k\}} \{\mathrm{SINR}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})\}\right\}, \tag{4}$$

where $\mathrm{SINR}_{k,k}(\mathbf{w}, \boldsymbol{\alpha})$ and $\mathrm{SINR}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ are the corresponding SINRs for decoding UE_k's message at UE_k itself and at UE_ℓ, which are defined as

$$\mathrm{SINR}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\Psi_k(\mathbf{w}, \boldsymbol{\alpha})}, \tag{5a}$$

$$\mathrm{SINR}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) = \frac{|\mathbf{h}_\ell^H \mathbf{w}_k|^2}{\alpha_{\ell,k}\Theta_{\ell,k}(\mathbf{w})}. \tag{5b}$$

The interference-plus-noise expressions in the denominators of (5a) and (5b) are defined as

$$\Psi_k(\mathbf{w}, \boldsymbol{\alpha}) \triangleq \sum_{\forall k'\in\mathcal{K}\setminus\{k\}} (1 - \alpha_{k,k'})|\mathbf{h}_k \mathbf{w}_{k'}|^2 + \sigma_k^2, \tag{6a}$$

$$\Theta_{\ell,k}(\mathbf{w}) \triangleq \sum_{\forall k'\in\mathcal{K}\setminus\{k\}} |\mathbf{h}_\ell \mathbf{w}_{k'}|^2 + \sigma_\ell^2, \tag{6b}$$

[2]The benefits of such an arrangement of the channel gains are two-fold: (i) under the most cases of channel gain with the pathloss involved, the SIC is applied more efficiently at the near user than at the far user, (ii) by sorting the channel gains, the complexity of the combination problem is significantly reduced.

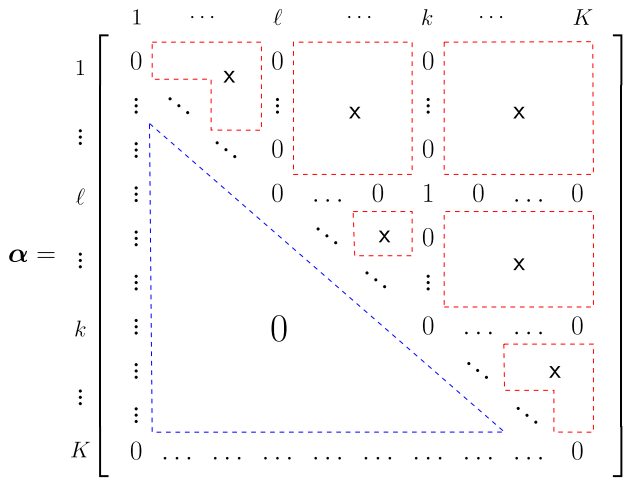


FIGURE 2. Structure of a pairing matrix.

where k' is used to indicate any UE different from UE_k, i.e., $\mathrm{UE}_{k'}$ may precede or succeed UE_k, while ℓ denotes the index of UEs that precede UE_k in $\mathcal{K}$, as defined in (1).

For simplicity, we redefine $\gamma_k(\mathbf{w}, \boldsymbol{\alpha}) \triangleq \mathrm{SINR}_k(\mathbf{w}, \boldsymbol{\alpha})$ and $\gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \triangleq \mathrm{SINR}_{k,k}(\mathbf{w}, \boldsymbol{\alpha})$. We further define $\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ to be the *effective* $\mathrm{SINR}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ for decoding UE_k's message at UE_ℓ:

$$\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) = \frac{|\mathbf{h}_\ell^H \mathbf{w}_k|^2}{(\alpha_{\ell,k} + \varepsilon)\Theta_{\ell,k}(\mathbf{w})}, \tag{7}$$

where a small number ε is used to avoid numerical failure when divided by zero if $\alpha_{\ell,k} = 0$, while guaranteeing that $\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ is extremely large in that case.

As illustrated in Fig. 2, assuming that the two UEs UE_ℓ and UE_k) with $\|\mathbf{h}_\ell\|^2 > \|\mathbf{h}_k\|^2$, (i.e., $\ell < k$) are paired to apply the NOMA technique, we obtain the following binary values $\alpha_{\ell,k} = 1$, $\alpha_{k,\ell} = 0$, and $\alpha_{\ell',\ell} = \alpha_{k,k'} = 0$, $\forall \ell' \in \{\mathcal{K}|\ell' \leq \ell - 1\}$, $\forall k' \in \{\mathcal{K}|k' \geq k + 1\}$. It is noteworthy that the symbol $\times$ in Fig. 2 denotes the elements with $\alpha_{\ell',k'} \in \{0, 1\}, \forall\{\ell' \neq \ell, k' \neq k\}$. In this case, UE_ℓ would decode UE_k's message and remove this message from its observation before decoding its own message. Accordingly, the expression in (4) indicates that the SINR for the UE_ℓ is determined by $\gamma_\ell(\mathbf{w}, \boldsymbol{\alpha}) = \gamma_{\ell,\ell}(\mathbf{w}, \boldsymbol{\alpha})$, since $\alpha_{\ell',\ell} = 0$ and $\gamma_{\ell',k}(\mathbf{w}, \boldsymbol{\alpha}) \to \infty$, $\forall \ell' \in \mathcal{K}$. Meanwhile, the SINR for UE_k becomes $\gamma_k(\mathbf{w}, \boldsymbol{\alpha}) = \min\{\gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha}), \gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})\}$ owing to the fact that $\gamma_{\ell',k}(\mathbf{w}, \boldsymbol{\alpha}) \to \infty$, $\forall \ell' \in \mathcal{K}\setminus\{\ell\}$. Recall that an important criterion for our UE pairing approach is that each UE can be paired with at most one UE. In summary, matrix $\boldsymbol{\alpha}$ satisfies the following linear constraints

$$\alpha_{\ell,k} + \sum_{\ell'=1}^{\ell-1} \alpha_{\ell',\ell} + \sum_{k'=k+1}^{K} \alpha_{k,k'} + \sum_{\substack{\ell'=1 \\ \ell'\neq\ell}}^{k-1} \alpha_{\ell',k}$$

$$+\sum_{\substack{k'=\ell+1\\k'\neq k}}^{K}\alpha_{\ell,k'}\leq 1,\ \forall(\ell,k)\in\{\mathcal{K}\times\mathcal{K}|\ell<k\},\quad(8)$$

$$\alpha_{\ell,k}=0,\ \forall(\ell,k)\in\{\mathcal{K}\times\mathcal{K}|\ell\geq k\}.\quad(9)$$

Remark 1: The arrangement of UEs in the descending channel gain can reduce the number of variables in binary matrix $\boldsymbol{\alpha}$. In particular, when the UEs are sorted, those with value 1 are always allocated above the main diagonal, based on the SIC technique in a NOMA system. Therefore, the variable matrix $\boldsymbol{\alpha}$ is an upper triangular matrix.

B. POWER CONSUMPTION MODEL

In this paper, power consumption includes the power for data transmission in the form of electromagnetic radiation, operation, signal processing at each UE, and consumption at the circuit [29], [30]. In particular, the total power consumption in our system can be modeled as follows:

$$P_{\mathrm{T}}(\mathbf{w})=P_{\mathrm{D}}(\mathbf{w})+P_{\mathrm{C}},\quad(10)$$

where $P_{\mathrm{D}}(\mathbf{w})$ is the power for data processing at the UE and transmission from the BS to all UEs, and P_{C} is the power consumed by the operation and circuitry.

- First, the power $P_{\mathrm{D}}(\mathbf{w})$ for the transmission and processing of data is composed of two parts as [29], [31]

$$P_{\mathrm{D}}(\mathbf{w})=\underbrace{\sum_{\forall k\in\mathcal{K}}\frac{1}{g_{\mathrm{PA}}}\|\mathbf{w}_k\|^2}_{\text{radiated power}}+\underbrace{\sum_{\forall k\in\mathcal{K}}P_k^{\mathrm{s}}}_{\text{signal processing power}}.\quad(11)$$

  The *radiated power* is the power for transmitting data from the BS to UE_k using electromagnetic radiation, with $g_{\mathrm{PA}}\in[0,1]$ being the power amplifier (PA) efficiency at the BS [32]. The *signal processing power* is the power consumed for data processing including the waveform generation, synchronization, and precoding/beamforming vector computation at UE_k.
- Second, the power consumption for circuit operation P_{C} can be expressed as [29], [31]

$$P_{\mathrm{C}}=P_{\mathrm{BS}}^{\mathrm{cir}}+P_{\mathrm{BS}}^{\mathrm{a}}+\sum_{\forall k\in\mathcal{K}}P_k^{\mathrm{cir}}.\quad(12)$$

  where $P_{\mathrm{BS}}^{\mathrm{cir}}$ and P_k^{cir} denote the power consumed by circuit operation at the BS and UE_k, respectively [29]. $P_{\mathrm{BS}}^{\mathrm{a}}$ is the power consumed when the BS is in the active mode, i.e., the power to operate the BS.

C. PROBLEM FORMULATION

From (4), the SE of UE_k (in nats/s/Hz)[3] is obtained as

$$R(\gamma_k(\mathbf{w},\boldsymbol{\alpha}))=\ln(1+\gamma_k(\mathbf{w},\boldsymbol{\alpha})),\quad k\in\mathcal{K}.\quad(13)$$

[3]For the sake of optimization, we use nats/s/Hz to devise the solution, while the SE in numerical results is presented in bits/s/Hz by scaling with $\ln 2$.

Subsequently, the sum SE is expressed as

$$F_{\mathrm{SE}}(\mathbf{w},\boldsymbol{\alpha})\triangleq\sum_{k\in\mathcal{K}}R(\gamma_k(\mathbf{w},\boldsymbol{\alpha})).\quad(14)$$

Therefore, the EE (in nats/J), defined as the ratio of the sum throughput (in nats/s) of all UEs to the total power consumption (in W), is expressed as

$$F_{\mathrm{EE}}(\mathbf{w},\boldsymbol{\alpha})\triangleq\frac{BF_{\mathrm{SE}}(\mathbf{w},\boldsymbol{\alpha})}{P_{\mathrm{T}}(\mathbf{w})},\quad(15)$$

where B is the system bandwidth.

We aim to maximize the EE under the power control and quality-of-service (QoS) requirement for each UE, formulated as

$$\max_{\mathbf{w},\boldsymbol{\alpha}}\ F_{\mathrm{EE}}(\mathbf{w},\boldsymbol{\alpha}),\quad(16a)$$

$$\text{s.t.}\ \sum_{k\in\mathcal{K}}\|\mathbf{w}_k\|^2\leq P_{\mathrm{BS}}^{\max},\quad(16b)$$

$$R(\gamma_k(\mathbf{w},\boldsymbol{\alpha}))\geq\bar{R}_k,\quad\forall k\in\mathcal{K},\quad(16c)$$

$$\alpha_{\ell,k}\in\{0,1\},\quad\forall k,\ell\in\mathcal{K},\quad(16d)$$

$$\alpha_{\ell,k}=0,\ \forall(\ell,k)\in\{\mathcal{K}\times\mathcal{K}|\ell\geq k\},\quad(16e)$$

$$\alpha_{\ell,k}+\sum_{\ell'=1}^{\ell-1}\alpha_{\ell',\ell}+\sum_{k'=k+1}^{K}\alpha_{k,k'}+\sum_{\substack{\ell'=1\\\ell'\neq\ell}}^{k-1}\alpha_{\ell',k}+\sum_{\substack{k'=\ell+1\\k'\neq k}}^{K}\alpha_{\ell,k'}\leq 1,\ \forall(\ell,k)\in\{\mathcal{K}\times\mathcal{K}|\ell<k\}.\quad(16f)$$

Clearly, constraint (16b) guarantees the maximum BS transmit power. Constraints (16d)-(16f) describe the criteria for UE pairing, wherein each UE can pair with at most another UE under the condition of the upper triangular matrix as mentioned in Remark 1. Constraint (16c) ensures the minimum per-UE rate, $\bar{R}_k,\ \forall k\in\mathcal{K}$. Because of the nonconcave objective function (16a), nonconvex constraint (16c), and binary constraints (16d)-(16f), the problem (16a) belongs to a mixed-integer nonconvex program; hence, it is generally difficult to solve directly [33], [34].

III. PROPOSED SOLUTION FOR EE MAXIMIZATION

It is observed that the interference-plus-noise terms in (16a) contain the strong compounds of binary variables and complex beamforming vectors. Hence, it is highly challenging to transform the problem into popular convex forms, such as semi-definite programming and second-order cone programming. To facilitate problem solving, we first relax the binary variables in (16a), hence, constraint (16d) becomes $0\leq\alpha_{k,\ell}\leq 1,\forall k,\ell\in\mathcal{K}$. Subsequently, we introduce a new variable $\omega_k,\ k\in\mathcal{K}$ as a soft SINR for UE_k, and apply it to (13) and (14). The sum SE is simplified to

$$F_{\mathrm{SE}}(\mathbf{w},\boldsymbol{\alpha})=\sum_{k\in\mathcal{K}}R(\gamma_k(\mathbf{w},\boldsymbol{\alpha}))\geq\sum_{k\in\mathcal{K}}R(\omega_k),\quad(17)$$

with a new constraint imposed by

$$\gamma_k(\mathbf{w},\boldsymbol{\alpha})\geq\omega_k,\ \forall k\in\mathcal{K},\quad(18)$$

Accordingly, the lower bound of the EE in (15) can be expressed as

$$\bar{F}_{\mathrm{EE}}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}) = \frac{B}{P_{\mathrm{T}}(\mathbf{w})} \sum_{k \in \mathcal{K}} R(\omega_k), \tag{19}$$

where $\boldsymbol{\omega} \triangleq \{\omega_k\}_{\forall k \in \mathcal{K}}$. Therefore, the relaxed problem for (16a) can be formulated as

$$\max_{\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}} \ \bar{F}_{\mathrm{EE}}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}), \tag{20a}$$
$$\text{s.t. (16b), (16e), (16f), (18),} \tag{20b}$$
$$0 \le \alpha_{\ell,k} \le 1, \ \forall k, \ell \in \mathcal{K}, \tag{20c}$$
$$\omega_k + 1 \ge \exp(\bar{R}_k), \ \forall k \in \mathcal{K}. \tag{20d}$$

Because the new variable $\boldsymbol{\omega}$ is used as an alternative for the SINR, the nonconvex constraint in (16c) is transformed into a linear constraint, as shown in (20d). In addition, the binary constraints are relaxed to the linear forms shown in (16e), (16f) and (20c). However, problem (20a) is still nonconvex owing to the nonconcave objective function (20a) and nonconvex constraint (18), which must be approximated to solve it.

A. CONVEXIFICATION FOR FEASIBLE SET OF PROBLEM (17)

Theorem 1: The solution to (20a) can be obtained iteratively by solving the following problem at iteration $(i+1)$:

$$\max_{\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu}} \ \bar{F}_{\mathrm{EE}}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}), \tag{21a}$$
$$\text{s.t. (16b), (16e), (16f), (20c), (20d),} \tag{21b}$$
$$|\mathbf{h}_\ell^H \mathbf{w}_k|^2 \le \mu_{\ell,k}, \ \forall \ell, k \in \mathcal{K}, \tag{21c}$$
$$\mathcal{A}_{k,k}^{(i)}(\mathbf{w}|\mathbf{H}, \Psi_k(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})) - \bar{\mathcal{A}}_{k,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}|\mathbf{H}, \Psi_k^{(i)}, f_{k,k}(\boldsymbol{\alpha}, \boldsymbol{\mu})) \ge \omega_k, \ k \in \mathcal{K}, \tag{21d}$$
$$\mathcal{A}_{\ell,k}^{(i)}(\mathbf{w}|\mathbf{H}, \Phi_{\ell,k}^{(i)}) - \bar{\mathcal{A}}_{\ell,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}|\mathbf{H}, \Phi_{\ell,k}^{(i)}, f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu})) \ge \omega_k, \ \ell, k \in \mathcal{K}, \ell \ne k, \tag{21e}$$

where $\boldsymbol{\mu} \triangleq [\mu_{\ell,k}]_{\forall \ell,k \in \mathcal{K}}$ is introduced as a new variable, whereas the channel matrix $\mathbf{H} \triangleq [\mathbf{h}_k]_{\forall k \in \mathcal{K}} \in \mathbb{C}^{N \times K}$, $\Psi_k^{(i)} \triangleq \Psi_k(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})$, $\Phi_{\ell,k}^{(i)} \triangleq (\alpha_{\ell,k}^{(i)} + \varepsilon)\Theta_{\ell,k}(\mathbf{w}^{(i)})$ and

$$\mathcal{A}_{\ell,k}^{(i)}(\mathbf{w}|\mathbf{A}, C) \triangleq 2\frac{(\mathbf{a}_\ell^H \mathbf{w}_k^{(i)}) * (\mathbf{a}_\ell^H \mathbf{w}_k)}{C}, \ \ell, k \in \mathcal{K},$$
$$\bar{\mathcal{A}}_{\ell,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}|\mathbf{A}, C, f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu})) \triangleq \frac{|\mathbf{a}_\ell^H \mathbf{w}_k^{(i)}|^2}{C^2} f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu}),$$
$$f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu}) \triangleq \begin{cases} \sigma_\ell^2 + \sum_{\forall k' \in \mathcal{K} \setminus \{k\}} g_{\mathrm{pr}}^{(i)}(1 - \alpha_{\ell,k'}, \mu_{\ell,k'}), & \text{if } \ell = k, \\ \alpha_{\ell,k}\sigma_k^2 + \sum_{\forall k' \in \mathcal{K} \setminus \{k\}} g_{\mathrm{pr}}^{(i)}(\alpha_{\ell,k}, \mu_{\ell,k'}), & \text{if } \ell \ne k, \end{cases}$$

with the predefined matrix $\mathbf{A} \in \mathbb{C}^{N \times K}$ with $\mathbf{a}_\ell$ being the ℓ-th column of $\mathbf{A}$. C is a constant scalar, and $\mathbf{X} \in \mathbb{C}^{K \times K}$ and $\mathbf{Y} \in \mathbb{C}^{K \times K}$ are the matrices of the variables. ■

Proof: Please see Appendix B. ■

It is evident that the feasible set of problem (21a) is convex, because it merely contains the linear and second-order-cone (SOC) constraints. However, we must address the nonconcave objective function before deriving a solution to (20a).

B. PROPOSED ALGORITHM FOR SOLVING PROBLEM (14)

The throughput in the objective function (19) is easily obtained as in [31, Theorem 3]:

$$B\sum_{k \in \mathcal{K}} R(\omega_k) = B\ln\Big(\prod_{k \in \mathcal{K}}(1 + \omega_k)\Big) = B\ln|\mathbf{I} + \boldsymbol{\Omega}| := \bar{F}_{\mathrm{SE}}(\boldsymbol{\Omega}), \tag{22}$$

where $\boldsymbol{\Omega} \triangleq \mathrm{diag}([\omega_1, \omega_2, \cdots, \omega_k])$. Subsequently, an auxiliary variable ϑ can be used for power control as the following convex constraint:

$$P_{\mathrm{T}}(\mathbf{w}) \le \vartheta. \tag{23}$$

Hence, the objective function (21a) can be represented as a concave-over-linear form, i.e., $\frac{\bar{F}_{\mathrm{SE}}(\boldsymbol{\Omega})}{\vartheta}$. By applying the Dinkelbach transformation, sub-problem (21a) at iteration $(i+1)$ can be equivalently expressed as

$$\max_{\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu}, \vartheta} \ \bar{F}_{\mathrm{SE}}(\boldsymbol{\Omega}) - u^{(i)}\vartheta, \tag{24a}$$
$$\text{s.t. (16b), (16e), (16f), (20c), (20d), (21c), (21d), (21e), (23),} \tag{24b}$$

where $u^{(i)} \triangleq \frac{\bar{F}_{\mathrm{SE}}(\boldsymbol{\Omega}^{(i)})}{\vartheta^{(i)}}$ is the optimal value achieved in the previous iteration. Based on the properties of IA and Dinkelbach methods [35], [36], the optimality of sub-problem (21a) and (24a) holds, resulting in the optimum of problem (20a) at the convergence of an iterative algorithm. However, the feasibility for sub-problem (24a) is numerically difficult to obtain at the beginning of the loop in the algorithm, particularly because of the QoS constraint (20d):

To solve problem (24a), an initial starting point $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\omega}^{(0)}, \boldsymbol{\mu}^{(0)}, \vartheta^{(0)})$ is required. Hence, we solve an alternative problem that generates a feasible point for (24a):

$$\max_{\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu}, \mathbf{v}} \ V \triangleq \sum_{\forall k \in \mathcal{K}} v_k, \tag{25a}$$
$$\text{s.t. (16b), (16e), (16f), (20c), (21c), (21d), (21e),} \tag{25b}$$
$$\omega_k + 1 - \exp(\bar{R}_k) \ge v_k, \quad \forall k \in \mathcal{K}, \tag{25c}$$
$$v_k \le 0, \quad \forall k \in \mathcal{K}, \tag{25d}$$

where $\mathbf{v} = \{v_k\}_{\forall k \in \mathcal{K}}$ is a new slack variable. It can be foreseen that the optimal value V approximately reaches zero when the equalities of constraints (25c) and (25d) hold. This is equivalent to the fact that the QoS constraint in (20d) is satisfied; therefore, the initial feasible point $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\omega}^{(0)}, \boldsymbol{\mu}^{(0)}, \vartheta^{(0)})$ can be obtained.

When the iterative algorithm for problem (20a) terminates, the value of $\boldsymbol{\alpha}$ still belongs to the closed interval [0, 1], i.e., may not be an exact binary value (zero or one). In fact, after a limited number of iterations, the iterative algorithm reaches the convergence when a difference between two consecutive objective values is tolerant. Therefore, the optimal solution with real value of $\boldsymbol{\alpha}$ when solving problem (20a) is not a solution for problem (16a). To transform $\boldsymbol{\alpha}$ obtained by solving (20a) into a binary value, we use a rounding function:

$$\alpha^{\star}_{k,\ell} = \lfloor \alpha^{(i)}_{k,\ell} + \frac{1}{2} \rfloor, \quad \forall k, \ell \in \mathcal{K}, \tag{26}$$

where $\lfloor x \rfloor$, $x \in \mathbb{R}$ returns the maximum integer not larger than x. Subsequently, problem (20a) is resolved using the fixed binary value $\boldsymbol{\alpha}$ to obtain the optimal solution for problem (16a). The proposed iterative algorithm for solving the EE maximization problem (16a) under the perfect CSI assumption at the BS is detailed in Algorithm 1.

Algorithm 1 Proposed Algorithm to Solve Problem (16a)

Phase 1:

1: **Initialization:** Set $i := 0$ and solve (25a) to generate the initial starting point $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\omega}^{(0)}, \boldsymbol{\mu}^{(0)}, \vartheta^{(0)})$.
2: **repeat**
3: Solve (24a) to obtain $(\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star}, \boldsymbol{\omega}^{\star}, \boldsymbol{\mu}^{\star}, \vartheta^{\star})$.
4: Update $(\mathbf{w}^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\omega}^{(i+1)}, \boldsymbol{\mu}^{(i+1)}, \vartheta^{(i+1)}) := (\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star}, \boldsymbol{\omega}^{\star}, \boldsymbol{\mu}^{\star}, \vartheta^{\star})$.
5: Set $i := i + 1$.
6: **until** Convergence {Obtain the optimal solution for problem (20a)}
7: **Output-1:** The optimal solution $(\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star})$ with the continuous value of $\boldsymbol{\alpha}$.

Phase 2:

8: Recover the binary value of $\boldsymbol{\alpha}^{\star}$, using (26).
9: Repeat steps 1-6 with the fixed value $\boldsymbol{\alpha}^{\star}$ (obtained at step 8) to compute $\mathbf{w}^{\star}$.
10: **Output-2:** The optimal solution $(\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star})$ for problem (16a).

C. CONVERGENCE AND COMPLEXITY ANALYSIS

1) CONVERGENCE ANALYSIS

The convergence behavior of the proposed algorithm reflects the properties of the IA method and Dinkelbach transformation [35], [36]. According to [36], the optimum successive convex problem presented in (21a) is obtained from that presented in (24a). Therefore, problems (21a) and (24a) have the same optimum and feasible set. Let $\mathcal{X} = (\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu})$ and $\mathcal{F}^{(i)}$ be the feasible point and feasible set of approximated problem (21a) at iteration i, respectively. It is proven that $\sup_{\mathcal{X} \in \mathcal{F}^{(i+1)}} \bar{F}_{\mathrm{EE}}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega}) \geq \sup_{\mathcal{X} \in \mathcal{F}^{(i)}} \bar{F}_{\mathrm{EE}}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\omega})$ for a connected set $\mathcal{F}^{(i)}$ [37]. Hence, the program in (21a) monotonically improves the objective values for problems (20a) and (16a) as well as provides at least a local optimal solution satisfying the Karush-Kuhn-Tucker conditions at the convergence. In practice, the algorithm can be numerically terminated at a finite number of iterations when $u^{(i+1)} - u^{(i)} < 10^{-3}$.

2) COMPLEXITY ANALYSIS

According to [38], the complexity of a convex problem can be measured by the number of constraints and variables, even when the objective function is not linear. As can be seen, problem (24a) comprises $(5K^2 + K + 2)$ SOC/linear constraints and $(2K^2 + NK + K + 1)$ variables. Consequently, the per-iteration complexity for solving problem (24a) is $\mathcal{O}((2K^2 + NK + K)^{2.5}((5K^2 + 3K + 1)^2 + 2K^2 + NK + K))$.

IV. ROBUST TRANSMISSION DESIGN

In Section III, the proposed design for EE maximization is analyzed under the ideal assumption of perfect CSI, which is typically difficult to achieve in practice. In real world scenarios, it often happens that the channel is not known a priori at the BS owing to the complex urban environment. The EE optimization of a system that is robust to the imperfect CSI condition has thus attracted attention, e.g., for a MIMO two-way relay network in [39], for a full-duplex multi-user multi-cell MIMO network in [40], and for a MISO non-orthogonal multiple access system in [28]. For practical purposes, this section develops a robust design for EE optimization of the proposed NOMA-aided downlink network under channel uncertainty, which has not been considered in the literature so far. Since the real channel $\mathbf{h}_k$ is assumed to be unknown, we first decompose it as [17], [28], [40]

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \Delta \mathbf{h}_k, \tag{27}$$

where $\hat{\mathbf{h}}_k$ and $\Delta \mathbf{h}_k$ are the channel estimate and estimation error, respectively. In this section, the BS is only aware of $\hat{\mathbf{h}}_k$ provided by the estimation scheme, whereas the estimation error is undetermined and independent of the channel estimate.[4] Based on the estimation theory, the error can be considered as a random variable that follows a complex Gaussian distribution [41], [42], i.e., $\Delta \mathbf{h}_k \sim \mathcal{CN}(0, \epsilon_k \mathbf{I})$ with the variance ϵ_k being modeled as

$$\epsilon_k \triangleq \delta \rho_k^{-\lambda}, \tag{28}$$

where the SNR at UE_k is supposed to be $\rho_k \triangleq \frac{P_{\mathrm{BS}}^{\max}}{\sigma_k^2}$ for the worst-case design, $\delta \geq 0$ and $\lambda > 0$. It is clear that $\delta = 0$ corresponds to the perfect CSI scenario.

The SINR at UE_k under channel uncertainty can be rewritten as

$$\hat{\gamma}_k(\mathbf{w}, \boldsymbol{\alpha}) = \min \left\{ \hat{\gamma}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}), \min_{\ell \in \mathcal{K} \setminus \{k\}} \{\hat{\gamma}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})\} \right\}, \tag{29}$$

where $\hat{\gamma}_{k,k}(\mathbf{w}, \boldsymbol{\alpha})$ denotes the SINR for decoding UE_k's message at UE_k itself and $\hat{\gamma}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ stands for the *effective* SINR for decoding UE_k's message at UE_ℓ (see (7)), which are

[4] Unless otherwise stated, the SIC procedure is assumed to be perfect in this paper.

expressed as

$$\hat{\gamma}_{k,k}(\mathbf{w},\boldsymbol{\alpha}) = \frac{|\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2}{\hat{\Psi}_k(\mathbf{w},\boldsymbol{\alpha})}, \tag{30a}$$

$$\hat{\gamma}_{\ell,k}(\mathbf{w},\boldsymbol{\alpha}) = \frac{|\hat{\mathbf{h}}_\ell^H \mathbf{w}_k|^2}{(\alpha_{\ell,k}+\varepsilon)\hat{\Theta}_{\ell,k}(\mathbf{w})}, \tag{30b}$$

where $\hat{\Psi}_k(\mathbf{w},\boldsymbol{\alpha})$ and $\hat{\Theta}_{\ell,k}(\mathbf{w},\boldsymbol{\alpha})$ are defined as

$$\hat{\Psi}_k(\mathbf{w},\boldsymbol{\alpha}) = \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} (1-\alpha_{k,k'})|\hat{\mathbf{h}}_k \mathbf{w}_{k'}|^2 + \sum_{\forall k' \in \mathcal{K}} \epsilon_k \|\mathbf{w}_{k'}\|^2 + \sigma_k^2, \tag{31a}$$

$$\hat{\Theta}_{\ell,k}(\mathbf{w}) = \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} |\hat{\mathbf{h}}_\ell \mathbf{w}_{k'}|^2 + \sum_{\forall k' \in \mathcal{K}} \epsilon_\ell \|\mathbf{w}_{k'}\|^2 + \sigma_\ell^2. \tag{31b}$$

Subsequently, the sum SE can be rewritten as

$$\hat{F}_{\mathrm{SE}}(\mathbf{w},\boldsymbol{\alpha}) = \sum_{k\in\mathcal{K}} R(\hat{\gamma}_k(\mathbf{w},\boldsymbol{\alpha})). \tag{32}$$

Therefore, a robust design problem for EE maximization can be formulated as

$$\max_{\mathbf{w},\boldsymbol{\alpha}} \ \frac{B\hat{F}_{\mathrm{SE}}(\mathbf{w},\boldsymbol{\alpha})}{P_{\mathrm{T}}(\mathbf{w})}, \tag{33a}$$

$$\text{s.t. (16b), (16d)} - \text{(16f)}, \tag{33b}$$

$$R(\hat{\gamma}_k(\mathbf{w},\boldsymbol{\alpha})) \geq \bar{R}_k, \quad \forall k \in \mathcal{K}. \tag{33c}$$

By utilizing the same steps (17)-(19) to derive (20a), the relaxed robust design problem for EE maximization can be expressed as

$$\max_{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\omega}} \ \bar{F}_{\mathrm{EE}}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\omega}), \tag{34a}$$

$$\text{s.t. (16b), (16e), (16f), (20c), (20d),} \tag{34b}$$

$$\hat{\gamma}_k(\mathbf{w},\boldsymbol{\alpha}) \geq \omega_k, \ \forall k \in \mathcal{K}. \tag{34c}$$

To convexify (34a), we introduce the following proposition:

Proposition 1: By applying **Theorem 1** and the steps in (22)–(24a), the sub-problem in a successive convex program providing a minorant maximization for problem (34a) at iteration $(i+1)$ is formulated as

$$\max_{\substack{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\omega},\\ \boldsymbol{\mu},\boldsymbol{\tau},\vartheta}} \ \bar{F}_{\mathrm{SE}}(\boldsymbol{\Omega}) - u^{(i)}\vartheta, \tag{35a}$$

$$\text{s.t. (16b), (16e), (16f), (20c), (20d), (23),} \tag{35b}$$

$$|\hat{\mathbf{h}}_\ell^H \mathbf{w}_k|^2 \leq \mu_{\ell,k}, \ \forall \ell, k \in \mathcal{K}, \tag{35c}$$

$$\|\mathbf{w}_k\|^2 \leq \tau_k, \forall k \in \mathcal{K}, \tag{35d}$$

$$\begin{aligned} &\mathcal{A}_{k,k}^{(i)}(\mathbf{w}|\hat{\mathbf{H}}, \hat{\Psi}_k^{(i)}) \\ &- \Xi_{k,k}\big(\mathbf{w}^{(i)}|\hat{\mathbf{H}}, \hat{\Psi}_k^{(i)}\big) \sum_{\forall k' \in \mathcal{K}} \epsilon_k \tau_{k'} \\ &- \bar{\mathcal{A}}_{k,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\mu}|\hat{\mathbf{H}}, \hat{\Psi}_k^{(i)}, f_{k,k}(\boldsymbol{\alpha},\boldsymbol{\mu})) \\ &\geq \omega_k, \ k \in \mathcal{K}, \end{aligned} \tag{35e}$$

$$\begin{aligned} &\mathcal{A}_{\ell,k}^{(i)}(\mathbf{w}|\hat{\mathbf{H}}, (\alpha_{\ell,k}^{(i)}+\varepsilon)\hat{\Theta}_{\ell,k}(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})) \\ &- \Xi_{\ell,k}\big(\mathbf{w}^{(i)}|\hat{\mathbf{H}}, \hat{\Phi}_{\ell,k}^{(i)}\big) \sum_{\forall k' \in \mathcal{K}} \epsilon_\ell g_{\mathrm{pr}}^{(i)}(\alpha_{\ell,k}, \tau_{k'}) \\ &- \bar{\mathcal{A}}_{\ell,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\mu}|\hat{\mathbf{H}}, \hat{\Phi}_{\ell,k}^{(i)}, f_{\ell,k}(\boldsymbol{\alpha},\boldsymbol{\mu})) \\ &\geq \omega_k, \ \ell, k \in \mathcal{K}, \ell \neq k, \end{aligned} \tag{35f}$$

where $\boldsymbol{\tau} \triangleq [\tau_k]_{\forall k\in\mathcal{K}}$ is introduced as a new variable, whereas the channel estimates $\hat{\mathbf{H}} \triangleq [\hat{\mathbf{h}}_k]_{\forall k\in\mathcal{K}} \in \mathbb{C}^{N\times K}$, $\hat{\Psi}_k^{(i)} \triangleq \hat{\Psi}_k(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})$, $\hat{\Phi}_{\ell,k}^{(i)} \triangleq (\alpha_{\ell,k}^{(i)}+\varepsilon)\hat{\Theta}_{\ell,k}(\mathbf{w}^{(i)})$ and

$$\Xi_{\ell,k}\big(\mathbf{w}|\mathbf{A}, C\big) \triangleq \frac{|\mathbf{a}_\ell^H \mathbf{w}_k|^2}{C^2}.$$

Proof: Please see Appendix C. ■

To derive a solution for the robust design problem (33a), we utilize Algorithm 1 with the following modifications. In Step 1, instead of solving (25a), the initial starting point is generated by solving the following problem:

$$\max_{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\omega},\boldsymbol{\mu},\boldsymbol{\tau},\mathbf{v}} \ V \triangleq \sum_{\forall k\in\mathcal{K}} v_k, \tag{36a}$$

$$\begin{aligned} &\text{s.t. (16b), (16e), (16f), (20c), (25c),} \\ &\text{(25d), (35c)} - \text{(35f),} \end{aligned} \tag{36b}$$

Subsequently, the convex problem (24a) is replaced by (35a) in Step 3, and Step 4 is performed to update the hextuple variable $(\mathbf{w}^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\omega}^{(i+1)}, \boldsymbol{\mu}^{(i+1)}, \boldsymbol{\tau}^{(i+1)}, \vartheta^{(i+1)}) := (\mathbf{w}^\star, \boldsymbol{\alpha}^\star, \boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\tau}^\star, \vartheta^\star)$. The customized algorithm is presented in Algorithm 2.

Algorithm 2 Proposed Algorithm to Solve Problem (16a) Under Channel Uncertainty

Phase 1:

1: **Initialization:** Set $i := 0$ and solve (36a) to generate $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\omega}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\tau}^{(0)}, \vartheta^{(0)})$.
2: **repeat**
3: Solve (35a) to obtain $(\mathbf{w}^\star, \boldsymbol{\alpha}^\star, \boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\tau}^\star, \vartheta^\star)$.
4: Update $(\mathbf{w}^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\omega}^{(i+1)}, \boldsymbol{\mu}^{(i+1)}, \boldsymbol{\tau}^{(i+1)}, \vartheta^{(i+1)}) := (\mathbf{w}^\star, \boldsymbol{\alpha}^\star, \boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\tau}^\star, \vartheta^\star)$.
5: Set $i := i+1$.
6: **until** Convergence {Obtain the optimal solution for (34a)}
7: **Output-1:** The optimal solution $(\mathbf{w}^\star, \boldsymbol{\alpha}^\star)$ with the continuous value of $\boldsymbol{\alpha}$.

Phase 2:

8: Recover the binary value of $\boldsymbol{\alpha}^\star$, using (26).
9: Repeat steps 1-6 with $\boldsymbol{\alpha}^\star$ obtained at step 8 to compute $\mathbf{w}^\star$.
10: **Output-2:** The optimal solution $(\mathbf{w}^\star, \boldsymbol{\alpha}^a r)$ for problem (16a).

V. NUMERICAL RESULTS

To evaluate the proposed methods, we consider a DL cellular system comprising a centrally located BS equipped with N antennas and K single-antenna UEs. The Monte Carlo framework is used to investigate the performance of the considered schemes. The channel is assumed to undergo flat fading; the small-scale fading is modeled as a random variable that follows a circularly-symmetric complex normal

TABLE 1. Simulation Parameters.

Parameter	Value
Radius of the small-cell, R	100 m
System bandwidth	20 MHz
Noise power at UEs	-174 dBm/Hz
PL between BS and UEs, $\mathrm{PL}_{\mathrm{BS,UE}}$	$145.4 + 37.5\log_{10}(d_{\mathrm{BS,UE}})$ dB
Nearest distance limit from BS to the UEs	≥ 10 m
Power budget at BS, $P_{\mathrm{BS}}^{\max}$	18 dBm
Number of UEs, K	6
Number of antennas at BS, N	4
Required QoS level, $\bar{R}_k = \bar{R}_\ell = \bar{R}, \forall k, \ell$	1 bits/s/Hz
BS's power in active mode, $P_{\mathrm{BS}}^{\mathrm{a}}$	10 W
Power for circuit operation at BS, $P_{\mathrm{BS}}^{\mathrm{cir}}$	1 W
PA efficiency at BS, g_{PA}	0.3
UEs' power for circuit operation, $P_k^{\mathrm{cir}} = P_\ell^{\mathrm{cir}}, \forall k, \ell$	0.1 W
UEs' power for signal processing, $P_k^{\mathrm{s}} = P_\ell^{\mathrm{s}}, \forall k, \ell$	0.1 W

distribution, i.e., $\mathbf{h} \sim \mathcal{CN}(0, \mathrm{PL}_{\mathrm{BS,UE}}\mathbf{I})$, where $\mathrm{PL}_{\mathrm{BS,UE}}$ is the path loss (PL) between the BS and UE with distance $d_{\mathrm{BS,UE}}$. Other simulation parameters are shown in Table 1, in which the power consumption and PA efficiency parameters are obtained from [29]. The proposed algorithms are terminated when the difference of objective values between two successive iterations is smaller than 10^{-3}. To quantify the effectiveness of the proposed scheme, we compare the proposed scheme with the following benchmark schemes:

- Exhaustive Search for Pairing (ESP): This scheme considers all possible cases of $\boldsymbol{\alpha}$ to derive a set of sub-problems. Subsequently, each sub-problem corresponding to each value of $\boldsymbol{\alpha}$ is solved using the proposed algorithm. Finally, the optimal solution is selected as it provides the best performance among the sub-problems.
- Random Pairing (RaP): In this scheme, the proposed algorithm is utilized to obtain the beamforming vectors for a sub-problem, wherein the UE association matrix $\boldsymbol{\alpha}$ is fixed by generating random values for the upper triangle part of $\boldsymbol{\alpha}$.
- Beamforming design (BFD): For the conventional BFD, $\boldsymbol{\alpha}$ is set to an all-zero matrix; subsequently, the proposed algorithm is used to compute the power control for beamforming vectors.

Among these three schemes, the ESP method would yield the highest EE, albeit having the highest complexity. Without using UE pairing, the BFD method will provide the lowest EE but require the lowest complexity due to the simple design. RaP is considered as an intermediate solution for the two extreme cases with a performance worse than that of ESP but with a much lower complexity as specified in Section V-A.

A. EE PERFORMANCE AND COMPLEXITY

Fig. 3 shows the cumulative distribution function (CDF) of the average EE of the proposed method (Alg. 1) and the three aforementioned schemes. The CDF results are obtained with 1000 different random channel realizations. As expected, the EE of the BFD scheme is much worse than that of the other three NOMA-based schemes, with an EE difference from 0.2 to 1.2 Mbits/J at the midpoint. As expected, the proposed algorithm yields an EE performance similar to that of ESP, with an offset of 0.1 Mbit/J, while outperforming the other

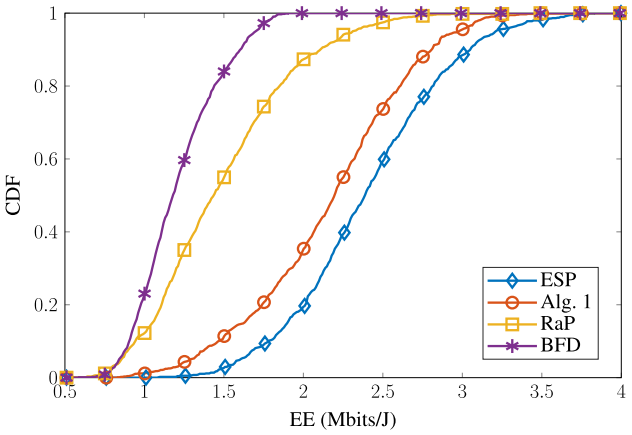


FIGURE 3. Cumulative distribution function of the EE (Mbits/J).

schemes, i.e., the BFD and RaP. Under various random channel responses, Fig. 3 validates the adaption of the proposed method for EE optimization (when compared with ESP), and verifies the advantages of pairing schemes in NOMA-based systems.

In addition to EE performance, computational complexity is a key metric for evaluating the effectiveness of algorithms. In particular, the convergence rate, which must be intuitively investigated via simulations, is typically used as a fundamental criterion of computational complexity. Fig. 4 shows the convergence behavior under a typical channel scenario with different numbers of UEs, e.g., $K = 6$ and $K = 10$. It is noteworthy that the ESP comprises many sub-problems, each of which is equivalent to a problem using the RaP scheme. Because the proposed algorithm is designed to generalize the UE pairing methods, it is sufficient to consider the RaP scheme in terms of the number of iterations while omitting the ESP scheme. In general, in Alg. 1, the RaP and BFD schemes require a similar number of iterations to reach the convergence, i.e., approximately 10 iterations. The jumping phenomenon in the convergence behavior of the proposed algorithm is caused by the first step of phase 2, in which the relaxed values of matrix $\boldsymbol{\alpha}$ were recovered into binary ones. Nevertheless, the saturation values for the objective functions are quickly reached after about 10 iterations and remain constant afterwards, providing a fast convergence rate.

Another crucial criterion of computational complexity is the per-iteration complexity in terms of big-O as a function of the number of variables and that of SOC/linear constraints [38]. We present the comparison of Alg. 1 and considered schemes based on this criterion in Table 2. Accordingly, the complexity per iteration for both the ESP (in a sub-problem) and RaP (in a problem/sub-problem) schemes is typically expressed as $\mathcal{O}((v_2^{2.5}(c_2^2 + v_2)))$. Meanwhile, the BFD method has a similar big-O expression; however, it requires a smaller number of constraints, i.e., $K^2 + 2K + 2$. This confirms that the order of this complexity expression is the same as that of the complexity required by the proposed algorithm,

TABLE 2. Comparison of Complexity.

Schemes	Sub-problems	Constraints	Variables	Complexity
Alg. 1	1	$c_1 = 5K^2 + K + 2$	$v_1 = 2K^2 + NK + K + 1$	$\mathcal{O}(v_1^{2.5}(c_1^2 + v_1))$
ESP	$s_2 = \sum_{i=1}^{\lfloor K/2 \rfloor} \binom{K}{i} \frac{(2i)!}{2^i i!}$	$c_2 = K^2 + 2K + \lfloor K/2 \rfloor + 2$	$v_2 = K^2 + NK + K + 1$	$\mathcal{O}(s_2 v_2^{2.5}(c_2^2 + v_2))$
RaP	1	$c_3 = c_2$	$v_3 = v_2$	$\mathcal{O}(v_3^{2.5}(c_3^2 + v_3))$
BFD	1	$c_4 = K^2 + 2K + 2$	$v_4 = v_2$	$\mathcal{O}(v_4^{2.5}(c_4^2 + v_4))$

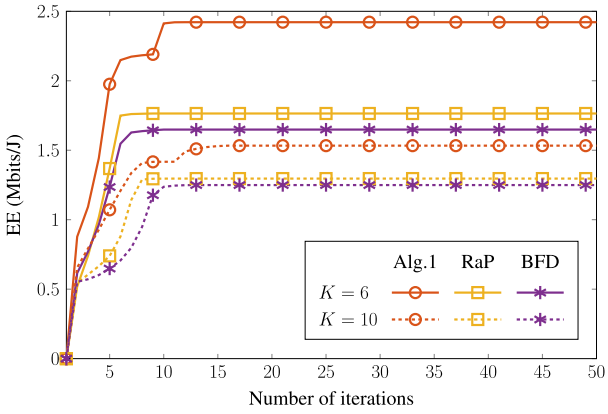


FIGURE 4. Convergence behavior on the EE performance.

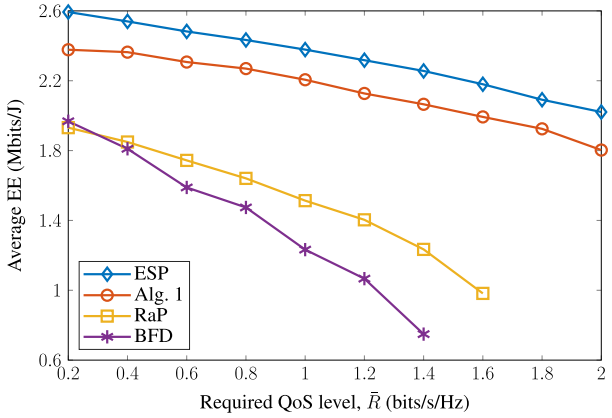


FIGURE 5. Average EE (Mbits/J) as a function of the required QoS level.

as mentioned in Section III-C. With the presence of a large K, the per-iteration complexities for all the methods above are comparable with each other. However, the ESP method requires a considerably long operation time owing the large number of sub-problems. Meanwhile, the proposed method and other methods merely address an original problem at one time.

B. EFFECT OF QoS REQUIREMENT

In this subsection, we present the effects of the QoS requirement and maximum power budget at the BS on the EE performance. Fig. 5 shows the average EE as a function of the minimum bit rate (or the QoS threshold), $\bar{R}$, in bits/s/Hz. As expected, the EE of all methods degrades as the minimum bit rate $\bar{R}$ increases. For the same QoS threshold, the adaptive capability of the proposed algorithm as compared with the BFD and RaP schemes is confirmed, providing the performance gains of at least 0.5 Mbits/J. Intuitively, it is feasible to obtain a good solution based on Alg. 1 for higher QoS requirements, i.e., $\bar{R} \in \{1.8, 2.0\}$. More interestingly, Fig. 5 shows that the performance of the proposed method is similar to that of the ESP, with a small EE loss of approximately 0.15 Mbits/J.

To further comprehend the EE behavior under changes in the QoS requirement, we plot the CDF of $\bar{R}$, as shown in Fig. 6. It can be observed that the probabilities of obtaining a feasible point for the four considered strategies are inversely proportional to $\bar{R}$. The 95-percentile points of the BFD and RaP schemes are achieved at approximately $\bar{R} = 1.6$ and 2 bits/s/Hz, respectively, rendering it difficult to obtain the feasible points. By contrast, Alg. 1 and the ESP methods

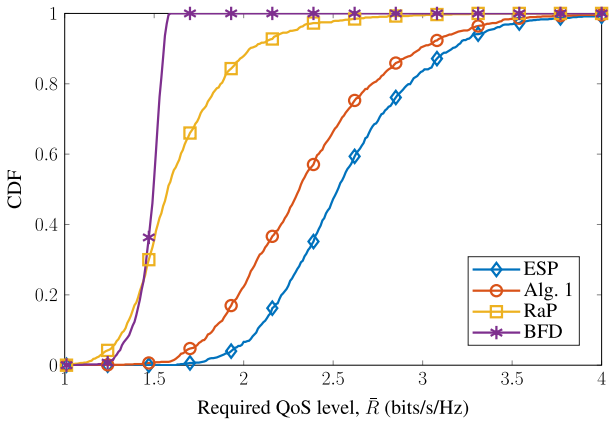


FIGURE 6. Cumulative distribution function of the required QoS.

have a 95-percentile point around 3 bits/s/Hz. These results clearly reflect the observation in Fig. 5, thereby confirming the advantage of the proposed scheme for the NOMA-based system.

C. EFFECT OF MAXIMUM POWER BUDGET AT BS

Fig. 7 shows the change in EE when the maximum power budget at the BS, i.e., $P_{\text{BS}}^{\max}$, increases. As shown, the proposed method outperforms both the BFD and RaP schemes while providing an EE similar to that obtained using ESP. In particular, compared with the BFD scheme, the performance gain of the RaP scheme at the floor level is approximately 0.4 Mbits/J, which is less than that of the proposed method, i.e., approximately 1.4 Mbits/J. This implies the capability of the

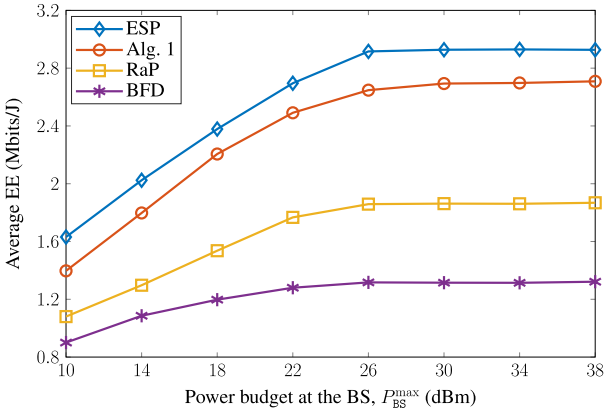


FIGURE 7. Average EE (Mbits/J) as a function of the BS power budget $P_{\text{BS}}^{\max}$.

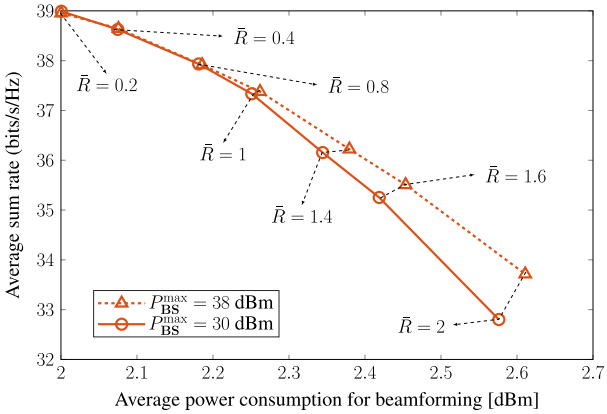


FIGURE 8. Relationship between power consumption and sum rate with $\bar{R} \in [0.2, 2]$ bits/s/Hz.

proposed approach in obtaining a good solution for NOMA-based systems. In addition, Fig. 7 shows that the EE of all schemes reaches saturated values when the maximum power budget is sufficiently high, i.e., $P_{\text{BS}}^{\max} = 26$ dBm.

To obtain insights into the EE behavior of the proposed method, Fig. 8 shows the relationship between the sum rate and average total power consumption, W, with respect to different values of minimum bit rates and maximum power budgets at the BS, e.g., $\bar{R} \in [0.2, 2]$ bits/s/Hz and $P_{\text{BS}}^{\max} \in \{30, 38\}$ dBm, respectively. It is noteworthy that the power consumption is directly proportional to the minimum bit rate, whereas the sum rate shows a reverse trend. In fact, the higher the power resource, the greater is the degree of freedom provided to satisfy the QoS constraint. This is verified by the observation that the gap between $P_{\text{BS}}^{\max} = 30$ dBm and $P_{\text{BS}}^{\max} = 38$ dBm increases with the QoS threshold, beginning from $\bar{R} > 1$ bits/s/Hz. When $\bar{R} < 1$ bits/s/Hz, the values of the sum rate and power consumption in the two cases of $P_{\text{BS}}^{\max} = 30$ dBm and $P_{\text{BS}}^{\max} = 38$ dBm are similar to each other. Consequently, the EE is unchanged, thereby depicting a sufficient power budget. This result validates the observations from Fig. 7.

D. EFFECT OF NUMBER OF UEs

To further investigate the effectiveness of the proposed algorithm, we now evaluate the EE performance of the proposed algorithm with respect to the number of UEs. Since the BFD yields the worst EE performance, as mentioned at the beginning of this section, we therefore consider the RaP method as the baseline for UE pairing. In addition to the Alg. 1 and RaP scheme, we investigate the EE of the following typical greedy-based pairing schemes.[5]

- Greedy pairing for channel gain difference (GP-CGD): This refers to the pairing scheme in [12], wherein if the k-th and $(K - k + 1)$-th UEs are paired, then it corresponds to $\alpha_{k,K-k+1} = 1, 1 \le k \le \lfloor K/2 \rfloor$.
- Greedy pairing for dynamic zone boundary (GP-DZB): As in the pairing scheme in [14], the boundary of the inner and outer zones is dynamically determined using the sorted list of channel gain. Accordingly, two sets of UEs are established: the strong-zone set includes the $\lfloor K/2 \rfloor$ strongest UEs from the BS, and the weak-zone set contains the remaining $K - \lfloor K/2 \rfloor$ UEs. Subsequently, the strongest UEs in the strong-zone and weak-zone sets are popped out and paired to each other until the strong-zone set is empty. This results in $\alpha_{k,K-\lfloor K/2 \rfloor+k} = 1, 1 \le k \le \lfloor K/2 \rfloor$.
- Consecutive paring (CoP): This refers to the pairing scheme in [13], where two consecutive UEs are paired with each other, and the $(2k-1)$-th and $(2k)$-th UEs are paired based on $\alpha_{2k-1,2k} = 1, 1 \le k \le \lfloor K/2 \rfloor$.

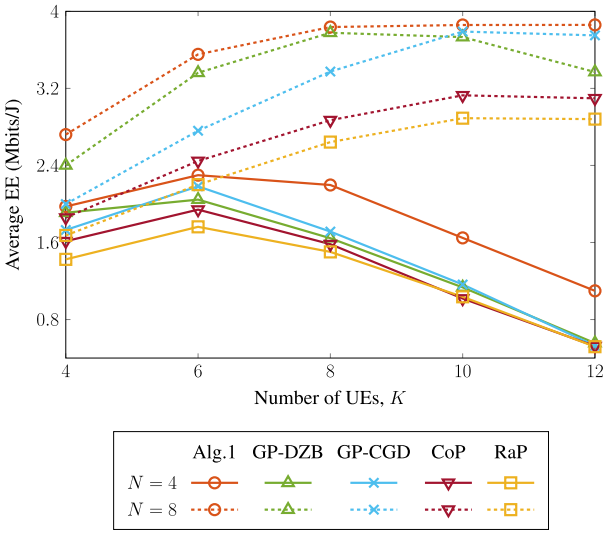


FIGURE 9. Average EE as a function of the number of UEs K at fixed $\bar{R} = 0.5$ bits/s/Hz.

[5]The greedy-pairing methods based on the channel gains primarily depend on the large-scale fading rather than on the small-scale fading, and thus, they are good representatives of classical NOMA schemes with the UE pairing using the statistical CSI.

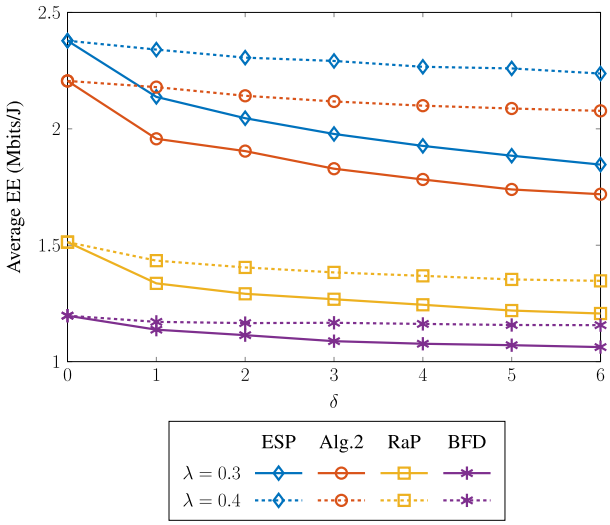


FIGURE 10. Average EE (Mbits/J) as a function of stochastic error level.

Fig. 9 illustrates the behavior of the EE of different pairing schemes when the number of UEs K changes, with a fixed value of the minimum bit rate, $\bar{R} = 0.5$ bits/s/Hz. It is noteworthy that Alg. 1 is generalized to all pairing schemes; therefore, the three greedy-based methods can be straightforwardly employed by changing the values of matrix $\boldsymbol{\alpha}$ in the proposed algorithm. As shown in Fig. 9, Alg. 1 significantly outperforms the other pairing schemes, even in the case where the degree of freedom is insufficient, i.e., with the number of antennas at the BS $N = 4$. This gap quickly enlarges when K increases, starting from $K \geq 6$. With a high N ($N = 8$), while the EE of all the five paring schemes increases in proportion to K, the proposed method is still superior to all greedy-based ones, thereby verifying the adaptability and stability of UE pairing with respect to various numbers of UEs in the cell. These simulation results confirm that the proposed method enables a unified UE pairing framework that is applicable for all greedy-based schemes as well as random pairing. By solving the joint optimization problem, it provides the best performance among the existing pairing schemes under different network topologies. In addition, as afforded by the RaP method, the GP-based and CoP schemes utilize the pairing matrix provided for computing the beamforming vectors, resulting in an equivalent computational complexity in terms of big-O. This validates the advantages of the proposed algorithm over existing methods for UE pairing.

E. EFFECT OF IMPERFECT CSI

In this subsection, we evaluate the effectiveness of the proposed robust design under a practical scenario in the presence of imperfect CSI. Fig. 10 presents the average EE as a function of the stochastic error level δ, where δ is described in (28) with parameters $\lambda \in \{0.3, 0.4\}$. For simplicity, we adopt the three aforementioned schemes, i.e., the ESP, RaP and BFD schemes.

As shown in Fig. 10, the average EE of all four cases with $\lambda = 0.4$ is lower than those with $\lambda = 0.3$. This shows that the system performance degrades as the accuracy of channel estimation decreases, i.e., with $\lambda = 0.4$. Furthermore, this corresponds to a larger estimation error compared with $\lambda = 0.3$. Recall that $\delta = 0$ implies perfect CSI cases. Hence, one obtains the same EE values of each scheme at $\delta = 0$ for any value of λ. For each considered scheme, a gap between two cases, i.e., $\lambda = 0.3$ (solid lines) and $\lambda = 0.4$ (dashed lines), is observed to increase with the growth of δ. In this practical scenario, the proposed scheme performs significantly better than the RaP and BFD schemes, while providing an average EE similar to that of the ESP scheme. Hence, this verifies the robustness of the proposed algorithm against channel uncertainty.

VI. CONCLUSION

In this paper, we formulated a generalized framework for a hybrid design of beamforming and NOMA methods in a DL network by constructing a matrix of UE pairing binary variables. In fact, this auxiliary matrix facilitated the optimization problem in terms of EE by reducing the number of binary variables and constraints. Our proposed formulation was devised to yield an optimal UE pairing solution, i.e., it may pair two arbitrary UEs in a cell despite the geographical conditions. Therefore, the entire channel spectrum can be efficiently exploited instead of splitting it into sub-channels as in the conventional method. Given that the EE maximization problem belongs to mixed-integer nonconvex programming, which has scarcely been solved by previous approaches, we thus derived a low-complexity iterative algorithm (Algorithm 1) based on the IA method and Dinkelbach transformation.

Numerical results with realistic parameters were presented to demonstrate the effectiveness of the proposed algorithm. The proposed algorithm achieves a significant EE gain over two benchmark schemes, BFD and RaP, which are the traditional beamforming design and traditional UE pairing for NOMA, respectively, while providing a fast convergence rate and maintaining an acceptable complexity. In addition, the advantages of the proposed algorithm over the existing schemes were validated via simulation results based on the QoS requirement, maximum BS power budget, and number of UEs. For the imperfect CSI scenario, we introduced a robust design developed from the proposed algorithm, namely Algorithm 2, to address a realistic scenario where the channel estimation is inaccurate. An iterative low-complexity algorithm was verified via a stochastic estimation error model to demonstrate its robustness and adaptability over the classical RaP and BFD schemes, while providing an EE performance comparable to the ESP scheme.

APPENDIX A
AUXILIARY FUNCTIONS FOR INNER APPROXIMATIONS

We first introduce the following two approximate functions that are useful for the proofs in Appendices B and C:

- The function $|x|^2/y$ with $(x, y) \in \mathbb{R}^2_{++}$ has a lower bound around a feasible point $(x^{(i)}, y^{(i)})$, as provided in [43], i.e.,

$$\frac{|x|^2}{y} \geq 2\frac{x^{(i)}x}{y^{(i)}} - \frac{|x^{(i)}|^2}{(y^{(i)})^2}y. \tag{A.1}$$

- For a product function $g_{\mathrm{pr}}(x, y) \triangleq xy$, an upper bound of $g_{\mathrm{pr}}(x, y)$ around a feasible point $(x^{(i)}, y^{(i)})$ can be obtained by applying [44, Eq. (B.1)] as

$$g_{\mathrm{pr}}(x, y) \leq \frac{y^{(i)}}{2x^{(i)}}x^2 + \frac{x^{(i)}}{2y^{(i)}}y^2 \triangleq g^{(i)}_{\mathrm{pr}}(x, y). \tag{A.2}$$

APPENDIX B
PROOF OF THEOREM 1

For a nonconvex problem with the nonconcave objective function and nonconvex constraints in the form of problem (20a), it is possible to transform the problem into a successive program with sub-problems having feasible sets by applying the inner convex approximation [45]. In particular, the feasible set in (20a) is nonconvex owing to the nonconvexity of (18). Therefore, the remainder of the proof is to address the nonconvex constraint (18). First, constraint (18) with the minimum function $\gamma_k(\mathbf{w}, \boldsymbol{\alpha})$ on the left-hand side (LHS) can be expressed as two sub-constraints:

$$\begin{cases} \gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \omega_k, & \forall k \in \mathcal{K}, \quad \text{(B.1a)} \\ \gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \omega_k, & \forall \ell, k \in \mathcal{K}, \ell \neq k. \quad \text{(B.1b)} \end{cases}$$

Subsequently, by applying (A.1) to the LHSs of (3) and (3), the lower bounds of $\gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha})$, $\forall k \in \mathcal{K}$ and $\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$, $\forall \ell, k \in \mathcal{K}$, $\ell \neq k$ at iteration $(i+1)$ are formulated as

$$\begin{cases} \gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{k,k}(\mathbf{w}|\mathbf{H}, \Psi^{(i)}_k) \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{H}, \Psi^{(i)}_k, \Upsilon_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{H})), \quad \text{(B.2a)} \\ \gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{\ell,k}(\mathbf{w}|\mathbf{H}, \Phi^{(i)}_{\ell,k}) \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{H}, \Phi^{(i)}_{\ell,k}, \Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{H})), \quad \text{(B.2b)} \end{cases}$$

where the nonconvex function $\tilde{\mathcal{A}}^{(i)}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A}, C, \Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}))$ is expressed as

$$\begin{aligned} &\tilde{\mathcal{A}}^{(i)}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A}, C, \Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A})) \\ &\quad \triangleq \frac{|\mathbf{a}^H_\ell \mathbf{w}^{(i)}_k|^2}{C^2}\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A}), \quad \text{(B.3)} \\ &\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A}) \\ &\quad \triangleq \begin{cases} \sum_{\forall k' \in \mathcal{K}\setminus\{k\}}(1-\alpha_{\ell,k'})|\mathbf{a}^H_\ell \mathbf{w}_{k'}|^2 \\ \qquad\qquad +\sigma^2_\ell, \quad \text{if } \ell = k, \\ \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} \alpha_{\ell,k}|\mathbf{a}^H_\ell \mathbf{w}_{k'}|^2, \\ \qquad\qquad +\alpha_{\ell,k}\sigma^2_\ell, \quad \text{if } \ell \neq k. \end{cases} \quad \text{(B.4)} \end{aligned}$$

Clearly, the nonconvexity of $\tilde{\mathcal{A}}^{(i)}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\mathbf{A}, C, \Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}))$ is caused by the nonconvexity of $\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$. Therefore, $\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ must be convexified. In this regard, the upper bounds of $\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ corresponding to cases in (B.3) are given as

$$\tilde{\Upsilon}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}) \triangleq \begin{cases} \sigma^2_\ell + \sum_{\forall k' \in \mathcal{K}\setminus\{k\}}(1-\alpha_{\ell,k'})\mu_{\ell,k'}, \\ \qquad\qquad \text{if } \ell = k, \\ \alpha_{\ell,k}\sigma^2_\ell + \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} \alpha_{\ell,k}\mu_{\ell,k'}, \\ \qquad\qquad \text{if } \ell \neq k, \end{cases} \tag{B.5}$$

with an SOC constraint imposed as:

$$|\mathbf{h}^H_\ell \mathbf{w}_k|^2 \leq \mu_{\ell,k}, \ \forall \ell, k \in \mathcal{K}. \tag{B.6}$$

By applying (A.2) to (B.5), the upper bound of $\Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ can be derived as

$$f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu}) \triangleq \begin{cases} \sigma^2_\ell + \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} g^{(i)}_{\mathrm{pr}}(1-\alpha_{\ell,k'}, \mu_{\ell,k'}), \\ \qquad\qquad \text{if } \ell = k, \\ \alpha_{\ell,k}\sigma^2_\ell + \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} g^{(i)}_{\mathrm{pr}}(\alpha_{\ell,k}, \mu_{\ell,k'}), \\ \qquad\qquad \text{if } \ell \neq k. \end{cases} \tag{B.7}$$

Finally, the LHSs of (3) and (3) are linearly approximated as

$$\begin{cases} \gamma_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{k,k}(\mathbf{w}|\mathbf{H}, \Psi^{(i)}_k) \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}|\mathbf{H}, \Psi^{(i)}_k, f_{k,k}(\boldsymbol{\alpha}, \boldsymbol{\mu})), \quad \text{(B.8a)} \\ \gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{\ell,k}(\mathbf{w}|\mathbf{H}, \Phi^{(i)}_{\ell,k}) \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}|\mathbf{H}, \Phi^{(i)}_{\ell,k}, f_{\ell,k}(\boldsymbol{\alpha}, \boldsymbol{\mu})). \quad \text{(B.8a)} \end{cases}$$

This results in constraints (21d) and (21e) with constraint (B.6) expressed as (21c), thereby completing the proof.

APPENDIX C
PROOF OF PROPOSITION 1

To convexify problem (34a), we must address the feasible set and objective function. As shown, the feasible set of problem (34a) is nonconvex owing to the nonconvex constraint (34c), whereas the objective function performed as a ratio function of sum-log over quadratic form is nonconcave.

First, we address constraint (34c). Similar to (3), we can rewrite (34c) as

$$\begin{cases} \hat{\gamma}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \omega_k, & \forall k \in \mathcal{K}, \quad \text{(C.1a)} \\ \hat{\gamma}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \omega_k, & \forall \ell, k \in \mathcal{K}, \ell \neq k. \quad \text{(C.1b)} \end{cases}$$

By adopting (3), the lower bounds for the LHSs of (B.8) can be derived as

$$\begin{cases} \hat{\gamma}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{k,k}(\mathbf{w}|\hat{\mathbf{H}}, \hat{\Psi}^{(i)}_k) \\ \qquad\qquad -\Xi_{k,k}(\mathbf{w}^{(i)}|\hat{\mathbf{H}}, \hat{\Psi}^{(i)}_k)\sum_{\forall \ell \in \mathcal{K}} \epsilon_k \|\mathbf{w}_\ell\|^2 \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\hat{\mathbf{H}}, \hat{\Psi}^{(i)}_k, \Upsilon_{k,k}(\mathbf{w}, \boldsymbol{\alpha}|\hat{\mathbf{H}})), \quad \text{(C.2a)} \\ \hat{\gamma}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \geq \mathcal{A}^{(i)}_{\ell,k}(\mathbf{w}|\hat{\mathbf{H}}, \hat{\Phi}^{(i)}_{\ell,k}) \\ \qquad\qquad -\Xi_{\ell,k}\big(\mathbf{w}^{(i)}|\hat{\mathbf{H}}, \hat{\Phi}^{(i)}_{\ell,k}\big)\alpha_{\ell,k}\sum_{\forall k' \in \mathcal{K}} \epsilon_\ell \|\mathbf{w}_{k'}\|^2 \\ \qquad\qquad -\tilde{\mathcal{A}}^{(i)}_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\hat{\mathbf{H}}, \hat{\Phi}^{(i)}_{\ell,k}, \Upsilon_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}|\hat{\mathbf{H}})), \quad \text{(C.2b)} \end{cases}$$

By applying (A.2) and the same steps (B.5)–(B.8), the LHSs of (B.8) can be convexified as

$$\begin{cases} \hat{\gamma}_{k,k}(\mathbf{w},\boldsymbol{\alpha}) \geq \mathcal{A}_{k,k}^{(i)}(\mathbf{w}|\hat{\mathbf{H}},\hat{\Psi}_k^{(i)}) \\ \qquad\qquad -\Xi_{k,k}\big(\mathbf{w}^{(i)}|\hat{\mathbf{H}},\hat{\Psi}_k^{(i)}\big)\sum_{\forall k'\in\mathcal{K}}\epsilon_k\tau_{k'} \\ \qquad -\bar{\mathcal{A}}_{k,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\mu}|\hat{\mathbf{H}},\hat{\Psi}_k^{(i)},f_{k,k}(\boldsymbol{\alpha},\boldsymbol{\mu})), \qquad \text{(C.3a)} \\ \hat{\gamma}_{\ell,k}(\mathbf{w},\boldsymbol{\alpha}) \geq \mathcal{A}_{\ell,k}^{(i)}(\mathbf{w}|\hat{\mathbf{H}},\hat{\Phi}_{\ell,k}^{(i)}) \\ \qquad\qquad -\Xi_{\ell,k}\big(\mathbf{w}^{(i)}|\hat{\mathbf{H}},\hat{\Phi}_{\ell,k}^{(i)}\big)\sum_{\forall k'\in\mathcal{K}}\epsilon_\ell g_{\mathrm{pr}}^{(i)}(\alpha_{\ell,k},\tau_{k'}) \\ \qquad -\bar{\mathcal{A}}_{\ell,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\mu}|\hat{\mathbf{H}},\hat{\Phi}_{\ell,k}^{(i)},f_{\ell,k}(\boldsymbol{\alpha},\boldsymbol{\mu})), \qquad \text{(C.3b)} \end{cases}$$

with newly introduced variables $\boldsymbol{\mu} \triangleq [\mu_{\ell,k}]_{\forall\ell,k\in\mathcal{K}}$ and $\boldsymbol{\tau} \triangleq [\tau_k]_{\forall k\in\mathcal{K}}$, thereby satisfying the following convex constraints:

$$|\hat{\mathbf{h}}_\ell^H\mathbf{w}_k|^2 \leq \mu_{\ell,k},\ \forall\ell,k\in\mathcal{K}, \qquad \text{(C.3a)}$$
$$\|\mathbf{w}_k\|^2 \leq \tau_k,\forall k\in\mathcal{K}. \qquad \text{(C.3b)}$$

It is clear that (B.8) and (C.3a) results in constraints (35c) – (35f) for problem (35a). Finally, the objective function (34a) can be straightforwardly addressed using (22) – (24a), thereby yielding problem (35a).

REFERENCES

[1] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 179–189, 2020.

[2] M. Vaezi, G. A. A. Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 4, pp. 900–919, Dec. 2019.

[3] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, 4th Quart., 2020.

[4] M. T. P. Le, G. C. Ferrante, T. Q. S. Quek, and M. Di Benedetto, "Fundamental limits of low-density spreading NOMA with fading," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4648–4659, Jul. 2018.

[5] M. T. P. Le, G. C. Ferrante, G. Caso, L. De Nardis, and M. Di Benedetto, "On information-theoretic limits of code-domain NOMA for 5G," *IET Commun.*, vol. 12, no. 15, pp. 1864–1871, Sep. 2018.

[6] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[7] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[8] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[9] W. Liang, Z. Ding, Y. Li, and L. Song, "User pairing for downlink non-orthogonal multiple access networks using matching algorithm," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5319–5332, Dec. 2017.

[10] A. Zakeri, A. Khalili, M. R. Javan, N. Mokari, and E. Jorswieck, "Robust energy-efficient resource management, SIC ordering, and beamforming design for MC MISO-NOMA enabled 6G," *IEEE Trans. Signal Process.*, vol. 69, pp. 2481–2498, 2021.

[11] X. Chen, F.-K. Gong, G. Li, H. Zhang, and P. Song, "User pairing and pair scheduling in massive MIMO-NOMA systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 788–791, Apr. 2018.

[12] L. Zhu, J. Zhang, Z. Xiao, X. Cao, and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 328–331, Apr. 2019.

[13] Y. Cheng, K. H. Li, K. C. Teh, and S. Luo, "Joint user pairing and subchannel allocation for multisubchannel multiuser nonorthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8238–8248, Sep. 2018.

[14] H. Zhang, D.-K. Zhang, W.-X. Meng, and C. Li, "User pairing algorithm with SIC in non-orthogonal multiple access system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[15] V.-P. Bui, P. Nguyen, H. Nguyen, V.-D. Nguyen, and O.-S. Shin, "Optimal user pairing for achieving rate fairness in downlink NOMA networks," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2019, pp. 575–578.

[16] T.-V. Nguyen, V.-D. Nguyen, D. B. da Costa, and B. An, "Hybrid user pairing for spectral and energy efficiencies in multiuser MISO-NOMA networks with SWIPT," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4874–4890, Aug. 2020.

[17] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.

[18] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O.-S. Shin, "Joint power control and user association for NOMA-based full-duplex systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8037–8055, Nov. 2019.

[19] D. Kudathanthirige and G. A. A. Baduge, "NOMA-aided multicell downlink massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 612–627, Jun. 2019.

[20] M. T. P. Le, L. Sanguinetti, E. Bjornson, and M.-G.-D. Benedetto, "Code-domain NOMA in massive MIMO: When is it needed?" *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4709–4723, May 2021.

[21] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. Xia, "Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5065–5079, Nov. 2019.

[22] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[23] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, J. Tang, A. G. Burr, Z. Ding, and O. A. Dobre, "Spectral-energy efficiency trade-off-based beamforming design for MISO non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6593–6606, Oct. 2020.

[24] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2019.

[25] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10231–10236, Dec. 2016.

[26] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2668–2673.

[27] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[28] F. Alavi, K. Cumanan, M. Fozooni, Z. Ding, S. Lambotharan, and O. A. Dobre, "Robust energy-efficient design for MISO non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7937–7949, Nov. 2019.

[29] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[30] S. Tombaz, A. Vastberg, and J. Zander, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.

[31] V. H. Nguyen, D. Nguyen, O. A. Dobre, S. K. Sharma, S. Chatzinotas, B. Ottersten, and O.-S. Shin, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[32] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

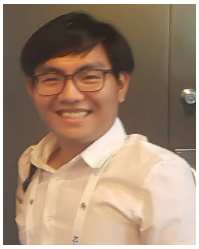
[33] J. Lee and S. Leyffer, *Mixed Integer Nonlinear Programming*. New York, NY, USA: Springer, 2012.

[34] M. Tawarmalani and N. V. Sahinidis, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*, vol. 65. Dordrecht, The Netherlands: Kluwer, 2002.

[35] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programms," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Jul. 1978.
[36] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.
[37] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, Y. Wu, and O.-S. Shin, "Joint antenna array mode selection and user assignment for full-duplex MU-MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 2946–2963, Jun. 2019.
[38] Y. Labit, D. Peaucelle, and D. Henrion, "SEDUMI INTERFACE 1.02: A tool for solving LMI problems with SEDUMI," in *Proc. IEEE Int. Symp. Comput. Aided Control Syst. Design*, Oct. 2002, pp. 272–277.
[39] Q. Li and L. Yang, "Robust optimization for energy efficiency in MIMO two-way relay networks with SWIPT," *IEEE Syst. J.*, vol. 14, no. 1, pp. 196–207, Mar. 2020.
[40] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, Apr. 2017.
[41] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[42] A. J. Maurer, J. Jalden, D. Seethaler, and G. Matz, "Vector perturbation precoding revisited," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 315–328, Jan. 2011.
[43] V.-D. Nguyen, T. Q. Duong, H. D. Tuan, O.-S. Shin, and H. V. Poor, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2220–2233, May 2017.
[44] V.-D. Nguyen, H. V. Nguyen, O. A. Dobre, and O.-S. Shin, "A new design paradigm for secure full-duplex multiuser systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, Jul. 2018.
[45] H. Tuy, *Convex Analysis and Global Optimization*, 2nd ed. Cham, Switzerland: Springer, 2016.

KHA-HUNG NGUYEN received the B.E. degree in electronic and telecommunication from the Posts and Telecommunications Institute of Technology, Hanoi, Vietnam, in 2019, and the M.S. degree in electrical engineering from Soongsil University, South Korea, in 2021.

From 2019 to 2021, he was with the Wireless Communication Laboratory, Soongsil University. His research interest includes wireless communications, with particular focus on optimization techniques, 5G systems, beamforming design, NOMA, and IRS communications.



HIEU V. NGUYEN (Member, IEEE) received the B.E. degree in electronics and telecommunications from The University of Danang - University of Science and Technology (UD-DUT), Vietnam, in 2011, and the M.E. and Ph.D. degrees in electronic engineering from Soongsil University, Seoul, South Korea, in 2016 and 2020, respectively.

From 2011 to 2013, he was with the UD-DUT as an Assistant Researcher and a Lecturer. Since 2014, he has been with the Wireless Communications Laboratory, Soongsil University. He is currently a Postdoctoral Researcher with the School of Electronic Engineering, Soongsil University. His research interests include wireless communications, with particular focus on optimization techniques and machine learning for wireless communications, such as UAV/drones communications, device-to-device communications, full-duplex radios, green communication systems, the IoT, and satellite networks.

MAI T. P. LE received the Ph.D. degree from the Sapienza University of Rome, Rome, Italy, in February 2019.

From 2015 to 2020, she was a Ph.D. Student and a Postdoctoral Researcher with the Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome. In 2016, she was a Visiting Researcher with the Singapore University of Technology and Design, Singapore, and in 2012, with Arizona State University, Tempe, AZ, USA. Since 2011, she has been with the Department of Electronics and Telecommunications, The University of Danang - University of Science and Technology, Da Nang, Vietnam, where she is currently a Lecturer. Her main research interests include information theory, mathematical theories, and their application in wireless communications. Her current research focuses on physical layer techniques for beyond 5G networks and THz communication.

LUCA SANGUINETTI (Senior Member, IEEE) received the Laurea degree *(cum laude)* in telecommunications engineering and the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2002 and 2005, respectively.

In 2004, he was a Visiting Ph.D. Student with the German Aerospace Center, Oberpfaffenhofen, Germany. From June 2007 to June 2008, he was a Postdoctoral Associate with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. From July 2013 to October 2017, he was with the Large Systems and Networks Group, Centrale Supelec, France. He is currently an Associate Professor with the Dipartimento di Ingegneria dell'Informazione, University of Pisa. He has coauthored two textbooks *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017) and *Foundations of User-Centric Cell-Free Massive MIMO* (2021). His research interests include communications and signal processing.

Dr. Sanguinetti is a member of the Executive Editorial Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a recipient of the 2018 Marconi Prize Paper Award in Wireless Communications and coauthored a paper that received the Young Best Paper Award from the ComSoc/VTS Italy Section. He was a co-recipient of two Best Conference Paper awards, the IEEE WCNC 2013 and the IEEE WCNC 2014. He was also a recipient of the FP7 Marie Curie IEF 2013, Dense Deployments for Green Cellular Networks. He was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE SIGNAL PROCESSING LETTERS, the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS Special Issue on Game Theory for Networks, and an Associate Editor of the IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS, series on Green Communications and Networking. He is also an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.

OH-SOON SHIN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 1998, 2000, and 2004, respectively.

From 2004 to 2005, he was a Postdoctoral Fellow with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. From 2006 to 2007, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea. In 2007, he joined the School of Electronic Engineering, Soongsil University, Seoul, where he is currently a Professor. His research interests include communication theory, wireless communication systems, and signal processing for communication.