energies

MDPI

Article

Rate Fairness and Power Consumption Optimization for NOMA-Assisted Downlink Networks

Kha-Hung Nguyen [1,†], Hieu V. Nguyen [1,†], Mai T. P. Le [2], Tuan X. Cao [3] and Oh-Soon Shin [1,*]

[1] School of Electronic Engineering and Department of Information Communication Convergence Technology, Soongsil University, Seoul 06978, Korea; hungnkptit@gmail.com (K.-H.N.); hieuvnguyen@soongsil.ac.kr (H.V.N.)
[2] Faculty of Electronics and Telecommunication Engineering, The University of Danang-University of Science and Technology, Danang 550000, Vietnam; lpmai@dut.udn.vn
[3] Department of Office Administration, The University of Danang, Danang 550000, Vietnam; cxtuan@ac.udn.vn
* Correspondence: osshin@ssu.ac.kr
† These authors contributed equally to this work.

Abstract: This paper considers a non-orthogonal multiple access (NOMA) downlink network, where a hybrid of NOMA and beamforming designs is developed to enhance the channel capacity. We aim to improve the system performance in terms of rate fairness and power consumption. Hence, a multi-objective problem with a joint optimization of user equipment pairing, power control, and quality-of-service requirements is addressed. To efficiently solve the problem, we propose two low-complexity algorithms based on the inner-approximation method, with the first algorithm using the relaxation method and the second one using graph theory. Numerical results are provided to demonstrate the effectiveness of the two proposed algorithms in comparison with the exhaustive search and existing methods.

Keywords: bipartite graph; inner approximation; non-orthogonal multiple access; power consumption; rate fairness

check for updates

Citation: Nguyen, K.H.; Nguyen, H.V.; Le, M.T.P.; Cao, T.X.; Shin, O.H. Rate Fairness and Power Consumption Optimization for NOMA-Assisted Downlink Networks. *Energies* **2021**, *14*, 58. https://dx.doi.org/10.3390/en14010058

Received: 19 November 2020
Accepted: 18 December 2020
Published: 24 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Non-orthogonal multiple access (NOMA) can be considered to be a promising multiple access technology for fifth-generation networks for accommodating the rapid growth of data traffic [1–4]. Unlike orthogonal multiple access, NOMA allows user equipment (UE) to share the same time-frequency resource; hence, the spectral efficiency (SE) and energy efficiency (EE) are improved. The NOMA technique is classified into two main categories: the power domain [5] and code domain [6]. Herein, we focus power-domain NOMA, referred to as NOMA for convenience. The key idea of power-domain NOMA is to allocate a higher power to the UE with poorer channel gain. Meanwhile, a UE with a strong channel gain is equipped with successive interference cancellation (SIC) to remove the signal intended to the UE with poorer channel gain before decoding its own signal [5,7].

1.1. Related Works and Motivation

Recently, many new approaches have been proposed for UE grouping/clustering [8–11]. The number of clusters is selected based on the cell size and channel gain in [8], whereas a clustering strategy based on channel correlation was investigated in [11]. In [9], a tensor model was proposed to manage UE association, in which clusters of UEs were formed by different circular zones. Owing to the channel conditions and computational capacity of UE, SIC is restricted to limited groups/clusters of UE [8,9].

To ease implementationin a small-sized network, UE pairing is essential. In [12], pairing of one near UE and another far UE was shown to be efficient. A cognitive radio inspired NOMA system with a single-antenna base station (BS) was also investigated in [12], where a cell is divided into inner and outer zones, including cognitive UEs (CUs)

with strong channel gain and primary UEs (PUs) with poor channel gain. A two-zone scheme assisted by the multiple-antenna technique was studied in [13], in which each pair comprised one inner-zone UE and the other outer-zone UE. The authors in [14] proposed a distance-based UE pairing scheme to improve the sum throughput, in which the far UEs were paired with an available nearest UEs using a distance threshold. Another pairing scheme based on the vector norms of channel responses was developed for a massive MIMO-NOMA system in [15]. The effect of UE pairing on the performance of NOMA-based systems was studied in [16], in which a CR-NOMA and an F-NOMA system were considered under fixed power allocation.

To fully exploit the advantages of UE pairing, many algorithms have been developed for pairing two arbitrary UEs without separate zones. In addition to random pairing scheme used in [17], UE pairing schemes based on channel condition arrangement have been proposed [17–19]. The typical pre-processing step is to place all UE into an unpaired set with sorted channel gains (SCG). Under the setting, various strategies have been proposed to pop and pair two UEs until the SCG is empty or only contains one UE. In [19], two UEs with the best and worst channel gains were popped and paired consecutively. In [18], two best-channel conditions were selected, whereas the first and middle UEs in the SCG were paired together in [17]. Unlike the aforementioned work, a more flexible approach was proposed [20], where any two users in a cell can be dynamically paired together to improve the fairness throughput among all UEs.

Most of the aforementioned works investigated NOMA-based systems to improve SE. In power-domain NOMA, power control is extremely challenging owing to SIC. In addition, the power consumption or EE has been considered to be one of the immediate concerns for the upcoming wireless networks, as indicated by the energy for information and communication technologies, which constitute approximately five percents of the total world's energy consumption [21,22]. Therefore, power consumption has been recently reported in a few studies. The authors in [10] investigated a joint optimization problem of user assignment and power allocation for a downlink (DL) NOMA network in terms of EE maximization. Another power allocation strategy was proposed in [23] to evaluate the EE of a broadcast DL NOMA network. These studies investigated the sub-channels assignment for UEs; hence, the UE association is still unknown for exploiting the channel capacity using NOMA. This motivates us to develop algorithms with a joint UE pairing and beamforming design that accounts for both the SE fairness and power consumption.

1.2. Contributions

In this paper, we consider a cellular DL NOMA network, where a multi-objective problem for rate fairness among all UEs and power consumption were investigated under the minimum per-UE rate requirement and signal-to-noise ratio (SNR) constraints for decoding the message. To further improve the channel capacity, the NOMA technique can be applied to two UEs in a cell, instead of UE pairing with two near and far groups. The resulting problem belongs to a mixed-integer non-convex class that is difficult to solve. Therefore, we devise two low-complexity algorithms, based on inner approximation (IA) and graph theory. The main contributions of this paper are described as follows:

- First, we investigate a hybrid system comprising NOMA and conventional beamforming to reap their advantages for a unique design. A beamforming design with SIC can be applied to a pair of UE, or conventional beamforming can be used for certain UE depending on the number of UEs and channel conditions. Accordingly, the approach is referred to as dynamic pairing because it enables the pairing of two arbitrary UEs in a cell.
- To manage the dynamic pairing, we introduce a new binary variable matrix for UE association. By using the NOMA principle and sorted channel gains, the computational complexity is reduced with the variables forced into an upper triangular matrix. Next, we formulate a problem for joint optimization of UE pairing and beamforming design, thereby enabling the selection of the objective function: the max-min rate

(MMR) and power consumption. However, a strong coupling of binary and continuous variables under non-convex constraints results in a hard class of mixed-integer non-convex programs.

- To efficiently solve the resulting problem, we propose two approaches based on the IA framework as well as the relaxation method and bipartite graph. The first approach is based on a combination of the relaxation and IA methods; it can transform the complicated original problem into a more tractable form without integer variables. Therefore, an iterative algorithm based on the IA framework is derived to obtain a local optimal solution at the least. To further reduce the time complexity, we develop an algorithm in which UE pairing and the beamforming design are managed using a bipartite graph and the IA method. For UE pairing, this approach uses the channel correlation as the weights of edges on the graph, resulting in a linear bottleneck assignment problem. Hence, we apply the binary search and augmenting path algorithm to obtain the UE pair, and then use the solution of the first approach for the beamforming design based on a specified integer variable value.

1.3. Paper Organization and Notation

The rest of this paper is organized as follows. The system model and problem formulation are described in Sections 2 and 3, respectively. Two proposed solutions based on the IA framework with the relaxation method and graph theory are detailed in Sections 4 and 5, respectively. For the benchmark purpose, the optimal solution based on exhaustive search is also studied in Section 5. Section 6 presents numerical results, whereas Section 7 concludes the paper.

Notation: Lowercase and uppercase letters in bold indicate vectors and matrices, respectively. $\mathbf{x}^H$ is the Hermitian transpose of $\mathbf{x}$. The set of all complex and real numbers are represented by $\mathbb{C}$ and $\mathbb{R}$, respectively. $\|\cdot\|$ stands for the Euclidean norm. $\mathbb{E}\{\cdot\}$ is the expectation and $\Re\{.\}$ denotes the real part of a complex number.

2. System Model

We consider a DL NOMA cellular system, where a BS is equipped N transmit antennas to serve K single-antenna UEs. As illustrated in Figure 1, NOMA technique can be applied to a pair of two arbitrary UEs in a set of UEs, denoted by $\mathcal{K} \triangleq \{1, \ldots, K\}$. For simplicity, we denotes the k-th UE by UE_k.

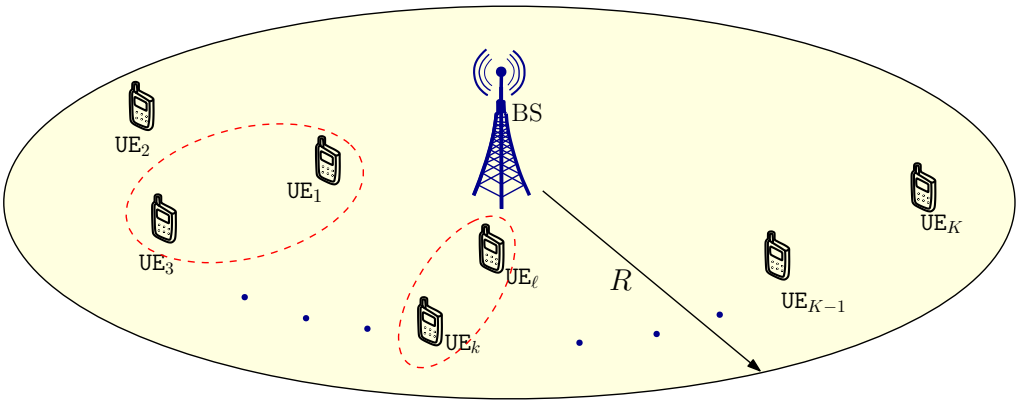

Figure 1. System model.

2.1. Data Transmission Model

The DL channel from BS to UE_k and the corresponding beamforming vector are denoted as $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ and $\mathbf{w}_k \in \mathbb{C}^{N\times 1}$, respectively. Without loss of generality, the channel gains of all UEs are sorted as $\|\mathbf{h}_1\|^2 \geq \|\mathbf{h}_2\|^2 \geq \cdots \geq \|\mathbf{h}_K\|^2$. The signal received at UE_k can be written as

$$y_k = \sum_{\forall \ell \in \mathcal{K}} \mathbf{h}_k^H \mathbf{w}_\ell x_\ell + n_k, \tag{1}$$

where x_k with $\mathbb{E}\{|x_k|^2\} = 1$ represents the symbol intended for UE_k, and n_k represents the additive white Gaussian noise (AWGN) at UE_k. To perform UE pairing, we introduce binary variables $\alpha_{k,\ell} \in \{0,1\}, \forall k, \ell \in \mathcal{K}$, which are expressed as

$$\alpha_{k,\ell} = \begin{cases} 1, & \text{if } \mathrm{UE}_k \text{ and } \mathrm{UE}_\ell \text{ are paired with the channel gains satisfying } \|\mathbf{h}_k\|^2 \geq \|\mathbf{h}_\ell\|^2, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

By defining $\boldsymbol{\alpha} \triangleq [\alpha_{k,\ell}]_{K\times K}$, it is clear that $\boldsymbol{\alpha}$ is an upper triangle matrix with the main diagonal of zeros. The signal-to-interference-plus-noise ratio (SINR) at UE_k can be written as

$$\gamma_k(\mathbf{w}, \boldsymbol{\alpha}) = \min\left\{\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha}), \min_{\forall \ell \in \mathcal{K}}\{\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})\}\right\}, \tag{3}$$

where $\mathbf{w} \triangleq [\mathbf{w}_k]_{k\in\mathcal{K}}$. The SINRs for decoding the UE_k's signals at itself and UE_ℓ are denoted by $\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha})$ and $\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$, respectively, which can be expressed as

$$\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha}) = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\chi_k(\mathbf{w}, \boldsymbol{\alpha})}, \tag{4a}$$

$$\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) = \frac{|\mathbf{h}_\ell^H \mathbf{w}_k|^2}{\alpha_{\ell,k}\Phi_{\ell,k}(\mathbf{w})}, \tag{4b}$$

where $\chi_k(\mathbf{w}, \boldsymbol{\alpha})$ and $\Phi_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})$ are respectively given as

$$\Phi_k(\mathbf{w}, \boldsymbol{\alpha}) = \sum_{\forall \ell \in \mathcal{K}\setminus\{k\}} (1 - \alpha_{k,\ell})|\mathbf{h}_k \mathbf{w}_\ell|^2 + \sigma_k^2, \tag{5a}$$

$$\Theta_{\ell,k}(\mathbf{w}) = \sum_{\forall k' \in \mathcal{K}\setminus\{k\}} |\mathbf{h}_\ell \mathbf{w}_{k'}|^2 + \sigma_\ell^2. \tag{5b}$$

It is seen that when UE_k and UE_ℓ are not paired together, i.e., $\alpha_{k,\ell} = \alpha_{\ell,k} = 0, \forall k, \ell \in \mathcal{K}$, only $\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha})$ is involved in Equation (3), since $\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha}) \to \infty$.

2.2. Power Consumption Model

To measure power-efficiency, we consider the power consumption for transmission with beamforming, which can be expressed as

$$P_{\mathrm{T}}(\mathbf{w}) = \underbrace{\sum_{\forall k \in \mathcal{K}} \frac{1}{\zeta_{\mathrm{PA}}} \|\mathbf{w}_k\|^2}_{\text{radiated power}}, \tag{6}$$

where ζ_{PA} is the efficiency of the power amplifier (PA) at the BS. In other words, the power consumption is the power for data transmission from the BS to all the UE in the form of electromagnetic wave radiation. It is clear that $P_{\mathrm{T}}(\mathbf{w})$ is a function of beamforming vectors.

Remark 1. *In addition to the power for beamforming, the power consumption in a wireless communication system includes the power for signal processing, device activation, and circuit operation [24,25]. However, these values are constant, and hence can be omitted in the objective function without affecting the search for the optimal solution. For convenience, we consider only the power consumption for beamforming in this paper.*

3. Problem Formulation

The achievable rate of UE_k in nats/s/Hz can be expressed as

$$R(\gamma_k(\mathbf{w}, \boldsymbol{\alpha})) = \ln(1 + \gamma_k(\mathbf{w}, \boldsymbol{\alpha})), \quad k \in \mathcal{K}. \tag{7}$$

We introduce a parameter $v \in \{0,1\}$ for the objective selection between rate fairness and power consumption optimization. As a result, a multi-objective problem can be formulated as

$$\max_{\mathbf{w},\boldsymbol{\alpha}} \quad v \min_{\forall k \in \mathcal{K}} R(\gamma_k(\mathbf{w},\boldsymbol{\alpha})) - (1-v)P_{\mathrm{T}}(\mathbf{w}), \tag{8a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} \|\mathbf{w}_k\|^2 \le P_{\mathrm{BS}}^{\max}, \tag{8b}$$

$$\alpha_{k,\ell} \in \{0,1\}, \quad \forall k,\ell \in \mathcal{K}, \tag{8c}$$

$$\alpha_{k,\ell} = 0, \quad \forall (k,\ell) \in \{\mathcal{K}\times\mathcal{K} | k \ge \ell\}, \tag{8d}$$

$$\sum_{k\in\mathcal{K}} \alpha_{k,\ell} \le 1, \quad \forall \ell \in \mathcal{K}, \tag{8e}$$

$$\sum_{\ell\in\mathcal{K}} \alpha_{k,\ell} \le 1, \quad \forall k \in \mathcal{K}, \tag{8f}$$

$$\alpha_{k,\ell} + \sum_{k'\in\mathcal{K}} \alpha_{\ell,k'} \le 1 \quad \forall (k,\ell) \in \{\mathcal{K}\times\mathcal{K} | k < \ell\}, \tag{8g}$$

$$\alpha_{k,\ell} + \sum_{\ell'\in\mathcal{K}} \alpha_{\ell',k} \le 1 \quad \forall (k,\ell) \in \{\mathcal{K}\times\mathcal{K} | k < \ell\}, \tag{8h}$$

$$|\mathbf{h}_k^H \mathbf{w}_k|^2 \ge \rho_{\mathrm{sn}} \sigma_k^2, \tag{8i}$$

$$R(\gamma_k(\mathbf{w},\boldsymbol{\alpha})) \ge (1-v)\bar{R}_k, \quad \forall k \in \mathcal{K}. \tag{8j}$$

Objective Equation (8a) can be selected for fairness rate maximization (or power consumption minimization) by setting $v = 1$ (or $v = 0$). Constraints Equation (8c–h) correspond to the NOMA criteria for user pairing. Specifically, Equation (8d) ensures that in a certain pair, UE with the better channel employs SIC to remove the interference from the remaining UE. Constraints Equation (8e–h) guarantee that each UE can be paired with at most one other UE. The signal-to-noise ratio (SNR) threshold ρ_{sn} in Equation (8i) is used for detecting the signal at the receivers, whereas constraint Equation (8j) provides the minimum bit rate, $\bar{R}_k$, such that UEs satisfy the QoS requirement. Owing to a strong coupling of continuous and binary variables under the non-convex objective function and constraints, Equation (8) is a mixed-integer non-convex problem, and thus, it is hard to obtain a globally optimal solution.

Remark 2. *Although the proposed dynamic pairing approach can further enhance the system performance, it leads to a large number of possible pairing cases. To overcome this issue and to accelerate the application of NOMA technique, the rest of the paper will devise two algorithms which provide low complexity and fast convergence.*

4. Proposed Solution Based on Relaxation Method

To convert Equation (8) to a tractable form, the binary variables in constraint Equation (8c) are relaxed to continuous ones as $0 \le \alpha_{k,\ell} \le 1, \forall k,\ell \in \mathcal{K}$. Then, the optimization problem Equation (8) can be rewritten as

$$\max_{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\lambda},\eta} \quad v\eta - (1-v)P_{\mathrm{T}}(\mathbf{w}), \tag{9a}$$

$$\text{s.t.} \quad 0 \le \alpha_{k,\ell} \le 1, \quad \forall k,\ell \in \mathcal{K}, \tag{9b}$$

$$R(\gamma_k(\mathbf{w},\boldsymbol{\alpha})) \ge \lambda_k, \quad \forall k \in \mathcal{K}, \tag{9c}$$

$$\lambda_k \ge (1-v)\bar{R}_k, \quad \forall k \in \mathcal{K}, \tag{9d}$$

$$\eta \ge \lambda_k, \quad \forall k \in \mathcal{K}, \tag{9e}$$

$$\textit{Equation}\ (8b),\ \textit{Equation}\ (8d\text{–}i), \tag{9f}$$

where η is a new variable representing the minimum rate among all UEs and $\boldsymbol{\lambda} \triangleq \{\lambda_1, \lambda_2, \ \ldots, \ \lambda_K\}$ denotes the rate of each UE. Furthermore, Equation (9d) is the QoS constraint, which is equivalent to Equation (8j).

It can be observed that constraints Equations (8i) and (9c) are non-convex and thus difficult to solve directly. To deal with the issue, we first introduce some useful IA functions as follows.

- We consider the multiplication function $f_{\text{mul}}(x,z) = xz$. Then an upper bound of $f_{\text{mul}}(x,z)$ is provided in Equation (B.1) in [26]:

$$f_{\text{mul}}(x,z) \leq \frac{x^{(i)}}{2z^{(i)}}z^2 + \frac{z^{(i)}}{2x^{(i)}}x^2 = f_{\text{mul}}^{(i)}(x,z), \quad x \geq 0, y \geq 0. \tag{10}$$

- The logarithm function $R(|x|^2/z) = \ln(1+|x|^2/z)$ with $x \in \mathbb{C}, y \in \mathbb{R}_+$, has a lower bound around a feasible point $(x^{(i)}, z^{(i)})$ [27]:

$$\ln(1+\frac{|x|^2}{z}) \geq R(\frac{|x^{(i)}|^2}{z^{(i)}}) - \frac{|x^{(i)}|^2}{z^{(i)}} + 2\frac{\Re\{x^{(i)} * x\}}{z^{(i)}} - \frac{|x^{(i)}|^2(|x|^2+z)}{(z^{(i)})^2(z^{(i)}+|x^{(i)}|^2)}. \tag{11}$$

- A lower bound of the quadratic function $f_{\text{qu}}(x) = |x|^2$ can be expressed as

$$f_{\text{qu}}(x) \geq 2\Re\{(x^{(i)}) * x\} - |x^{(i)}|^2, x \in \mathbb{C}. \tag{12}$$

Let us convexify the QoS constraint Equation (8c). By using Equation (12), Equation (8c) can be approximated as

$$|\mathbf{h}_k^H \mathbf{w}_k|^2 \geq 2\Re\{(\mathbf{h}_k^H \mathbf{w}_k^{(i)}) * (\mathbf{h}_k^H \mathbf{w}_k)\} - |\mathbf{h}_k^H \mathbf{w}_k^{(i)}|^2, \tag{13}$$

resulting in an alternative convex constraints:

$$2\Re\{(\mathbf{h}_k^H \mathbf{w}_k^{(i)}) * (\mathbf{h}_k^H \mathbf{w}_k)\} - |\mathbf{h}_k^H \mathbf{w}_k^{(i)}|^2 \geq \rho_{\text{sn}}\sigma_k^2, \ \forall k \in \mathcal{K}. \tag{14}$$

To convexify constraint Equation (9c), we can split Equation (9c) into two sub-constraints as

$$\begin{cases} R(\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha})) \geq \lambda_k, & \forall k \in \mathcal{K}, \text{ (15a)} \\ R(\gamma_{\ell,k}(\mathbf{w}, \boldsymbol{\alpha})) \geq \lambda_k, & \forall k \in \mathcal{K}. \text{ (15b)} \end{cases}$$

Using Equation (11), a lower bound of $\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha})$ around a feasible point $(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})$ at iteration $(i+1)$ is derived as

$$R(\gamma_{0,k}(\mathbf{w}, \boldsymbol{\alpha})) \geq \mathcal{A}_{0,k}^{(i)} + \mathcal{A}_{1,k}^{(i)}(\mathbf{w}) - \mathcal{A}_{2,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha}), \tag{16}$$

where $\mathcal{A}_{0,k}^{(i)}$, $\mathcal{A}_{1,k}^{(i)}(\mathbf{w})$, and $\mathcal{A}_{2,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha})$ are respectively given as

$$\begin{aligned} \mathcal{A}_{0,k}^{(i)} &= R(\gamma_{0,k}^{(i)}) - \gamma_{0,k}^{(i)}, \\ \mathcal{A}_{1,k}^{(i)}(\mathbf{w}) &= 2\frac{\Re\{(\mathbf{h}_k^H \mathbf{w}_k^{(i)}) * (\mathbf{h}_k^H \mathbf{w}_k)\}}{\chi_k^{(i)}}, \\ \mathcal{A}_{2,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha}) &= \Theta_k^{(i)}|\mathbf{h}_k^H \mathbf{w}_k|^2 + \Theta_k^{(i)} \sum_{\ell \in \mathcal{K} \setminus \{k\}} (1-\alpha_{k,\ell})|\mathbf{h}_k^H \mathbf{w}_\ell|^2 + \Theta_k^{(i)}\sigma_k^2, \end{aligned}$$

with $\gamma_{0,k}^{(i)} \triangleq \gamma_{0,k}(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})$, $\chi_k^{(i)} \triangleq \chi_k(\mathbf{w}^{(i)}, \boldsymbol{\alpha}^{(i)})$ and $\Theta_k^{(i)} \triangleq (\chi_k^{(i)})^{-1} - (\chi_k^{(i)} + |\mathbf{h}_k^H \mathbf{w}_k^{(i)}|^2)^{-1}$. Clearly, $\mathcal{A}_{2,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha})$ is still non-convex; hence, we apply Equation (10) to convexify $\mathcal{A}_{2,k}^{(i)}(\mathbf{w}, \boldsymbol{\alpha})$ as

$$\begin{aligned}\mathcal{A}_{2,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha}) &\leq \Theta_k^{(i)}\tau_{k,k} + \Theta_k^{(i)}\sum_{\ell\in\mathcal{K}\setminus\{k\}}(1-\alpha_{k,\ell})\tau_{k,\ell} + \Theta_k^{(i)}\sigma_k^2 \\ &\leq \Theta_k^{(i)}\tau_{k,k} + \Theta_k^{(i)}\sum_{\ell\in\mathcal{K}\setminus\{k\}} f_{\text{mul}}^{(i)}(1-\alpha_{k,\ell},\tau_{k,\ell}) + \Theta_k^{(i)}\sigma_k^2 \\ &\triangleq \tilde{\mathcal{A}}_{2,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau}),\end{aligned} \tag{17}$$

where $\boldsymbol{\tau} = \{\tau_{k,\ell}\}_{\forall k,\ell\in\mathcal{K}}$ is introduced as a new variable satisfying the following second-order cone (SOC) constraint:

$$|\mathbf{h}_k^H\mathbf{w}_\ell|^2 \leq \tau_{k,\ell},\ \forall k,\ell\in\mathcal{K}. \tag{18}$$

Therefore, the convex approximation for constraint Equation (15a) at iteration $(i+1)$ can be written as

$$\mathcal{A}_{0,k}^{(i)} + \mathcal{A}_{1,k}^{(i)}(\mathbf{w}) - \tilde{\mathcal{A}}_{2,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau}) \geq \lambda_k,\ \forall k\in\mathcal{K}. \tag{19}$$

It can be observed that constraints Equation (15a,b) take the same form. Hence, constraint Equation (15b) at iteration $(i+1)$ is approximated by the same steps as Equations (16)–(19), i.e., Equation (15b) is convexified as

$$\mathcal{B}_{0,\ell,k}^{(i)} + \mathcal{B}_{1,\ell,k}^{(i)}(\mathbf{w}) - \tilde{\mathcal{B}}_{2,\ell,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau}) \geq \lambda_k,\ \forall \ell,k\in\mathcal{K}, \tag{20}$$

where $\mathcal{B}_{0,\ell,k}^{(i)}$, $\mathcal{B}_{1,\ell,k}^{(i)}(\mathbf{w})$, and $\hat{\mathcal{B}}_{1,\ell,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha})$ are respectively defined as

$$\begin{aligned}\mathcal{B}_{0,\ell,k}^{(i)} &\triangleq R(\gamma_{\ell,k}^{(i)}) - \gamma_{\ell,k}^{(i)}, \\ \mathcal{B}_{1,\ell,k}^{(i)}(\mathbf{w}) &\triangleq 2\frac{\Re\{(\mathbf{h}_\ell^H\mathbf{w}_k^{(i)})*(\ell_k^H\mathbf{w}_k)\}}{\alpha_{\ell,k}^{(i)}\Phi_{\ell,k}^{(i)}}, \\ \tilde{\mathcal{B}}_{2,\ell,k}^{(i)}(\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau}) &\triangleq \Xi_{\ell,k}^{(i)}\tau_{\ell,k} + \Xi_{\ell,k}^{(i)}\sum_{k'\in\mathcal{K}\setminus\{k\}} f_{\text{mul}}^{(i)}(\alpha_{\ell,k},\tau_{\ell,k'}) + \Xi_{\ell,k}^{(i)}\alpha_{\ell,k}\sigma_\ell^2,\end{aligned}$$

where $\gamma_{\ell,k}^{(i)} \triangleq \gamma_{\ell,k}(\mathbf{w}^{(i)},\boldsymbol{\alpha}^{(i)})$, $\Phi_{\ell,k}^{(i)} \triangleq \Phi_{\ell,k}(\mathbf{w}^{(i)})$, and $\Xi_{\ell,k}^{(i)} \triangleq (\Phi_{\ell,k}^{(i)})^{-1} - (\Phi_{\ell,k}^{(i)} + |\mathbf{h}_\ell^H\mathbf{w}_k^{(i)}|^2)^{-1}$.

In summary, a successive convex program at iteration $(i+1)$, which provides minorant maximization for Equation (9), can be formulated as

$$cc \max_{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\lambda},\eta} \quad v\eta - (1-v)P_{\text{T}}(\mathbf{w}), \tag{21a}$$

$$\text{s.t.} \quad \textit{Equation}\ (8\text{b}), \textit{Equation}\ (8\text{d–h}), \textit{Equation}\ (9\text{b}),, \textit{Equation}\ (9\text{d,e}) \textit{Equation}\ (14), \textit{Equations}\ (18)-(20). \tag{21b}$$

Clearly, when the iterative algorithm terminates, the value of the relax variable $\boldsymbol{\alpha}$ is approximately the binary value, and not a solution of Equation (8). To address this point, the binary values of $\boldsymbol{\alpha}$ for the optimal solution of Equation (8) is recovered by the rounding function Equation (22), i.e.,

$$\alpha_{k,\ell}^\star = \lfloor \alpha_{k,\ell}^{(i)} + \frac{1}{2} \rfloor, \quad \forall k,\ell\in\mathcal{K}, \tag{22}$$

where $\lfloor x \rfloor$ returns the maximum integer that is not larger than x, $x\in\mathbb{R}$.

For implementation purposes, it is necessary to find an initial feasible point for Equation (21). This can be achieved by solving the following sub-problem:

$$cc \max_{\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\tau},\boldsymbol{\lambda},\omega} \quad \omega, \tag{23a}$$

$$\text{s.t.} \quad \lambda_k - \bar{R}_k \geq \omega, \quad \forall k\in\mathcal{K}, \tag{23b}$$

$$\omega \leq 0, \quad \forall k\in\mathcal{K}, \tag{23c}$$

$$\textit{Equation}\ (8\text{b}), \textit{Equation}\ (8\text{d–h}) \textit{Equation}\ (9\text{b}), \textit{Equation}\ (14), \textit{Equations}\ (18)-(20), \tag{23d}$$

where ω is a new slack variable. Constraint Equation (23b) guarantees that the initial feasible point can satisfy the QoS requirement. It can be seen that a starting point $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\tau}^{(0)})$ is obtained when ω is approximately zero. The proposed iterative algorithm based on the IA method to solve the problem Equation (8) is summarized in Algorithm 1. In particular, the solution of continuous variables for the relaxation problem is found in Phase 1, while Phase 2 is devoted to re-compute the solution for the original problem with a fixed value of the integer variable $\boldsymbol{\alpha}$.

Convergence analysis: Our proposed algorithm based on the IA method has the convergence properties as in [28]. Let $\mathcal{F}^{(i)}$ be the feasible point set of Equation (21) at iteration i, then the properties of IA method indicate that the optimal solution $\hat{\mathcal{X}}^{(i)} \in \mathcal{F}^{(i)}$ obtained at iteration i satisfies $\hat{\mathcal{X}}^{(i)} \in \mathcal{F}^{(i+1)}$ as well. In addition, $\mathcal{F}^{(i)}$ is a connected set [29]. Therefore, our proposed algorithm can provide at least a local optimal solution that satisfies the Karush-Kuhn-Tucker conditions.

Complexity analysis: The approximated problem Equation (21) includes $5K^2 + 4K + 1$ SOC/linear constraints and $2K^2 + NK + 1$ variables, which complicates the solving of Equation (21) in terms of the big-O expressed as $\mathcal{O}((2K^2 + NK + 1)^{2.5}((5K^2 + 4K + 1)^2 + 2K^2 + NK + 1))$ [30].

Algorithm 1 Proposed Algorithm for Solving Problem Equation (8) based on IA Method.

Phase 1:

1: **Input-1:** $\mathbf{h}_k$, $\bar{R}_k$ $(k \in \mathcal{K})$, $P_{\mathrm{BS}}^{\max}$, and ρ_{sn}.
2: **Initialization:** Set $i := 0$.
3: **Get an initial point:** Execute the sub-problem Equation (23) to obtain an initial point $(\mathbf{w}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\tau}^{(0)})$ for Equation (21).
4: **repeat**
5: Solve the optimization problem Equation (21) to find $(\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star}, \boldsymbol{\tau}^{\star})$ at the i-th iteration.
6: Update $(\mathbf{w}^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\tau}^{(i+1)}) := (\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star}, \boldsymbol{\tau}^{\star})$.
7: Set $i := i + 1$.
8: **until** Convergence
9: Use the rounding function Equation (22) to recover $\boldsymbol{\alpha}^{\star}$.
10: **Output-1:** The optimal solution $\boldsymbol{\alpha}^{\star}$ for the binary variable $\boldsymbol{\alpha}$.

Phase 2:

11: **Input-2:** $\mathbf{h}_k$, $\bar{R}_k$ $(k \in \mathcal{K})$, $P_{\mathrm{BS}}^{\max}$, ρ_{sn}, and $\boldsymbol{\alpha}^{*}$.
12: Reset $i := 0$ and repeat steps 1–7 with binary $\boldsymbol{\alpha}^{\star}$ to obtain the exact beamforming vector $\mathbf{w}^{\star}$.
13: **Output-2:** The optimal solution $(\mathbf{w}^{\star}, \boldsymbol{\alpha}^{\star})$.

5. Proposed Solutions Based on Integer-Variable Reduction

This section presents the reduction in the computational complexity using integer-variable reduction (IVR). Unlike the relaxation method, this approach first determines the value of $\boldsymbol{\alpha}$ for UE pairing; subsequently, power allocation is computed by solving a sub-problem with respect to the given $\boldsymbol{\alpha}$. We investigate two methods to find the UE pairs: (i) a graph-based IVR method to reduce the complexity; and (ii) an IVR using exhaustive search for benchmarking.

5.1. Proposed Solution Using Graph-Based IVR

In this subsection, we propose a low-complexity pairing scheme based on channel vectors. Accordingly, a dynamic pair of UE is determined based on a high channel correlation, to which the SIC technique can be effectively applied. It is observed that when

the UEs with high channel correlation are paired to each other, the BS has more degree of freedom for beamforming design to mitigate the inter-pair interference.

To implement this method, we first define a weight matrix $\mathbf{C} \triangleq [c_{k,\ell}]_{\forall k,\ell \in \mathcal{K}} \in \mathbb{R}^{K \times K}$, in which each entry $c_{k,\ell}$ denoting the channel correlation between the k-th and ℓ-th UEs is calculated as

$$c_{k,\ell} = \frac{| < \mathbf{h}_k, \mathbf{h}_\ell > |}{\|\mathbf{h}_k\| \|\mathbf{h}_\ell\|}, \tag{24}$$

where $< \mathbf{x}, \mathbf{y} >$ denotes the inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$. Subsequently, the UE pairs can be determined by solving the following linear bottleneck assignment problem (LBAP):

$$\max_{\boldsymbol{\alpha}} \quad \min_{\forall k \in \mathcal{K}} \sum_{\ell \in \mathcal{K}} \alpha_{k,\ell} c_{k,\ell}, \tag{25a}$$

$$\text{s.t.} \quad \textit{Equation } (8c), \textit{Equation } (8e), \textit{Equation } (8f). \tag{25b}$$

It is noteworthy that when $c_{k,\ell} = 0$, no edge exists between the k-th and ℓ-th UEs. To solve problem Equation (25), we use the bipartite graph and binary search methods as follows.

5.1.1. Bipartite Graph for UE Pairing

We employ a balanced bipartite graph $\mathcal{G} = (\mathcal{K}_1, \mathcal{K}_2; \mathcal{D})$, where $\mathcal{K}_1$ and $\mathcal{K}_2$ are the vertex sets of all UEs with $|\mathcal{K}_1| = |\mathcal{K}_2| = K$, and $\mathcal{D}$ is the set of edges, i.e., $\mathcal{D} = \{\{k, \ell\} | k, \ell \in \mathcal{K} \text{ and } k \neq \ell\}$, with the weight of an edge $\{k, \ell\}$ being $c_{k,\ell}$. UE pairing can be performed by using the augmenting path algorithm [31]. Some definitions for developing the pairing algorithm are as follow:

- A matching $\mathcal{M}$ is a subset of $\mathcal{D}$ without any two edges sharing the same vertex. For convenience, we define an unmatched set $\bar{\mathcal{M}} = \mathcal{D} \backslash \mathcal{M}$. Furthermore, a maximum matching corresponds to a special case, in which $\mathcal{M}$ is enlarged to the maximum.
- A free vertex is defined as a vertex not contained in $\mathcal{M}$.
- An alternating path $\mathcal{P}$ is a path in graph $\mathcal{G}$, such that two arbitrary successive edges in $\mathcal{P}$ are from both sets $\mathcal{M}$ and $\bar{\mathcal{M}}$.
- An alternating path $\mathcal{P}$ becomes an augmenting path when both the starting and ending points of $\mathcal{P}$ are free vertexes.

We can use the augmenting path algorithm to expand $\mathcal{M}$ until a maximum matching is obtained. In particular, at each iteration of an iterative algorithm, the augmenting path $\mathcal{P}$ is determined for a given matching $\mathcal{M}$ ($\mathcal{M} = \emptyset$ at the beginning), and then, $\mathcal{M}$ is updated as $\mathcal{M} = (\mathcal{M} \backslash \mathcal{P}) \cup (P \backslash M)$. The algorithm terminates when no augmenting path is found. However, a maximum matching with the augmenting path algorithm merely addresses the maximum pairs of UEs, which is insufficient for obtaining the optimal solution for problem Equation (25) under a balanced bipartite graph in a polynomial-time complexity. Hence, the augmenting path algorithm is integrated into a binary search method as will be described in the following subsection.

5.1.2. Binary Search with Augmenting Path Algorithm

The optimal solution for problem Equation (25) is a maximum matching with channel correlation fairness among UE pairs. In other words, the pairs of UEs are determined such that the minimum channel correlation of UE pairs in the achieved maximum matching is maximized. Moreover, the maximal value for the weight fairness is one of the entries in $\mathbf{C}$. Therefore, the binary search algorithm [32] is used to find the solution for Equation (25). The search region is given by an auxiliary vector $\boldsymbol{v} \triangleq [v_i]_{i \in \{1, \ldots, K^2\}} \in \mathbb{R}^{1 \times K^2}$, of which the components involving all entries of matrix $\mathbf{C}$ are sorted in the ascending order, i.e., $v_1 \leq v_2 \leq \cdots \leq v_{K^2}$. Let $\mathrm{id}_{\min}$ and $\mathrm{id}_{\max}$ be the lower and upper indexes for the search region on vector $\mathbf{v}$, respectively. Clearly, $\mathrm{id}_{\min} = 1$ and $\mathrm{id}_{\max} = K^2$ at the initialization.

The search region is denoted by $\mathcal{S}(\mathrm{id}_{\min}, \mathrm{id}_{\max}) \triangleq [v_{\mathrm{id}_{\min}}, \cdots, v_{\mathrm{id}_{\max}}]$. At each iteration, the index of the median in the search region is determined as

$$\mathrm{id}_{\mathrm{med}} = \left\lfloor \frac{\mathrm{id}_{\min} + \mathrm{id}_{\max}}{2} \right\rfloor. \tag{26}$$

Subsequently, the augmenting path algorithm is executed on sub-graph $\tilde{\mathcal{G}} = (\mathcal{K}_1, \mathcal{K}_2; \tilde{\mathcal{D}}) \subseteq \mathcal{G}$ to find the maximum matching $\mathcal{M}$, where $\tilde{\mathcal{D}} \triangleq \mathcal{D} \backslash \{\{k, \ell\} | c_{k,\ell} < v_{\mathrm{id}_{\mathrm{med}}}\}$. The search region is updated as

$$\mathcal{S}(\mathrm{id}_{\min}, \mathrm{id}_{\max}) = \begin{cases} \mathcal{S}(\mathrm{id}_{\mathrm{med}}, \mathrm{id}_{\max}), & \text{if } |\mathcal{M}| = K \\ \mathcal{S}(\mathrm{id}_{\min}, \mathrm{id}_{\mathrm{med}}), & \text{if } |\mathcal{M}| < K. \end{cases} \tag{27}$$

The binary search algorithm terminates when $\mathcal{S}(\mathrm{id}_{\min}, \mathrm{id}_{\max})$ contains only one element, i.e., $\mathrm{id}_{\mathrm{med}} = \mathrm{id}_{\min} = \mathrm{id}_{\max}$, which results in the value of $\boldsymbol{\alpha}$. After obtaining the value of $\boldsymbol{\alpha}$, we solve the optimization problem Equation (21) with the given $\boldsymbol{\alpha}$ to obtain the solution for Equation (8). The proposed algorithm using graph-based IVR is summarized in Algorithm 2. By applying binary search and augmenting path methods for the bipartite graph, a dynamic pairing solution is computed in Phase 1, at which the binary value of variable $\boldsymbol{\alpha}$ is determined to satisfy the constraints of the original problem. In Phase 2, a convex approximation program is solved under the binary value of $\boldsymbol{\alpha}$ obtained in Phase 1.

Complexity Analysis: The binary search requires $\mathcal{O}(\log_2(K^2))$ steps to find the optimal value c_{med}, while the augmenting path algorithm takes $\mathcal{O}(K^3)$ [33]. Therefore, the total complexity for finding $\boldsymbol{\alpha}$ in problem Equation (25) is $\mathcal{O}(K^3 \log_2(K^2))$. With a given value of $\boldsymbol{\alpha}$, a sub-problem derived from problem Equation (21) has $K^2 + NK + 1$ variables and $K^2 + 3K + \lfloor K/2 \rfloor + 1$ SOC/linear constraints; consequently, the total complexity of Algorithm 2 is $\mathcal{O}(K^3 \log_2 K^2 + (K^2 + NK + 1)^{2.5}((K^2 + 3K + \lfloor K/2 \rfloor + 1)^2 + K^2 + NK + 1))$.

Algorithm 2 Proposed Algorithm for Solving Problem Equation (8) using Graph-based IVR.

Phase 1:

1: **Input-1:** $\mathbf{h}_k$, $(k \in \mathcal{K})$.
2: Compute the weight matrix $\boldsymbol{C}$ as in Equation (24).
3: Generate an auxiliary vector $\boldsymbol{v}$ which is sorted in an ascending order.
4: Set $\mathrm{id}_{\min} := 1$, $\mathrm{id}_{\max} := K^2$.
5: **repeat**
6: Compute $\mathrm{id}_{\mathrm{med}}$ as in Equation (26).
7: Build a sub-graph $\tilde{\mathcal{G}}$ using the median $v_{\mathrm{id}_{\mathrm{med}}}$.
8: Apply augmenting path algorithm to find the maximum matching $\mathcal{M}$ in $\tilde{\mathcal{G}}$.
9: **if** $|\mathcal{M}| = K$ **then**
10: Set $\mathrm{id}_{\min} := \mathrm{id}_{\mathrm{med}}$.
11: **else**
12: Set $\mathrm{id}_{\max} := \mathrm{id}_{\mathrm{med}}$.
13: **end if**
14: **until** $\mathrm{id}_{\min} = \mathrm{id}_{\max}$
15: Create $\boldsymbol{\alpha}$ corresponding to the given $\mathcal{M}$.
16: **Output-1:** The pairing solution $\boldsymbol{\alpha}^\star$ for the binary variable $\boldsymbol{\alpha}$.

Phase 2:

17: **Input-2:** $\mathbf{h}_k$, $\bar{R}_k$ $(k \in \mathcal{K})$, $P_{\mathrm{BS}}^{\max}$, ρ_{sn}, and $\boldsymbol{\alpha}^*$.
18: Apply Steps 1–7 in Algorithm 1 with the given value of $\boldsymbol{\alpha}^\star$ to find the beamforming vector $\mathbf{w}^\star$.
19: **Output-2:** The optimal solution $(\mathbf{w}^\star, \boldsymbol{\alpha}^\star)$.

5.2. Proposed Solution Based on Exhaustive Search

For benchmarking, this section presents the optimal solution using exhaustive search (ES), where all possible cases of $\boldsymbol{\alpha}$ are considered while satisfying constraints Equation (8c–h). For each case, a sub-problem is derived from Equation (8) by fixing the value of $\boldsymbol{\alpha}$. Instead of only one subproblem, we obtain many sub-problems as in the graph-based method. Subsequently, we use Equation (21) to find the solutions to the sub-problems. The solution $(\mathbf{w}, \boldsymbol{\alpha})$ that provides the best performance is selected as the optimal solution. It can be observed that each sub-problem can be solved with a fast convergence since it only needs to handle the continuous variables of the beamforming vectors. However, this approach typical incurs high computational complexity owning to the large number of sub-problems.

Complexity Analysis: The ES method examines $\sum_{k=1}^{\lfloor K/2 \rfloor} \binom{K}{k} \frac{(2k)!}{2^k k!}$ sub-problems as all possible cases of $\boldsymbol{\alpha}$. For a given value of $\boldsymbol{\alpha}$, a sub-problem based on Equation (21) contains the same numbers of variables and constraints as in Algorithm 2. Therefore, the total complexity of the ES method is determined as $\mathcal{O}\Big(\Big(\sum_{k=1}^{\lfloor K/2 \rfloor} \binom{K}{k} \frac{(2k)!}{2^k k!}\Big)(K^2+NK+1)^{2.5}((K^2+3K+\lfloor K/2 \rfloor+1)^2+K^2+NK+1)\Big)$.

6. Numerical Results

6.1. Parameter and Simulation Settings

We consider a DL cellular system where one BS equipped with N-antennas serves K single-antenna UEs. The BS is located at the center of the cell, whereas the UEs are uniformly distributed within radius R from the cell center. To evaluate the performance of the proposed algorithms, we adopt Monte-Carlo simulations. The channel from the BS to each UE includes the path loss, $\mathrm{PL}(d_{\mathrm{BS,UE}})$ with distance $d_{\mathrm{BS,UE}}$, and small-scale fading expressed as a Gaussian random variable $\mathbf{h} \sim \mathcal{CN}(0, \mathrm{PL}(d_{\mathrm{BS,UE}})\mathbf{I})$. The other simulation parameters are shown in Table 1,in which the realistic parameters of power consumption and PA efficiency are provided by [24]. We compare Algorithm 1 (Algorithm 1), Algorithm 2 (Algorithm 2) and the ES scheme with the existing schemes. Before UE pairing, we first sort all UEs in descending order of their channel gains. The pairing strategies are described as follows:

- Greedy pairing with difference fairness for channel gains (GP-DFCG): A pair comprises the k-th and $(K-\lfloor K/2 \rfloor + k)$-th UEs, resulting in $\alpha_{k,K-\lfloor K/2 \rfloor+k} = 1, 1 \le k \le \lfloor K/2 \rfloor$.
- Greedy pairing with the strongest-to-weakest channel-gain strategy (GP-SWCG): Considering all sorted UEs in an unpaired set, a pair is selected by popping the strongest and weakest channel-gain UEs until the unpaired set is empty or contains one UE, i.e., a pair including the k-th and $(K-k+1)$-th UEs, resulting in $\alpha_{k,K-k+1} = 1,\ 1 \le k \le \lfloor K/2 \rfloor$.
- Consecutive pairing (CP): As in [19], the $(2k-1)$-th and $2k$-th UEs are paired together, resulting in $\alpha_{2k-1,2k} = 1,\ 1 \le k \le \lfloor K/2 \rfloor$.

Furthermore, we also investigate two other typical schemes: (i) Random pairing, where two UEs in the cell are randomly paired with each other; (ii) Beamforming design, where conventional beamforming is applied to all UEs, resulting in an all-zero matrix $\boldsymbol{\alpha}$. Except for Algorithm 1, the sub-problems, which are derived as the cases of problem Equation (21) for achieved values of $\boldsymbol{\alpha}$, are solved for the beamforming vectors.

Table 1. Simulation Parameters.

Parameter	Value
Noise power spectral density	-174 dBm/Hz
Radius of the cell, R	100 m
Path loss from BS to UEs, $\mathrm{PL}_{\mathrm{BS,UE}}$	$128.1 + 37.6\log_{10}(d_{\mathrm{BS,UE}})$ dB
The nearest distance from between UEs and BS	10 m
Power budget at BS, $P_{\mathrm{BS}}^{\max}$	18 dBm
Rate threshold, $\bar{R}_k = \bar{R}_\ell = \bar{R}, \forall k, \ell \in \mathcal{K}$	1 bits/s/Hz
Number of UEs, K	6
Number of antennas at BS, N	4
SNR threshold at receivers, ρ_{sn}	0 dB
PA efficiency at BS, ζ_{PA}	0.3

6.2. Rate Fairness Performance

To evaluate the effectiveness of the proposed algorithms in terms of rate fairness, the cumulative distribution functions (CDFs) of all schemes are investigated with respect to the MMR. We use 1000 random channels to estimate the CDFs in Figure 2. In general, the GP-based and CP-based schemes exhibit better performance than random pairing and the conventional beamforming design, i.e., gains of 0.4–0.8 bps/Hz at the mid-point. As expected, the proposed algorithms outperform GP- and CP-based schemes. In particular, when comparing to the GP-DFCG scheme at the mid-point, the performance gains of Algorithms 1 and 2 are 0.7 and 0.3 bps/Hz, respectively. Except for beamforming, the other schemes provide high rate fairness for all UEs, exceeding 3 bps/Hz at the 95-percentile points. These CDFs spread in a wide range of MMR (approximately 2 bps/Hz from the 5- to 95-percentile points), adapting to the channel responses. This figure verifies the importance of the channel-gain disparities as well as the strategies of UE pairing for NOMA-based systems.

Figure 3 depicts the average MMR of the eight schemes along with the change in the budget at the BS. For comparison, the conventional beamforming and ES schemes are also investigated as two baselines. In comparison with the beamforming scheme, the random pairing scheme provides the least MMR gain, whereas the two proposed algorithms outperform the other existing schemes. The performance of Algorithm 1 is approximately 0.5 bps/Hz better than that of Algorithm 2, compensating for the computational complexity of jointly optimizing the mixed-integer variables. Interestingly, the performance of Algorithm 1 is quite close to that of the ES schemes (approximately 0.3 bps/Hz loss on average), thereby confirming the effectiveness of the proposed scheme.

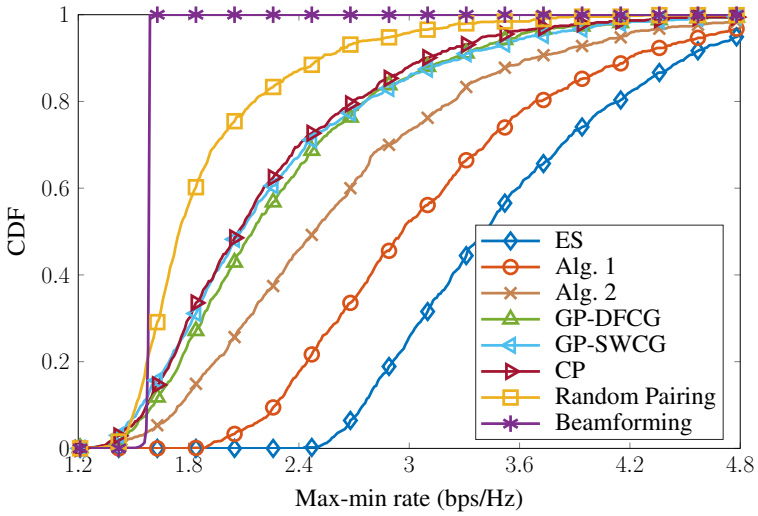

Figure 2. Cumulative distribution function of max-min rate.

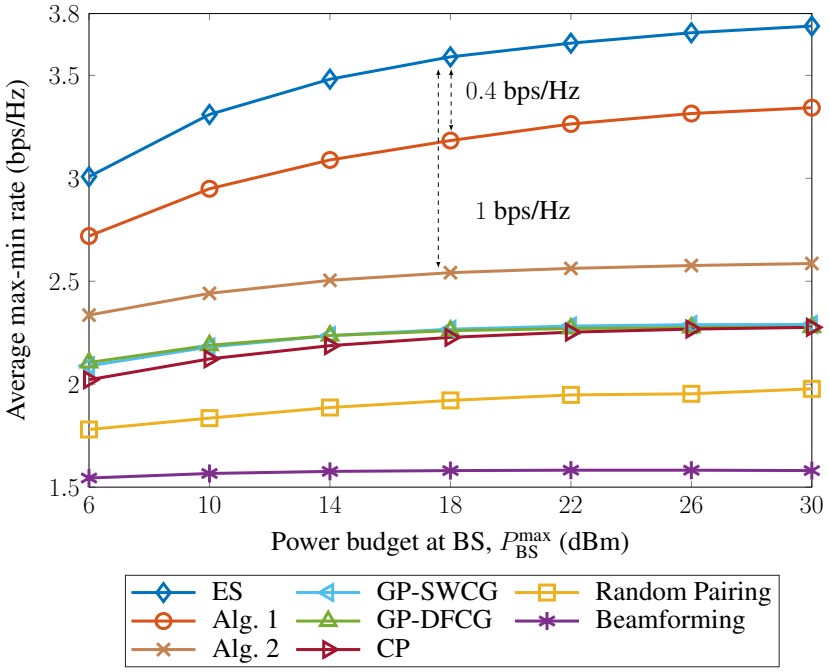

Figure 3. Average max-min rate versus power budget at BS.

6.3. Power Consumption Minimization

We first examine the minimum rate of all UEs and power-consumption percentage at the BS with the various values of power budget at the BS, $P_{\mathrm{BS}}^{\max}$, as illustrated in Figure 4. The minimum rates obtained using Algorithms 1 and 2 are only 1 bps/Hz when considering power consumption minimization (i.e., $v = 0$), which is exactly the minimum bit-rate threshold associated with the QoS requirements. This allows the BS to consume sufficient power and hence reduce the major part of the power budget. For example, the power-consumption percentage is lower than 1% of $P_{\mathrm{BS}}^{\max}$ and inversely proportional to $P_{\mathrm{BS}}^{\max}$. Meanwhile, when the design problem is targeted at the MMR $(v = 1)$, the minimum rate and power-consumption percentage obtained by Algorithms 1 and 2 increase along with $P_{\mathrm{BS}}^{\max}$. The minimum rates are equal to MMRs of Algorithms 1 and 2, consistent with the results in Figure 3. As shown, the power-consumption percentages for the MMR are much higher than those for power consumption minimization; however, the entire power budget is not used. Typically, at $P_{\mathrm{BS}}^{\max} = 18$ dBm, the power consumption percentages of Algorithms 1 and 2 are 56.7% and 97.8%, respectively. This supports interference management for achieving rate fairness among all UEs, resulting in a decrease in the power-consumed percentages and MMR saturation at high values of $P_{\mathrm{BS}}^{\max}$.

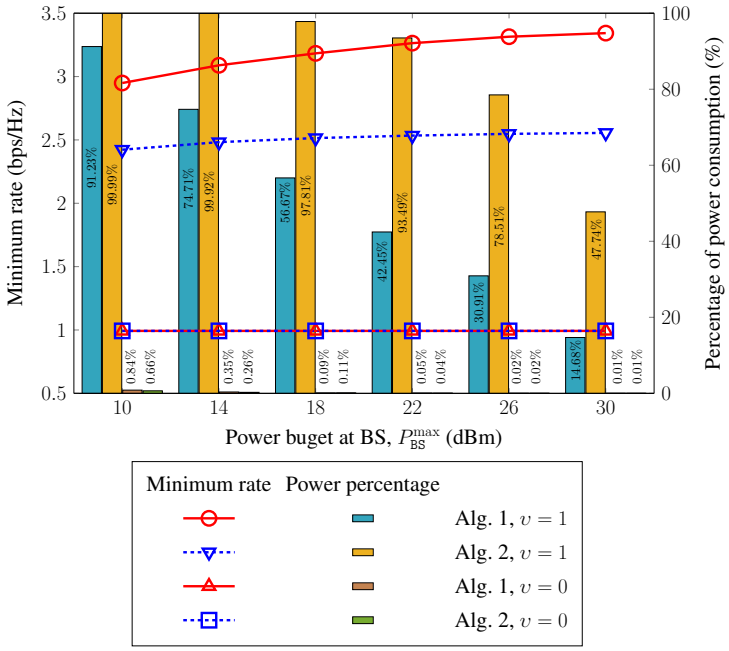

Figure 4. Minimum rate and percentage of power consumption with respect to power budget at BS.

To understand the behavior of the proposed algorithms, Figure 5 depicts the power consumption as a function of the QoS requirement. It is clear that the power consumption of all schemes increases with the minimum bit rate. As expected, the proposed algorithms outperform the others, and satisfy the high level of the minimum bit rate. In particular, from $\bar{R} = 2$ (bps/Hz), the proposed algorithms still yield good solution for satisfying the QoS requirement constraint, whereas the other existing schemes could not easily yield the feasible point. Figure 6 demonstrates the power consumption performance versus SNR thresholds. For convenience, we only examine the GP-SWCG scheme, which provides the best performance among the three greedy-based schemes. It is observed that the proposed algorithms provide performance comparable to that of the ES method even at a low SNR threshold. Compared with random pairing, the proposed schemes use approximately 4 dBm, while GP-SWCG provides more than 1 dBm of power saving. These results confirm the flexibility of the proposed methods for the UE pairing to enhance the rate fairness and reduce power consumption.

To further realize the effect of the SNR threshold on system performance, we investigate MMR and power-consumption as a function of SNR threshold, as shown in Figure 7. For power consumption minimization ($v = 0$), the percentage of consumed power is small and proportional to the SNR threshold, whereas the minimum rate remains almost constant. In this case, both proposed algorithms maintain sufficient consumed power to satisfy the QoS requirement, i.e., the minimum rate of 1 bps/Hz. Specifically, for a low SNR threshold, the power consumption percentage of both Algorithms 1 and 2 are less than 1%, whereas the percentage of consumed power for the higher SNR threshold increases to satisfy the requirement for decoding messages, i.e., approximately 15% at $\rho_{\mathrm{sn}} = 25$ dB. In addition, for the MMR problem ($v = 1$), the percentage of power consumption differs significantly from that of the power-minimization case. In particular, more than 90% of power budget is required for both algorithms at $\rho_{\mathrm{sn}} = 25$ dB. Moreover, when ρ_{sn} increases, the MMR performance degrades slightly. This is because the power allocated to each UE is increased to satisfy the SNR threshold at all UEs, resulting in more interference among the UEs and a decrease in the minimum rate. However, between $\rho_{\mathrm{sn}} = 0$ and $\rho_{\mathrm{sn}} = 25$ dB, the minimum rates obtained by Algorithms 1 and 2 reduces by only approximately 0.2 bps/Hz, indicating the adaption of the proposed schemes even under a strict received-SNR condition for NOMA-based beamforming design.

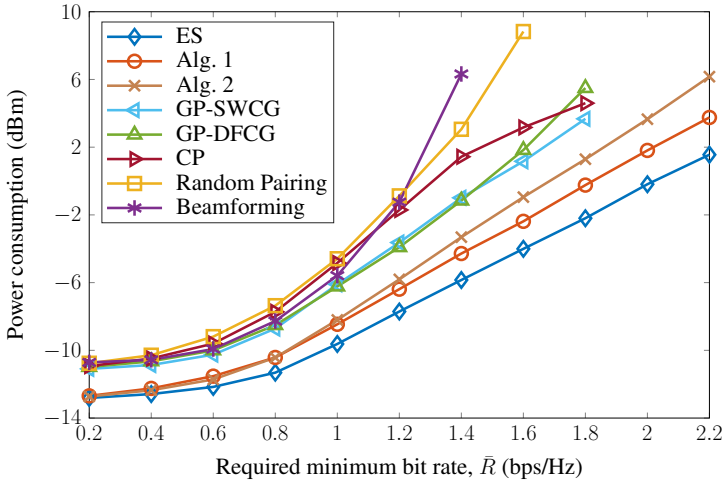

Figure 5. Power consumption versus minimum bit rate.

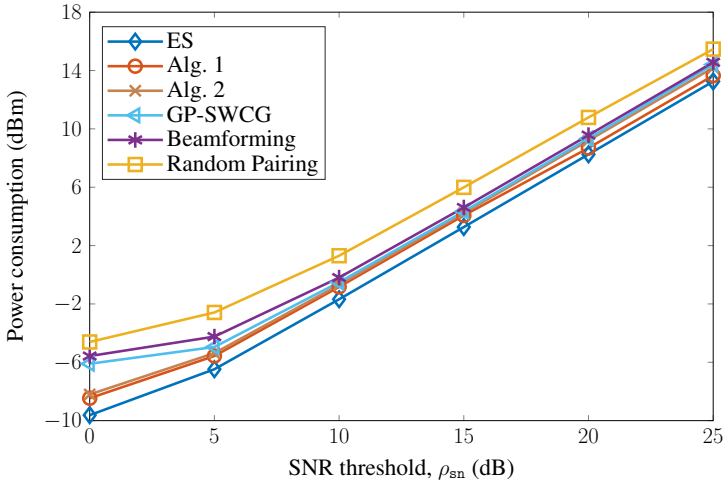

Figure 6. Power consumption versus received SNR threshold.

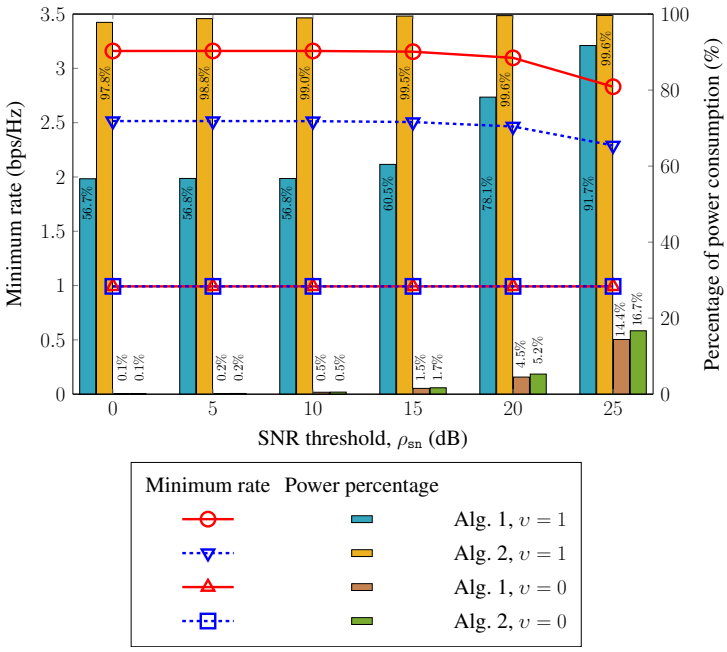

Figure 7. Minimum rate and percentage of power consumption versus SNR threshold.

6.4. Convergence Behaviors

We investigate the convergence behaviors of the proposed algorithms by referencing to the beamforming design. Figure 8 shows the MMR performance of the three schemes at $P_{\mathrm{BS}}^{\max} = 18$ dBm and $\rho_{\mathrm{sn}} = 0$ dB. Because the ES method contains many sub-problems, it is difficult to show its convergence behavior; hence, the ES method is omitted in this figure. As previously mentioned, Algorithm 1 includes two phases where the relaxed variable $\boldsymbol{\alpha}$ is rounded after the first phase, and the optimal solution for the original problem is obtained in the second phase. Hence, the convergence line of Algorithm 1 in Figure 8 consists of a bumpy period before reaching the saturation value. Furthermore, the two proposed algorithms achieve 99% performance with fast convergence. In particular, Algorithm 1 converges almost from iteration 13, whereas Algorithm 2 and beamforming require only a few iterations for a faster convergence rate (i.e., convergence at the 10-th iteration). However, it can be concluded that both proposed algorithms provide fast convergence.

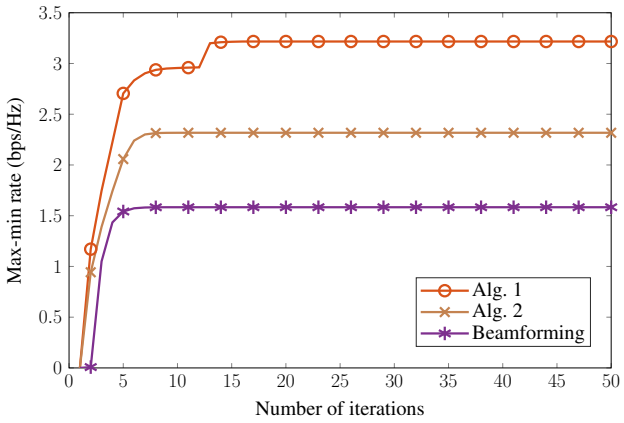

Figure 8. Convergence behaviors of MMR.

7. Conclusions

We studied a hybrid of NOMA and conventional beamforming design, where a dynamic pairing technique was investigated with two arbitrary UEs in a cell that can be paired together. The aim of this paper was to improve the performance in terms of both fairness rate among all UEs and power consumption. To tackle the resulting mixed-integer non-convex problem, we proposed two low-complexity algorithms. The first algorithm was based on a combination of the relaxation and IA methods to jointly optimize UE pairing and beamforming. In the second algorithm, we used the bipartite graph and binary search methods to achieve UE pairing before using the IA method to obtain the beamforming vectors, which can further reduce the computational complexity. Numerical results with realistic parameters demonstrated that the two proposed algorithms outperform the existing schemes in terms of both MMR and the power consumption.

Besides the advantage of novel dynamic UE pairing, some research gaps would be investigated for future works. Even when two proposed algorithms provide good performance and low complexity, an advanced optimization technique for NOMA could be expected to reduce the rate loss. In addition, the channel estimation is not perfect in practice; thus, a robust design under the channel uncertainty would be essential for UE pairing and beamforming in NOMA-based systems. Finally, the effectiveness of dynamic UE pairing in massive device networks is still unknown, which motivates us to develop a new method for the upcoming requirement and environment.

Author Contributions: Conceptualization, K.-H.N., H.V.N., and O.-S.S.; methodology, K.-H.N., H.V.N., and O.-S.S.; validation, K.-H.N., H.V.N., and O.-S.S.; investigation, K.-H.N., H.V.N., M.T.P.L., T.X.C., and O.-S.S.; writing—original draft preparation, K.-H.N. and H.V.N.; writing—review and editing, K.-H.N., H.V.N., M.T.P.L., T.X.C., and O.-S.S.; visualization, O.-S.S.; supervision, O.-S.S.; project administration, O.-S.S.; funding acquisition, O.-S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-00724, Development of Beyond 5G Mobile Communication Technologies (Ultra-Reliable, Low-Latency, and Massive Connectivity) and Combined Access Technologies for Cellular-based Industrial Automation Systems), in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1A2C1084834), and in part by Ministry of Education and Training under project number B2020-DNA-06.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dai, L.; Wang, B.; Yuan, Y.; Han, S.; Chih-Lin, I.; Wang, Z. Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends. *IEEE Commun. Mag.* **2015**, *53*, 74–81. [CrossRef]
2. Ding, Z.; Liu, Y.; Choi, J.; Sun, Q.; Elkashlan, M.I.C.; Poor, H.V. Application of non-orthogonal multiple access in LTE and 5G networks. *IEEE Commun. Mag.* **2017**, *55*, 185–191. [CrossRef]
3. Liu, Y.; Qin, Z.; Elkashlan, M.; Ding, Z.; Nallanathan, A.; Hanzo, L. Nonorthogonal multiple access for 5G and beyond. *Proc. IEEE* **2017**, *105*, 2347–2381. [CrossRef]
4. Saito, Y.; Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Li, A.; Higuchi, K. Non-orthogonal multiple access (NOMA) for cellular future radio access. In Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, Germany, 2–5 June 2013; pp. 1–5.
5. Islam, S.M.R.; Avazov, N.; Dobre, O.A.; Kwak, K. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 721–742. [CrossRef]
6. Le, M.T.P.; Ferrante, G.C.; Quek, T.Q.S.; Benedetto, M.D. Fundamental limits of low-density spreading NOMA with fading. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4648–4659. [CrossRef]
7. Saito, Y.; Benjebbour, A.; Kishiyama, Y.; Nakamura, T. System-level performance evaluation of downlink non-orthogonal multiple access (NOMA). In Proceedings of the 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), London, UK, 8–11 September 2013; pp. 611–615.
8. Ali, M.S.; Tabassum, H.; Hossain, E. Dynamic user clustering and power allocation for uplink and downlink nonorthogonal multiple access (NOMA) systems. *IEEE Access* **2016**, *4*, 6325–6343.
9. Nguyen, H.V.; Nguyen, V.-D.; Dobre, O.A.; Nguyen, D.N.; Dutkiewicz, E.; Shin, O.-S. Joint power control and user association for NOMA-based full-duplex systems. *IEEE Trans. Commun.* **2019**, *67*, 8037–8055. [CrossRef]
10. Fang, F.; Zhang, H.; Cheng, J.; Roy, S.; Leung, V.C.M. Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI. *IEEE J. Select. Areas Commun.* **2017**, *35*, 2874–2885. [CrossRef]
11. Ali, S.; Hossain, E.; Kim, D.I. Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation. *IEEE Access* **2017**, *5*, 565–577. [CrossRef]
12. Liang, W.; Ding, Z.; Li, Y.; Song, L. User pairing for downlink non-orthogonal multiple access networks using matching algorithm. *IEEE Trans. Commun.* **2017**, *65*, 5319–5332. [CrossRef]
13. Bui, V.-P.; Nguyen, P.X.; Nguyen, H.V.; Nguyen, V.-D.; Shin, O.-S. Optimal user pairing for achieving rate fairness in downlink NOMA networks. In Proceedings of the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 11–13 February 2019; pp. 575–578.
14. Mounchili, S.; Hamouda, S. Pairing distance resolution and power control for massive connectivity improvement in NOMA systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4093–4103. [CrossRef]
15. Chen, X.; Gong, F.; Li, G.; Zhang, H.; Song, P. User pairing and pair scheduling in massive MIMO-NOMA systems. *IEEE Commun. Lett.* **2018**, *22*, 788–791. [CrossRef]
16. Ding, Z.; Fan, P.; Poor, H.V. Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. *IEEE Trans. Veh. Technol.* **2016**, *65*, 6010–6023. [CrossRef]
17. Zhang, H.; Zhang, D.; Meng, W.; Li, C. User pairing algorithm with SIC in non-orthogonal multiple access system. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6.
18. Zhu, L.; Zhang, J.; Xiao, Z.; Cao, X.; Wu, D.D. Optimal user pairing for downlink non-orthogonal multiple access (NOMA). *IEEE Wirel. Commun. Lett.* **2019**, *8*, 328–331. [CrossRef]
19. Cheng, Y.; Li, K.H.; Teh, K.C.; Luo, S. Joint user pairing and subchannel allocation for multisubchannel multiuser nonorthogonal multiple access systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8238–8248. [CrossRef]
20. Nguyen, K.-H.; Nguyen, H.V.; Bui, V.-P.; Shin, O.-S. Dynamic user pairing for non-orthogonal multiple access in downlink networks. *arXiv* **2020**, arXiv:2006.14297.

21. Auer, G.; Blume, O.; Giannini, V.; Godor, I.; Imran, M.A.; Jading, Y.; Katranaras, E.; Olsson, M.; Sabella, D.; Skillermark, P.; et al. *Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown*; Technical Report INFSO-ICT-247733; Deliverable D2.3: Valencia, Spain, 2012; pp. 1–68.
22. Zappone, A.; Lin, P.-H.; Jorswieck, E.A. Energy efficiency in secure multi-antenna systems. *arXiv* **2015**, arXiv:1505.02385.
23. Zhang, Y.; Wang, H.; Zheng, T.; Yang, Q. Energy-efficient transmission design in non-orthogonal multiple access. *IEEE Trans. Veh. Technol.* **2017**, *66*, 2852–2857. [CrossRef]
24. Bjornson, E.; Sanguinetti, L.; Hoydis, J.; Debbah, M. Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer? *IEEE Trans. Wirel. Commun.* **2015**, *14*, 3059–3075. [CrossRef]
25. Tombaz, S.; Vastberg, A.; Zander, J. Energy- and cost-efficient ultra-high-capacity wireless access. *IEEE Wirel. Commun.* **2011**, *18*, 18–24. [CrossRef]
26. Nguyen, V.-D.; Nguyen, H.V.; Dobre, O.A.; Shin, O.-S. A new design paradigm for secure full-duplex multiuser systems. *IEEE J. Select. Areas Commun.* **2018**, *36*, 1480–1498. [CrossRef]
27. Nguyen, V.-D.; Duong, T.Q.; Tuan, H.D.; Shin, O.-S.; Poor, H.V. Spectral and energy efficiencies in full-duplex wireless information and power transfer. *IEEE Trans. Commun.* **2017**, *65*, 2220–2233. [CrossRef]
28. Marks, B.R.; Wight, G.P. A general inner approximation algorithm for nonconvex mathematical programs. *Oper. Res.* **1978**, *26*, 681–683. [CrossRef]
29. Nguyen, H.V.; Nguyen, V.-D.; Dobre, O.A.; Wu, Y.; Shin, O.-S. Joint antenna array mode selection and user assignment for full-duplex MU-MISO systems. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 2946–2963. [CrossRef]
30. Labit, Y.; Peaucelle, D.; Henrion, D. SEDUMI INTERFACE 1.02: A tool for solving LMI problems with SEDUMI. In Proceedings of the IEEE International Symposium on Computer Aided Control System Design, Glasgow, UK, 20 September 2002; pp. 272–277.
31. Burkard, R.; Dell'Amico, M.; Martello, S. *Assignment Problems*; Society for Industrial and Applied Mathematics, 2012. Available online: https://epubs.siam.org/doi/abs/10.1137/1.9781611972238 (accessed on 21 December 2020).
32. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2001.
33. Nguyen, H.V.; Duong, Q.; Nguyen, V.-D.; Shin, Y.; Shin, O.-S. Optimization of resource allocation for underlay device-to-device communications in cellular networks. *Peer-to-Peer Netw. Appl.* **2016**, *9*, 965–977. [CrossRef]