IEEE ComSoc IEEE Open Journal of the Communications Society

Received 9 May 2024; revised 17 June 2024; accepted 26 June 2024. Date of publication 1 July 2024; date of current version 25 July 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3421382

Optimal User Pairing Strategy for Minimum Power Utilization in Downlink Non-Orthogonal Multiple Access Systems

HASSAN NOOH [1], SEUNGHWAN WON [1] (Senior Member, IEEE), SOON XIN NG [2] (Senior Member, IEEE), MUHAMMAD FARHAN SOHAIL [3], MINKWAN KIM [4], AND MOHAMMED EL-HAJJAR [2] (Senior Member, IEEE)

[1]Connected Intelligence Research Group, University of Southampton Malaysia, Iskandar Puteri 79100, Malaysia

[2]Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K.

[3]Department of Electrical Engineering, National University of Modern Languages, Islamabad 44000, Pakistan

[4]Department of Aeronautical and Astronautical Engineering, Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ Southampton, U.K.

CORRESPONDING AUTHOR: M. EL-HAJJAR (e-mail: meh@ecs.soton.ac.uk)

This work was supported by the Engineering and Physical Sciences Research Council under Project EP/Y037243/1 and Project EP/X04047X/2.

ABSTRACT In this paper, we investigate the impact of user pairing on the power consumption of 2-user non-orthogonal multiple access (NOMA) systems in the downlink. We formulate the joint power allocation and user pairing problem as a mixed-integer programming problem with the objective of minimizing the total transmit power consumption. While the pairwise power allocation strategy is straightforward, for a system with $2K$ users and K NOMA pairs, there exist $\frac{(2K)!}{2^K \times K!}$ possible pairing strategies, resulting in a combinatorial search space that grows drastically with the number of users in the system. Hence, we propose an analytical approach to obtain the globally optimum user pairing strategy. Notably, our procedure has a linear time complexity of $\mathcal{O}(2K)$, which is a significant improvement over the suboptimal and computationally expensive methods in the existing literature. We demonstrate through extensive simulations that the proposed optimal pairing strategy can attain considerable performance gains in terms of power savings compared to benchmark schemes. In particular, in a typical deployment environment, 63% of the total power budget is saved at a mean received signal-to-noise ratio (SNR) of 15.7 dB among the users. Finally, we evaluate the energy efficiency (EE) of NOMA transmission compared to the EE achieved through orthogonal multiple access (OMA) transmission. We demonstrate that the EE gain of NOMA transmission compared to OMA is improved more than sixfold at convergence by adopting the power minimization approach studied in this work, rather than adopting the sum rate maximization approach found in the literature.

INDEX TERMS Non-orthogonal multiple access (NOMA), user pairing, power allocation, power minimization, energy efficiency.

I. INTRODUCTION

NON-ORTHOGONAL multiple access (NOMA) has been regarded as an enabling technology for beyond fifth-generation mobile systems [1], [2]. This is particularly true within the context of emerging new applications for next-generation communications such as non-terrestrial networks (NTNs) and Internet of Things (IoT) applications, where NOMA's salient capabilities in interference management and spectral efficiency enhancement are especially valuable. More specifically, NOMA offers superior spectral efficiency compared to orthogonal multiple access (OMA) schemes, such as orthogonal frequency division multiple access (OFDMA). This is largely due to NOMA's capability to allow multiple users to

share the same resource blocks (RBs) which may be in time, frequency, or code. The standard implementation of NOMA relies on superposition coding at the transmitter and successive interference cancellation (SIC) as a decoding strategy at the receiver [1]. 2-user NOMA is the typical deployment scenario, where the number of users sharing an RB is limited to a pair of users. The delay incurred and SIC error propagation are exacerbated when more users are multiplexed together [3].

The user pairing strategy is a crucial aspect of the NOMA system since it can significantly affect system performance. Equally crucial is the consideration of the implementation complexity incurred by the user pairing process. Therefore, there is a pressing need for low-complexity procedures for user pairing that guarantee optimal performance [4], [5]. However, deriving the optimal user pairing scheme along with a suitable power allocation method under any objective function usually results in a non-convex mixed-integer programming problem where the full search space grows combinatorially as the number of users increases. Accordingly, from a system design perspective, exhaustive search or computational methods with high complexity are infeasible in deployment scenarios.

Contrary to the plethora of works targeting sum rate maximization, there is a relative scarcity of research in the literature dedicated to power minimization via user pairing. In particular, the globally optimum solution to the user pairing problem with the objective of transmission power minimization remains an unsolved problem. Therefore, analyzing this user pairing problem in isolation holds considerable interest, as it pertains broadly to any NOMA downlink scenario. For instance, maintaining power efficiency while achieving quality of service (QoS) requirements is of utmost importance for NTN nodes such as high altitude platforms (HAPs) [6]. This is due to the limited power supply in these scenarios, as opposed to a conventional terrestrial base station (BS). The most relevant literature on this subject is discussed in Section I-A.

A. RELATED WORK

Early works on NOMA such as [3] and [7] identified the crucial impact of user pairing on the performance of NOMA. In particular, authors in [3] investigated user pairing under fixed power allocation NOMA (F-NOMA) and cognitive-radio-inspired NOMA (CR-NOMA). A general framework of dynamic user pairing was established in this work without an explicit user pairing strategy. In [8], the conventional matching algorithm was adopted to perform user pairing to maximize the achievable sum rate (ASR). Subsequently, in [9], the authors obtained the globally optimal user pairing strategy for maximizing ASR under the minimum rate constraints formulated as OMA achievable rates. This work derived the well-known result that the sum rate is maximized by pairing the users having the highest channel gains with the users experiencing the lowest channel gains in a nested manner. *In our work, we refer to this pairing scheme as near-far (NF) pairing*. In [10], another low-complexity user pairing scheme was proposed, where users within a cluster are paired successively, based on their power demand coefficients. This approach was shown to improve the decoding success probability of user pairs.

There have been numerous works investigating user pairing in subsequent works, relying on a variety of approaches. These include particle swarm optimization (PSO) [11], deep reinforcement learning (DRL) [12], quantum-inspired evolutionary strategies [13], multi-armed bandit algorithms [14], [15], and graph-theoretic algorithms [16]. These works largely focus on maximizing the sum rate of the system under various constraints. Moreover, in [17] the authors propose an iterative algorithm, which is shown to outperform some benchmark user pairing strategies within the formulated framework in terms of maximizing the minimum downlink rate in the context of cell-free massive multiple-input multiple-output (MIMO) systems. However, even if optimality is achieved (most of the works mentioned above report sub-optimal solutions), these algorithms are generally unable to cope with the increasing number of users as complexity becomes a concern.

In [18], the authors considered various factors affecting power consumption, involving sub-channel allocation and user pairing. The overall optimization problem was shown to be NP-hard, which motivated the adoption of a relax-then-adjust (RTA) algorithm to arrive at a sub-optimal solution, and this was shown to outperform the monotonic optimization (MO) method adopted in preceding works. In this work, the rate demands of the users are kept at a constant value. Most importantly, the authors made the crucial observation that, while users with relatively large channel gain differences were more likely to be multiplexed together, the power consumption is not necessarily minimized by adopting the NF scheme. Following this observation, the authors highlighted the significance of optimal user pairing for power minimization. Additionally, authors in [19] investigated power allocation in cache-aided NOMA systems. Two novel approaches – one based on a divide-and-conquer algorithm and another based on a DRL scheme were proposed.

Furthermore, authors in [20] provided a framework for achieving energy efficiency (EE) in downlink NOMA transmission in unmanned aerial vehicle (UAV) aided networks. Diverse user pairing schemes were explored in this paper, and the observation was made that the choice of user pairing has a considerable impact on the performance of the system. Following the results of [9], the authors adopted the NF pairing scheme in their primary analysis. However, the main focus of [20] was the optimization of the UAV altitude rather than user pairing. Building on the work done in [20], the same authors formulated a joint-optimization problem in [21] that performs altitude control and user pairing aiming to minimize power consumption. In this work, the cat swarm optimization (CSO) algorithm was adopted for the user pairing task. Some performance gains were reported compared to random pairing and PSO. However,

the complexity analysis revealed that both the CSO and PSO entail substantial computational complexity. These methods also do not guarantee optimality.

B. CONTRIBUTIONS AND PAPER ORGANIZATION

To the best of our knowledge, a framework that guarantees a provably optimal power allocation method along with its user pairing strategy for the power minimization problem does not exist in the literature. Against this backdrop, we note that a deeper analysis of the structure of the problem could provide a meaningful improvement over the suboptimal and relatively high-complexity approaches to this problem explored in prior works. To this end, the main contributions of this paper can be summarized as follows.

- We propose the analytical procedure to obtain the globally optimum power allocation method and user pairing scheme that minimizes transmission power consumption under minimum data rate requirements. We first derive the optimal power allocation strategy for an arbitrary pair of NOMA users. Then, we rigorously prove that the optimal pairing strategy for $K = 2$ user pairs generalizes for any K. Notably, the optimum pairing can be extracted with the proposed technique with $\mathcal{O}(2K)$ time complexity, growing linearly with the number of users in the system.
- We perform supplementary analyses that provide theoretical guarantees on power utilization under the problem constraints. Specifically, we focus on two critical aspects of the system. First, we prove that the power budget allocated per RB supports the required data rates for any arbitrary pair of users. Secondly, we determine the minimum power consumption threshold for the system. This threshold serves as an indicator of the lowest achievable power consumption that satisfies the data rate requirements, and its proximity to the operational region of practical systems can be easily computed with the users' channel state information.
- Simulation results show the potential of the proposed approach in achieving substantial power savings in typical deployment scenarios. In particular, 63% of the total power budget is saved at a mean received signal-to-noise ratio (SNR) of 15.7 dB among the users. The impact of various environmental profiles, along with variable amounts of users is also presented.
- We further analyze and discuss the impact of user pairing and associated objectives on the energy efficiency (EE) of NOMA transmission compared to the EE achieved through OMA transmission. We demonstrate that the EE gain of NOMA transmission is improved more than sixfold at convergence by adopting the power minimization approach studied in this work, in comparison to adopting the sum rate maximization approach found in the literature, such as in [9] and [20].

The remainder of this paper is organized as follows. In Section II, we introduce the system model, and provide the mathematical formulation of the joint power allocation and user pairing problem. In Section III we propose the analytical approach to obtain the user pairing strategy along with the power allocation method to solve this optimization problem. Further analyses regarding the analytical bounds of power utilization are also presented. Performance evaluation of the proposed pairing scheme with simulation results under typical deployment scenarios is presented in Section IV, followed by concluding remarks in Section V.

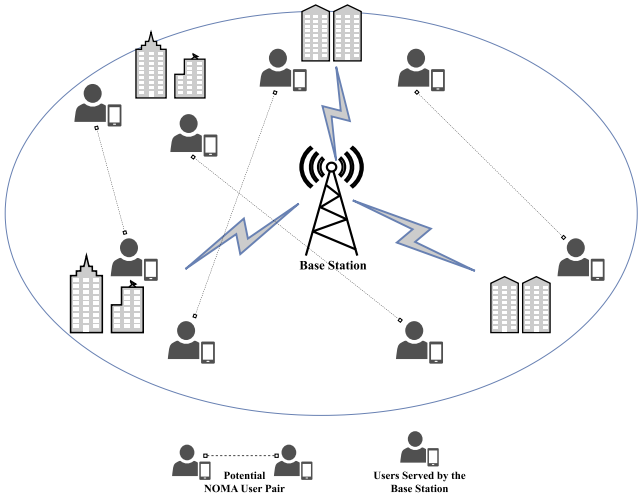

FIGURE 1. An illustration of the 2-user downlink NOMA transmission model, indicating possible user pairing configurations.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

We consider a downlink scenario with $2K$ users randomly distributed in a cell, where 2-user NOMA transmission is employed, resulting in K user pairs. In this model, it is assumed that the paired users receive the signal in the same RB (in time, frequency, or code), and different RBs are allocated orthogonal resources between themselves [3]. A generic model of this scenario is depicted in Fig. 1. For any pair of users that shares an RB constituted by individual users m, n, among these $2K$ users, the superimposed signal s transmitted by the BS can be expressed as

$$s = \sqrt{\alpha_n P_{\max}} s_n + \sqrt{\alpha_m P_{\max}} s_m, \tag{1}$$

where $s_k, k \in \{m, n\}$ indicates the signal intended for user k, and $E(|s_k|^2) = 1$, while $P_{\max}$ is the total power allocated to the RB and the ratios of $P_{\max}$ allocated to the user k are indicated by α_k.

The signal received by user n is

$$y_n = h_n\left(\sqrt{\alpha_n P_{\max}} s_n + \sqrt{\alpha_m P_{\max}} s_m\right) + \tilde{n}_n, \tag{2}$$

and the signal received by user m is

$$y_m = h_m\left(\sqrt{\alpha_n P_{\max}} s_n + \sqrt{\alpha_m P_{\max}} s_m\right) + \tilde{n}_m, \tag{3}$$

where $h_k, k \in \{m, n\}$ is their channel gain defined in (4) and $\tilde{n}_k$ denotes the noise at user k with power spectral density (PSD) σ^2.

The channel between the $k-$th user and the BS, denoted by h_k, can be expressed as [22]

$$h_k = \frac{\beta_k}{\sqrt{d_k{}^{\omega}}}, \tag{4}$$

where β_k is the Rician distributed small-scale fading coefficient experienced by user k, d_k is the distance between the user k and the BS, while ω denotes the path loss factor.

In general, we represent individual users in the system by $u_k, k \in \{1, 2, \ldots, 2K\}$. As in works such as [3], [9], [20], among many others, the channel gains corresponding to all $2K$ users in the system are assumed to be known at the BS such that $|h_1|^2 \geq |h_2|^2 \geq \cdots \geq |h_{2K}|^2 > 0$. Therefore, u_1 is the user with the highest channel gain, and u_{2K} has the lowest channel gain. In reference to a pair of users m and n, we assume that $|h_n|^2 \geq |h_m|^2$. In other words, user n is assumed to be the "strong" user in any given pair sharing an RB. This enables user n to perform SIC to remove the interference caused by the signal belonging to user m, then decode its own message. Meanwhile, the signal received by user m is decoded directly by treating $h_m\sqrt{\alpha_n P_{\max}}s_n$ as noise [3], [8], [9], [20], [21], [23]. Furthermore, the total power budget of the system that can be utilized across the K RBs is denoted by $\tilde{P}_{\max} \triangleq K \times P_{\max}$.

Based on the above, and assuming that the communication resources (e.g., bandwidth) are distributed evenly among the RBs, the achievable individual rates, $R_k^{\{m,n\}}, k \in \{m, n\}$ for the users m and n that form a pair can be expressed, respectively, as [3], [8], [9]

$$R_m^{\{m,n\}} = \log_2\left(1 + \frac{\alpha_m |h_m|^2}{\alpha_n |h_m|^2 + \gamma^{-1}}\right), \tag{5}$$

$$R_n^{\{m,n\}} = \log_2\left(1 + \alpha_n \gamma |h_n|^2\right), \tag{6}$$

where $\gamma \triangleq \frac{P_{\max}}{\sigma^2}$ defines the transmit SNR [7].

In contrast, in an OMA scheme, $2K$ RBs are required in order to provide coverage for the $2K$ users. Then, assuming that the resources are distributed evenly among the RBs as in our NOMA transmission model, this incurs a multiplexing loss factor of "$\frac{1}{2}$" on each user [9], assuming that the same amount of total resources are utilized in both cases.[1] Therefore, the achievable rate for any user k in an OMA scheme is given by

$$R_k^{\text{OMA}} = \frac{1}{2}\log_2\left(1 + \gamma |h_k|^2\right). \tag{7}$$

This formulation has been widely adopted in prior works in order to compare the performance of NOMA against OMA. Most notably, in major works on user pairing such as [3], [8], [9], and several subsequent works, such as [15], [20], [21], [23].

[1]This is in reference to the communication resources such as bandwidth, and does not include power.

B. PROBLEM FORMULATION

Our goal in this paper is to analytically derive the power allocation method and the user pairing strategy that minimizes power consumption at the BS within the context of the 2-user NOMA downlink scenario. The power allocation task involves choosing the power allocation coefficients α_n and α_m for the users in each RB. Second, we must select a suitable user pairing scheme, which is particularly challenging to address due to the combinatorial nature of the search space. Specifically, with $2K$ users in the system, there exist $\frac{(2K)!}{2^K \times K!}$ possible pairing schemes. In this subsection, we formulate this problem mathematically as a joint problem in power allocation and user pairing.

Given that there are K RBs to be shared,—each by 2 users—all possible user pairing schemes can be represented by a binary matrix with dimensions $2K \times 2K$, which we denote by $\mathbf{U}$. We construct this matrix such that the columns $m \in \{1, \ldots, 2K\}$ and rows $n \in \{1, \ldots, 2K\}$ are indexed by the users according to the descending order of their channel gains. E.g., the first row and column denoted by $\mathbf{U}_{1,*}$ and $\mathbf{U}_{*,1}$ respectively, are reserved for u_1. Likewise, the $2K$-th row and column, $\mathbf{U}_{2K,*}$ and $\mathbf{U}_{*,2K}$ are reserved for u_{2K}. Then the element in the n-th row of the m-th column of $\mathbf{U}$ can be extracted as $u_{n,m}$ in the following manner:

$$u_{n,m} = \begin{cases} 1, \text{ if user } n \text{ paired with user } m, & |h_n|^2 \geq |h_m|^2 \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

Hence, the general form of the optimization problem is formulated as

$$\underset{\{\alpha_n, \alpha_m, \mathbf{U}\}}{\text{minimize}} \quad \sum_{n=1}^{2K}\sum_{m=n+1}^{2K} u_{n,m}\left(\alpha_n + \alpha_m\right) \tag{9}$$

s.t.

$$\text{C1:} \quad u_{n,m}(\alpha_m + \alpha_n) \leq 1, \quad 1 \leq m, n \leq 2K \tag{9a}$$

$$\text{C2:} \quad R_n^{\{m,n\}} \geq R_n^{\text{OMA}}, \quad 1 \leq n \leq 2K \tag{9b}$$

$$\text{C3:} \quad R_m^{\{m,n\}} \geq R_m^{\text{OMA}}, \quad 1 \leq m \leq 2K \tag{9c}$$

$$\text{C4:} \quad u_{n,m} \in \{0, 1\}, \quad 1 \leq m, n \leq 2K \tag{9d}$$

$$\text{C5:} \quad u_{n,m} = u_{m,n}, \quad 1 \leq m, n \leq 2K \tag{9e}$$

$$\text{C6:} \quad u_{n,n} = 0, \quad 1 \leq n \leq 2K \tag{9f}$$

$$\text{C7:} \quad \sum_{n=1}^{2K} u_{n,m} = 1, \quad 1 \leq m \leq 2K \tag{9g}$$

$$\text{C8:} \quad \sum_{m=1}^{2K} u_{n,m} = 1, \quad 1 \leq n \leq 2K \tag{9h}$$

$$\text{C9:} \quad \alpha_m, \alpha_n > 0, \quad 1 \leq m, n \leq 2K. \tag{9i}$$

The constraint C1 ensures that the power budget $P_{\max}$ is not exceeded for any given pair, while the constraints C2 and C3 impose the minimum data rate requirement on each user to be their respective OMA-achievable rate as expressed in (7). Constraint C4 guarantees that the solution space contains only valid entries of $\mathbf{U}$ as required by (8). It can be easily seen that the upper and lower triangular portions of $\mathbf{U}$ contain the same information. For the calculation of the total

power utilized by a pairing scheme, we only consider the upper triangular portion of $\mathbf{U}$, as indicated by the objective function in (9). Therefore, the constraint C5 specifies that the solution space only contains permutations of the matrix such that $\mathbf{U}^T = \mathbf{U}$, and C6 ensures that the diagonal elements of $\mathbf{U}$ are set to 0 as users cannot form pairs with themselves. Additionally, this also guarantees that the user n with the higher channel gain in each pair is always indexed by row n of $\mathbf{U}$, while user m, with the lower channel gain, is always indexed by column m. Finally, with C7 and C8, we ensure that each row and column of $\mathbf{U}$ individually sum to 1, as each user must be uniquely contained among one of the K pairs.

Problem (9) is a non-convex mixed-integer programming problem [9], [21]. To solve it, we adopt a similar approach to that in [9], [12], and [24], by decoupling the power allocation task and the user pairing task to arrive at the full solution. However, it is important to note that while we draw inspiration from works such as [9], [12], and [24], our specific objective function and problem constraints differ significantly. These distinctions introduce substantial asymmetries and challenges that render the specific analytical methods employed in these works unsuitable for our scenario.

III. POWER ALLOCATION AND USER PAIRING

In this section, we propose the analytical procedure to solve the problem in (9). First, we propose the optimal power allocation method for any arbitrary pair of NOMA users. In light of this, we investigate the impact of user pairing on the overall power utilization of the system and propose the optimal user pairing scheme. Additionally, we also derive theoretical bounds on power utilization under the problem constraints.

A. OPTIMAL POWER ALLOCATION FOR AN ARBITRARY PAIR OF USERS

We derive the power allocation directly from the constraints, similar to the approaches taken in [9], [12], [20], and [21]. We require that each user in the system achieves their OMA achievable rate expressed in (7). Then, we know that the following inequalities must hold:

$$\log_2\left(1+\frac{\alpha_m|h_m|^2}{\alpha_n|h_m|^2+\frac{1}{\gamma}}\right)\geq R_m^{\mathrm{OMA}}, \tag{10}$$

$$\log_2\left(1+\alpha_n\gamma|h_n|^2\right)\geq R_n^{\mathrm{OMA}}. \tag{11}$$

After some algebraic manipulations, the lower bounds of α_m and α_n, denoted by $\alpha_m^{\min}$ and $\alpha_n^{\min}$, respectively, can be derived as

$$\alpha_m^{\min}=\frac{2^{R_m^{\mathrm{OMA}}}-1}{\gamma|h_m|^2}\left(1+\frac{2^{R_n^{\mathrm{OMA}}}-1}{|h_n|^2}|h_m|^2\right), \tag{12}$$

and

$$\alpha_n^{\min}=\frac{2^{R_n^{\mathrm{OMA}}}-1}{\gamma|h_n|^2}. \tag{13}$$

Observing (10) and (13), we define the following intermediate variables Φ_k and ϕ_k for u_k as

$$\Phi_k\triangleq\frac{2^{R_k^{\mathrm{OMA}}}-1}{\gamma|h_k|^2}, \tag{14}$$

and

$$\phi_k\triangleq\Phi_k\gamma|h_k|^2. \tag{15}$$

Then, $P_{\min}^{\{m,n\}}\triangleq\alpha_n^{\min}+\alpha_m^{\min}$ can be written as

$$P_{\min}^{\{m,n\}}=\Phi_n+\Phi_m+\Phi_n\phi_m. \tag{16}$$

In summary, (16) is the minimum transmit power, as a portion of $P_{\max}$ required by an arbitrary pair of users $\{m,n\}$, that meets their respective data rate demands, as required by the constraints C2 and C3 of the optimization problem in (9).

Additionally, we must ensure that the constraint C1 in (9a) is never violated under the optimal power allocation strategy proposed above. This is concisely articulated in Theorem 1.

Theorem 1: The maximum available power per RB denoted by $P_{\max}$ is sufficient to satisfy the rate requirements R_m^{OMA} and R_n^{OMA} for the users m and n that make up the NOMA pair $\{m,n\}$. Therefore, the condition $\alpha_m+\alpha_n\leq 1$ is always guaranteed, by choosing $\alpha_m=\alpha_m^{\min}$ and $\alpha_n=\alpha_n^{\min}$.

Proof: We begin by expressing Φ_k in its most explicit form by substituting the right-hand side of (7) into (17) as

$$\Phi_k=\frac{\sqrt{1+\gamma|h_k|^2}-1}{\gamma|h_k|^2}. \tag{17}$$

Notice that the term $\Phi(x)\triangleq\frac{\sqrt{1+x}-1}{x}$ is a monotonically decreasing function in x. In other words, the first derivative $\Phi'(x)<0,\quad\forall x\in\mathbb{R}$ and $x>0$. Additionally, recall that according to the NOMA principles postulated in Section II-A, we require that $|h_n|^2\geq|h_m|^2$. Therefore, we have $\Phi_m\geq\Phi_n$, (and $\phi_n\geq\phi_m$). Consequently, the inequality

$$\Phi_n+\Phi_m+\Phi_n\phi_m\leq\Phi_m+\Phi_m+\Phi_m\phi_m \tag{18}$$

must hold.

With the aforementioned definition of $\Phi(x)$, and using $\phi(x)\triangleq x\Phi(x)$, if we let $x=\gamma|h_m|^2$, the right-hand side of (18), which is a function of $\gamma|h_m|^2$ can be written as $2\Phi(x)+\Phi(x)\phi(x)$. By evaluating the right-hand side of (18) with these conditions, we get

$$\begin{aligned}2\Phi(x)+\Phi(x)\phi(x)&=2\frac{\sqrt{1+x}-1}{x}\\&\quad+\frac{\sqrt{1+x}-1}{x}\left(\sqrt{1+x}-1\right)\\&=2\frac{\sqrt{1+x}-1}{x}+\frac{x-2\sqrt{x+1}+2}{x}\\&=1,\end{aligned} \tag{19}$$

$\forall x\in\mathbb{R}$ and $x>0$.

Therefore, we obtain the upper bound of $P_{\min}^{\{m,n\}}$ as 1, since

$$\Phi_n+\Phi_m+\Phi_n\phi_m\leq 1. \tag{20}$$

This completes the proof. ■

In addition to obtaining the proof for Theorem 1, we make the following observation based on the above analysis. The upper bound, $P_{\min}^{\{m,n\}} = 1$ is reached for any γ when $|h_n|^2 = |h_m|^2$. However, subject to the exact values of the variables, $P_{\min}^{\{m,n\}}$ may approach 1 when $\gamma|h_n|^2 \approx \gamma|h_m|^2$. This is different from the aforementioned case which is only dependent on $|h_n|^2$ and $|h_m|^2$. This is because even when the channel gains have relatively high variation between them, an extremely small value of γ, induced by a sufficiently large value of σ^2, can be the dominating term.

As expected, larger values of the channel coefficients at a given value of γ imply lower power requirements, thereby driving the power allocation coefficients $\alpha_m^{\min}$ and $\alpha_n^{\min}$ down. However, since $\gamma > 0$ and $|h_k| > 0$ for $k \in \{1, 2, \ldots, 2K\}$, it is easily seen that (10) and (13) also satisfy the constraint C9 in (9i). In particular, these conditions require that the numerators and denominators of (10) and (13) are always positive, resulting in positive values of $\alpha_m^{\min}$ and $\alpha_n^{\min}$ in all cases.

B. USER PAIRING

The power allocation strategy provided in the closed-form expression (16), upheld by Theorem 1, which ensures that the power budget is never exceeded, effectively solves the optimization problem in (9) when $K = 1$. In this case, there is a single pair of users resulting in just one valid permutation of the user pairing matrix $\mathbf{U}$. Furthermore, for $K \geq 2$ the pairwise power allocation strategy for minimum power utilization given an arbitrary pairing scheme can also be retrieved by (16). In other words, for any given pair of users, the optimum values for the optimization variables α_m and α_n in (9) can be readily extracted by adopting $\alpha_m^{\min}$ and $\alpha_n^{\min}$, from (10) and (13), respectively. Building upon these results, in this subsection we propose the user pairing strategy that optimally solves the optimization problem in (9) for any K. Hence, here we are concerned with optimizing the remaining variable in (9), namely $\mathbf{U}$.

As a crucial first step, consider Lemma 1.

Lemma 1: Subject to the same constraints, the permutation of $\mathbf{U}$ that is required to solve the optimization problem in (9) is identical to that which solves the minimization problem with the objective function

$$\underset{\{\mathbf{U}\}}{\text{minimize}} \quad \sum_{n=1}^{2K} \sum_{m=n+1}^{2K} u_{n,m}\Big(\Phi_n \phi_m\Big), \tag{21}$$

which only considers the interaction term $\Phi_n\phi_m$.

Proof: After Section III-A, we know that the minimum power utilized for any possible pair is given by (16). Therefore, the objective function of the minimization problem in (9) can now be written as

$$\underset{\{\mathbf{U}\}}{\text{minimize}} \quad \sum_{n=1}^{2K} \sum_{m=n+1}^{2K} u_{n,m}\Big(\Phi_n + \Phi_m + \Phi_n\phi_m\Big). \tag{22}$$

From this, observe the fact that the first two terms of (16), namely Φ_n and Φ_m must form a constant term across all possible permutations of $\mathbf{U}$. Following this observation, an equivalent form of (22) is

$$\underset{\{\mathbf{U}\}}{\text{minimize}} \quad \sum_{n=1}^{2K} \sum_{m=n+1}^{2K} u_{n,m}\Big(\Phi_n \phi_m\Big) + \sum_{k=1}^{2K} \Phi_k, \tag{23}$$

where the constant term formed by Φ_n and Φ_m is isolated as a separate summation, which does not depend on the variable $\mathbf{U}$. In other words, Φ_k for each user $k, k \in \{1, \ldots, 2K\}$ appears exactly once when the total power consumption of a pairing scheme is calculated. Consequently, the term $\sum_{k=1}^{2K} \Phi_k$ can be treated as a constant when different pairing schemes are compared in terms of their total power consumption. As a result, the permutation of $\mathbf{U}$ that solves the optimization problem (9) also solves the optimization problem with the objective function (21) as asserted by the Lemma. This completes the proof. ■

Henceforth, we focus on finding the optimal pairing for (21) to obtain the optimal pairing scheme for our overall objective of solving problem (9). Furthermore, we employ prior definitions of the terms Φ_k, ϕ_k, $\Phi(x)$, and $\phi(x)$. Note that the product $\Phi_n\phi_m$ in the objective function (21) takes the form $\Phi(x)\phi(y)$, where $x = \gamma|h_n|^2$ and $y = \gamma|h_m|^2$, while $u_{n,m}$ refers to the pair formed by users m and n, where $|h_n|^2 \geq |h_m|^2$. Then, based on the fact that $|h_k|^2 \geq |h_{k+1}|^2$, we can establish that

$$\Phi_k \leq \Phi_{k+1}, \tag{24}$$

and

$$\phi_k \geq \phi_{k+1}, \tag{25}$$

for $k \in \{1, \ldots, 2K - 1\}$.

1) OPTIMAL PAIRING SCHEME WITH $2K = 4$ USERS

Now, we turn our attention towards deriving the optimal pairing strategy for the case when $K = 2$, corresponding to four users. There are three possible pairing schemes in this scenario, which result in the following options.

- *Option 1:* u_1 pairs u_2, u_3 pairs $u_4 \rightarrow S_1 \triangleq \Phi_1\phi_2 + \Phi_3\phi_4$,
- *Option 2:* u_1 pairs u_3, u_2 pairs $u_4 \rightarrow S_2 \triangleq \Phi_1\phi_3 + \Phi_2\phi_4$,
- *Option 3:* u_1 pairs u_4, u_2 pairs $u_3 \rightarrow S_3 \triangleq \Phi_1\phi_4 + \Phi_2\phi_3$,

where $S_P, P \in \{1, 2, 3\}$ is defined as the explicit sums obtained by the objective function in (21) when the pairing scheme is represented by Option P. The pairing schemes outlined in Options 1, 2, and 3 are illustrated in Fig. 2.

Theorem 2: In a NOMA system with four users, the pairing scheme that minimizes the power consumption is Option 2, i.e., $S_2 \leq S_1$ and $S_2 \leq S_3$.

Proof: First, let us explicitly rewrite the inequalities in (24) and (25) for the case with four users for our visual aid as

$$\Phi_1 \leq \Phi_2 \leq \Phi_3 \leq \Phi_4, \tag{26}$$

and

$$\phi_1 \geq \phi_2 \geq \phi_3 \geq \phi_4. \tag{27}$$

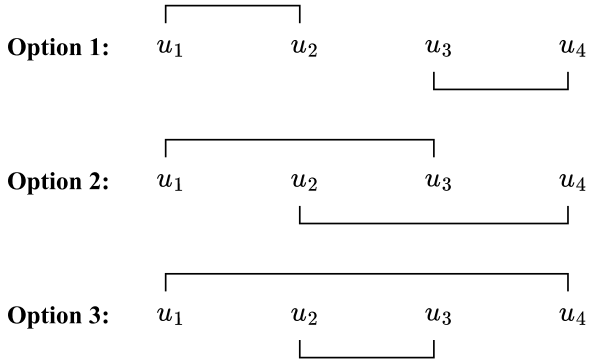

FIGURE 2. Illustration of user pairings obtained from Options 1, 2, and 3.

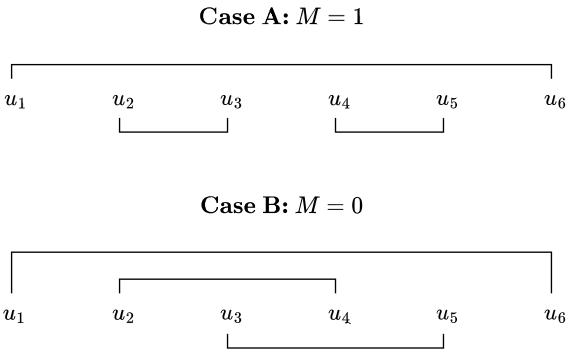

FIGURE 3. Illustration of a six-user scenario where $M = 1$ for the initial user pairing scheme in Case A, which is immediately improved to Case B through the application of Lemma 2, where $M = 0$.

Now, let us compare Option 1 and Option 2. If we let $a = \Phi_1\phi_2$, $b = \Phi_3\phi_4$, $c = \Phi_1\phi_3$, and $d = \Phi_2\phi_4$, asserting the opposite of Theorem 2 we get

$$\begin{aligned} S_1 &< S_2, \\ \Phi_1\phi_2 + \Phi_3\phi_4 &< \Phi_1\phi_3 + \Phi_2\phi_4, \\ a + b &< c + d. \end{aligned} \tag{28}$$

According to (26) and (27), it is clear that $a \geq c$ and $b \geq d$. We see that this does not hold, implying that the left-hand side of (14) must be greater than or equal to the right-hand side. Therefore, $S_2 \leq S_1$.

Now, let us compare Options 2 and 3. We can apply the same procedure here. Asserting that $S_2 > S_3$, we get:

$$\begin{aligned} S_3 &< S_2, \\ \Phi_1\phi_4 + \Phi_2\phi_3 &< \Phi_1\phi_3 + \Phi_2\phi_4, \\ \Phi_2[\phi_3 - \phi_4] &< \Phi_1[\phi_3 - \phi_4]. \end{aligned} \tag{29}$$

Since $\phi_3 \geq \phi_4$ and $\Phi_2 \geq \Phi_1$, this also results in a contradiction. Therefore, $S_2 \leq S_3$. ■

2) OPTIMAL USER PAIRING SCHEME FOR ANY $2K = N$

We now generalize Theorem 2 to the case with any even number of users. First, we focus on proving Lemma 2. Before that, the following definitions are provided.

$$\mathcal{G}_N \triangleq \{u_1, u_2, \ldots, u_K\}, \tag{30}$$

$$\mathcal{G}_M \triangleq \{u_{K+1}, u_{K+2}, \ldots, u_{2K}\}, \tag{31}$$

where $\mathcal{G}_N$ represents the set of users in the system with the K highest channel gains, and $\mathcal{G}_M$ represents the set comprising the users with the K lowest channel gains.

Lemma 2: In a NOMA system with $2K$ users, the pairing scheme that solves the optimization problem in (9) does not contain any pairings among users in $\mathcal{G}_N$, or among users in $\mathcal{G}_M$.

Proof: Lemma 2 states that the optimal pairing scheme cannot be a pairing scheme in which any pairings exist within $\mathcal{G}_M$, or equivalently, within $\mathcal{G}_N$. For any number M of pairings within $\mathcal{G}_N$, there must also exist M pairings within $\mathcal{G}_M$. Now, if we compare any single pair found within $\mathcal{G}_N$ with any pair within $\mathcal{G}_M$ keeping all other pairings unchanged, this results in a scenario with four users where the pairing scheme is given by Option 1. As a consequence of Theorem 2, we can immediately find a better overall pairing scheme by swapping over the two pairs in consideration to form pairs in the form of Option 2. This decreases M by 1. This can be done for all M pairs that may exist within $\mathcal{G}_N$ and equivalently in $\mathcal{G}_M$, until $M = 0$, in which case each pair must contain a user from $\mathcal{G}_N$ and one from $\mathcal{G}_M$. Then, it must be the case that any pairing scheme with $M = 0$ utilizes less power than that if $M \geq 1$. Thus, it requires that the optimal pairing cannot allow pairs to be found within $\mathcal{G}_N$ and $\mathcal{G}_M$, as asserted by Lemma 2. ■

As an example of Lemma 2, consider the arbitrary pairing scheme illustrated for the six-user scenario depicted in Fig. 3 as Case A. In this case, $\mathcal{G}_N = \{u_1, u_2, u_3\}$ and $\mathcal{G}_M = \{u_4, u_5, u_6\}$. There is one user pair found within each set. Namely, $u_{2,3}$ within $\mathcal{G}_N$ and $u_{4,5}$ within $\mathcal{G}_M$. Therefore, $M = 1$. When considering these two pairs in isolation, they are in the form of Option 1, since $|h_2|^2 \geq |h_3|^2 \geq |h_4|^2 \geq |h_5|^2$. Then, according to Theorem 2, the overall power consumption of Case A can be immediately lowered by switching these two pairs to take the form of Option 2 which results in the overall pairing scheme depicted as Case B in Fig. 3 where M is now reduced to 0.

Theorem 3: In a NOMA system with $2K$ users in which the channel gains of the users are known, the pairing scheme that solves the minimization problem in (9) is to pair u_k with u_{k+K}, $k \in \{1, 2, \ldots, K\}$. Therefore, the user pairing matrix $\mathbf{U}$ contains the elements

$$u_{n,m} = \begin{cases} 1, & n = k, m = k + K, \quad k \in \{1, 2, \ldots, K\}, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

Proof: After Lemma 2, we are restricted to only considering user pairs between $\mathcal{G}_N$ and $\mathcal{G}_M$. Based on the sequences

$$\begin{aligned} \Phi_{\mathcal{G}_N} &\triangleq \{\Phi_1, \Phi_2, \ldots, \Phi_K\}, \\ \phi_{\mathcal{G}_M} &\triangleq \{\phi_{K+1}, \phi_{K+2}, \ldots, \phi_{2K}\}, \end{aligned} \tag{33}$$

the optimal pairing is given by the permutations of $\Phi_{\mathcal{G}_N}$ and $\phi_{\mathcal{G}_M}$ that minimize their dot product $\Phi_{\mathcal{G}_N} \cdot \phi_{\mathcal{G}_M}$. The rearrangement inequality [25] gives the lower bound

when both sequences are sorted oppositely. Therefore, as established by the inequalities in (26) and (27), the minimum sum is given by

$$\Phi_1\phi_{K+1} + \Phi_2\phi_{K+2} + \cdots + \Phi_K\phi_{2K}. \tag{34}$$

This can only be achieved by the pairing scheme denoted by $u_{k,k+K}$, as stated by Theorem 3. This result directly solves the minimization problem in (21), and consequently also the minimization problem in (9) due to Lemma 1. Note that the rearrangement inequality can also be utilized to prove $S_2 < S_3$. ■

As emphasized in [4] and [5], low-complexity design for user pairing is necessary for NOMA's potential as a feasible component of future wireless communication systems. In this regard, the above approach which proves global optimality in minimizing power utilization is an attractive solution, since the complexity of the procedure to obtain the optimal pairing is $\mathcal{O}(2K)$. Given the sorted list of channel gains, the only complexity that arises is the computation of the pairwise power allocation according to (10) and (13), which grows linearly with the number of users in the system, similar to the case in [9]. Within this context, linear complexity in the user pairing task is considered to be within the low-complexity regime. This is in contrast to other alternatives, such as matching theory based user pairing [8], swarm optimization techniques [11], [21], multi-armed bandit algorithms [14], [15], graph theoretic user pairing [16], iterative approaches such as in [17], monotonic optimization [18] which incur significantly higher computational complexities. Also, the proposed techniques do not entail strict guarantees of global optimality for the considered objective functions, in most of these cases. Nevertheless, the computational complexity of our procedure can be further reduced owing to the fact that the power allocation required for each user pair is independent of all other pairs. As a result, once the optimal pairs have been identified according to Theorem 3, the power allocation can be carried out in parallel. Hence, the scalability of the system can be greatly improved to accommodate massive amounts of users, by leveraging any available parallel computation enabled hardware architectures at the transmitter. In the ideal case where enough parallel computing nodes are available, near-constant time performance can be achieved when power can be allocated to the K resource blocks simultaneously.

C. TIGHT UPPER BOUND ON POWER SAVINGS

With the optimal solution for (9) in place, as a supplementary piece of analysis, we aim to establish a tight upper bound on the power that can be saved per RB by utilizing NOMA compared to an OMA scheme. Equivalently, this involves understanding the lower bound of $P_{\min}^{\{m,n\}}$.

We know that $P_{\min}^{\{m,n\}} \leq 1$ based on Theorem 1, and so the quantity $(1 - P_{\min}^{\{m,n\}})P_{\max}$ can be thought of as the power that is saved per RB in comparison to OMA.

Theorem 4: $\lim_{\gamma\to\infty} P_{\min}^{\{m,n\}} = \frac{|h_m|}{|h_n|}$. In other words, at least $\frac{|h_m|}{|h_n|}$ of the total power $P_{\max}$ allocated to the pair $\{m, n\}$ ($|h_n|^2 \geq |h_m|^2$) is utilized to meet their minimum rate requirements $R_k^{\text{OMA}}, k \in \{m, n\}$.

Proof: See Appendix V. ■

The primary implication of Theorem 4 is that the power that can be saved by a user pair sharing an RB compared to OMA is directly proportional to the difference in their channel gains. This is in line with Theorem 1, where the total power budget of the RB $P_{\max}$ is required when $|h_m|^2 = |h_n|^2$, which is directly corroborated by Theorem 4, according to which $P_{\min}^{\{m,n\}} = 1$, even as $\gamma \to \infty$ if $|h_m|^2 = |h_n|^2$.

Furthermore, the total power utilized by some pairing scheme (represented by $\tilde{P}$) as a portion of its total power budget (represented by $\tilde{P}_{\max} \triangleq K \times P_{\max}$) in the limit can be extracted from its user pairing matrix from the power utilization ratio, $\frac{\tilde{P}}{\tilde{P}_{\max}}$ as

$$\lim_{\gamma\to\infty} \frac{\tilde{P}}{\tilde{P}_{\max}} = \frac{1}{K}\sum_{n=1}^{2K}\sum_{m=n+1}^{2K} u_{m,n}\left(\frac{|h_m|}{|h_n|}\right) \tag{35}$$

After Theorem 3, we know that the quantity $\sum_{n=1}^{2K}\sum_{m=n+1}^{2K} u_{m,n}(\frac{|h_m|}{|h_n|})$ must be minimized by choosing user pairs $u_{k,k+K}$, according to (32). As a result, for any $2K$ users served by the 2-user NOMA transmission model we have adopted in this paper, the minimum power utilization ratio as $\gamma \to \infty$ is given by

$$\frac{1}{K}\sum_{k=1}^{K}\frac{|h_{k+K}|}{|h_k|}, \tag{36}$$

when the users are optimally paired. This limit can be calculated for any arbitrary pairing scheme in a similar manner.

IV. PERFORMANCE EVALUATION AND DISCUSSIONS

In this section, we comprehensively evaluate the performance of the optimal user pairing strategy through simulations. Each data point in our plots is the average performance based on 10^3 independent Monte Carlo runs of random user distributions under the stated channel model. First, the optimal pairing scheme is benchmarked against low-complexity pairing schemes commonly found in the literature. In the latter half of this section, we present further analyses and discussions regarding the achievable EE in NOMA transmission, and we compare it to OMA. The simulation parameters are listed in Table 1.

A. PERFORMANCE OF THE OPTIMAL USER PAIRING METHOD

We benchmark the power consumption of our proposed pairing scheme against the following NOMA user pairing strategies:

- The NF scheme, the well-known pairing scheme derived in [9] for maximizing the sum rate, where u_k is paired with u_{2K-k+1} for $k \in \{1, 2, \ldots, K\}$,

TABLE 1. Simulation parameters.

Parameter	Value
Number of users, N	50
Coverage radius	1000 m
Path loss exponent, ω	2, 2.7, 3, 3.5
Rician K-factor	0, 3, 5
Power budget per RB, $P_{\max}$	0 dBm
Noise PSD, σ^2	0 to -140 dBm

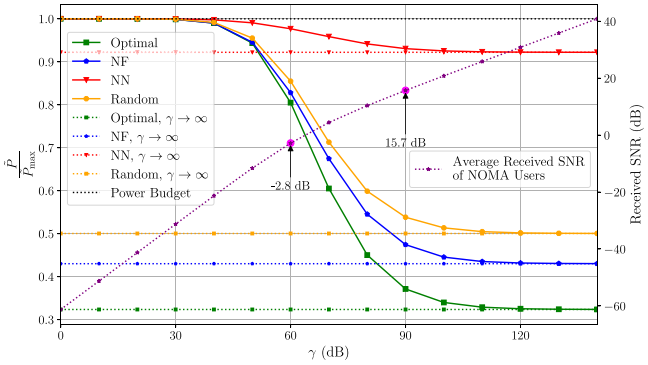

FIGURE 4. The power utilization ratio $\frac{\tilde{P}}{\tilde{P}_{\max}}$ of various pairing schemes against γ under the simulation settings in Table 1 with $\omega = 2$ and the $K-$factor fixed at 0. For each pairing scheme, $\frac{\tilde{P}}{\tilde{P}_{\max}}$ as $\gamma \to \infty$ as predicted by Theorem 4 is also marked. The secondary axis indicates the average received SNR of the users at each value of γ.

- the near-near (NN) scheme, as adopted in [10] and benchmarked in [20], where user pairs are chosen such that the user u_k is paired with the user u_{k+1}, for $k \in \{k{:}k = 2j-1, j \in \mathbb{N}, 1 \le j \le K\}$, and
- random pairing.

The quantity $\frac{\tilde{P}}{\tilde{P}_{\max}}$ defined in Section III-C is adopted as the key performance metric in our numerical simulations, which is the total transmission power consumed as a portion of the total power budget. In each case, the power allocated to each user pair is determined by (16).

In our simulations, we obtain γ, by fixing $P_{\max}$ at 0 dBm while varying the noise power σ^2 from 0 to -140 dBm, as detailed in Table 1. However, the specific values of γ do not correspond exclusively to these values. They can also be representative of any other combinations of $P_{\max}$ and σ^2 that yield the same power ratio. For example, a γ value of 90 dB is achieved with $P_{\max}$ set to 0 dBm and σ^2 at -90 dBm, as in our simulations, and the same γ can be obtained with $P_{\max} = 30$ dBm and $\sigma^2 = -60$ dBm.

In Fig. 4, we benchmark the power utilization ratio of the aforementioned pairing schemes across a wide range of γ values. In line with [9], the Rician K-factor is set to 0, which corresponds to the Rayleigh fading environment, and the path loss exponent ω is set to 2. For reference, the mean received SNR of the users is plotted in the secondary axis for each value of γ. We compute the received SNR of the user m in each pair, denoted by γ_r^m as

$$\gamma_r^m = \frac{\alpha_m |h_m|^2}{\alpha_n |h_m|^2 + \gamma^{-1}}, \tag{37}$$

and for each user n, denoted by γ_r^n, as

$$\gamma_r^n = \alpha_n \gamma |h_n|^2. \tag{38}$$

For instance, at $\gamma = 60$ dB the mean received SNR among the users is -2.8 dB, and 15.7 dB at $\gamma = 90$ dB as indicated in Fig. 4.[2] At $\gamma = 90$ dB we observe 63% power savings from the power budget with the optimal pairing scheme, which is a significant improvement over the benchmark schemes. We also indicate the limit of the power utilization ratios as $\gamma \to \infty$ for each pairing scheme, predicted by Theorem 4. At $\gamma \approx 120$ dB, we can observe $\frac{\tilde{P}}{\tilde{P}_{\max}}$ converging to these limits. Here, we can observe about 67% reduction in power consumption from the maximum power budget when the optimal pairing scheme is adopted. This can result in considerable power savings in typical deployment scenarios such as HAPs in NTNs, where efficient transmission power management is especially crucial [6]. The NF scheme achieves 57% power savings, which is also a substantial amount. In contrast, the NN scheme consumes a considerable portion of the power budget even under the most favorable conditions. This is evidenced by the "NN, $\gamma \to \infty$" curve, which indicates that 92% of the total power budget is required to satisfy the rate demands, as $\gamma \to \infty$.

Furthermore, it should be highlighted that considering the signal attenuation between the BS and the user resulting from factors such as path loss and shadowing effects, relatively high values of γ should be expected with typical values of $P_{\max}$ and σ^2 for the received SNR to be at reasonable levels, as evident from Fig 4. In particular, the noise power σ^2 is usually assumed to be ≈ -174 dBm/Hz [18], [26], [27]. As for $P_{\max}$, [20] and [21] adopted 30 dBm. Some variations of $P_{\max}$ for typical NOMA deployment scenarios are 33 dBm [12] and 23 dBm [27]. Taking the lowest possible value among these options for $P_{\max}$, we have γ (dB) = 197 dB, assuming $\sigma^2 = -174$ dB. Therefore, it is reasonable to assume that γ, at unity bandwidth, may reach as high as ≈ 200 dB. For instance, $P_{\max} = 30$ dB yields γ (dB) = 204 dB. However, the noise power is amplified proportional to the channel bandwidth. For instance, if we were to utilize a 1 MHz band, this would incur a 60 dB increase in the noise power. In any case, it should be noted that narrowband systems, typically operating at bandwidths in the range of 180 kHz to 200 kHz are usually preferred for NOMA applications such as IoT access [27], which further minimizes the possibility of bandwidth-induced impairment of γ.

Fig. 5 portrays the impact of the small-scale fading coefficient modeled by the Rician distribution with $\omega = 2$. We ignore the NN scheme and random pairing in this performance comparison. It is observed that more power is saved relative to the NF scheme in severely fading environments, which corresponds to low values of the Rician $K-$factor. This is corroborated by Theorem 4 as lower values

[2]Note that the slight non-linearity observed on the mean received SNR curve is due to the interference caused by the user n in each NOMA pair, in the calculation of γ_r^m in (38).

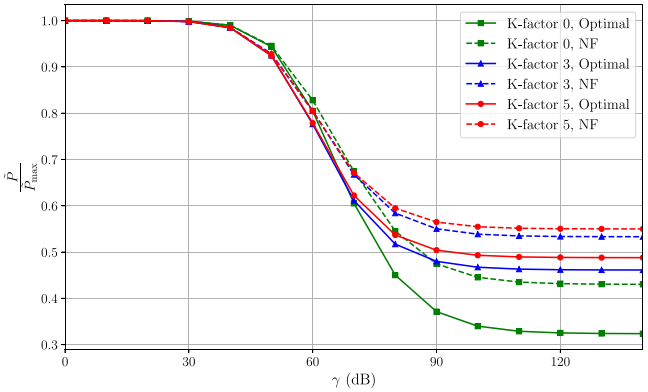

FIGURE 5. Effect of various $K-$factors on the power utilization profiles of the optimal pairing and NF pairing, for the simulation parameters in Table 1 with $\omega = 2$.

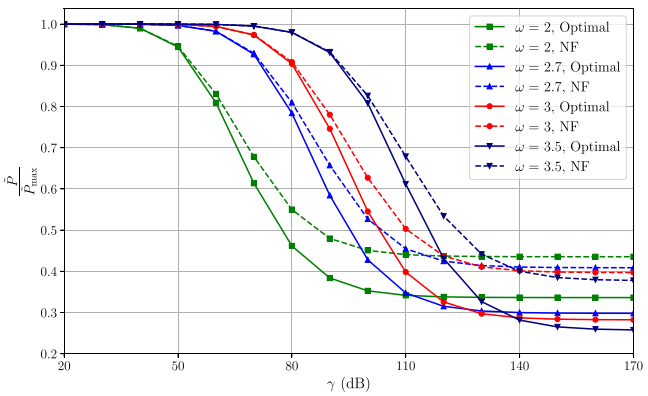

FIGURE 6. Effect of different values of the path loss exponent ω on the power utilization profile for $K-$factor = 0. Here, the noise PSD is varied from -20 dBm to -170 dBm. Other parameters are as listed in Table 1.

of $K-$factor encourage high variance among the channel gains experienced by the users.

Fig. 6 illustrates the impact of the path loss exponent ω, on the convergence speed of $\frac{\tilde{P}}{P_{\max}}$. A range of ω values are compared, which capture typical path loss dynamics of various environmental profiles, such as urban cells ($\omega \approx$ 2.7 to 3.5), and shadowed urban cells ($\omega \approx 3$ to 5) [28]. Specifically, we consider $\omega = 2, 2.7, 3$, and 3.5 in Fig. 6. In this figure, while maintaining $P_{\max}$ at 0 dBm as in previous figures, we extend the range of σ^2 to be from -20 dBm to -170 dBm. The Rician $K-$factor is set to 0. We observe that $\frac{\tilde{P}}{P_{\max}}$ converges at significantly higher γ values for larger values of ω, as expected. Furthermore, at $\gamma = 110$ dBm, the optimal pairing scheme achieves a power utilization ratio of ≈ 3.4, at $\omega = 2$ as well as $\omega = 2.7$, with $\omega = 3$ relatively close by with $\frac{\tilde{P}}{P_{\max}} \approx 0.4$. At $\omega = 3.5$ we still observe around 40% power reduction to the maximum power budget, at $\frac{\tilde{P}}{P_{\max}} \approx 0.6$. An interesting observation that follows from this is that the optimal pairing scheme, when implemented at some value of ω, can outperform the NF pairing scheme adopted at lower values of ω in terms of power savings. For instance, at γ as low as 100 dB, we observe that the optimal pairing scheme adopted in $\omega = 2.7$ attains a lower

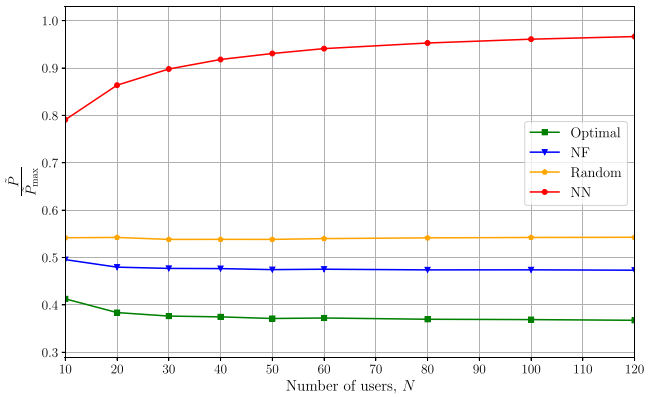

FIGURE 7. Impact of increasing the total number of users N on the power utilization profiles of various pairing schemes. System parameters in Table 1 are adopted, with $\gamma = 90$ dB, $\omega = 2$, and the $K-$factor set to 0.

$\frac{\tilde{P}}{P_{\max}}$ than the NF pairing scheme at $\omega = 2$. At $\gamma = 130$ dB, the optimal pairing schemes corresponding to all the values of ω have started to achieve more power savings than all of their NF scheme counterparts.

It should be pointed out that for both Fig. 5 and Fig. 6, the curves presented on the same plots for different channel parameters do not necessarily attain the same received SNRs or sum rates, which are directly influenced by the channel coefficients. However, the optimal pairing scheme and the benchmarked NF pairing scheme do achieve the same sum rates under the same channel parameters as is the case with Fig. 4. Thus, the discussion we have presented in these figures is deemed to be appropriate in the context of system design considerations such as power budget design.

The validity of Theorem 1 is also demonstrated from Figs. 4, 5, and 6, since the power utilization ratio $\frac{\tilde{P}}{P_{\max}}$ does not exceed its upper bound of 1.

To conclude this subsection, in Fig. 7, we demonstrate how the pairing schemes perform as the number of users increases. The NN pairing scheme suffers severely as the number of users, N increases. This is also understood through Theorem 4 as the variation between adjacent users decreases as we increase the number of users in a fixed coverage region. Guided by Fig. 4, we select the value of γ to be equal to 90 dB in this simulation, as we expect meaningful performance gains in this region, even though $\frac{\tilde{P}}{P_{\max}}$ has not fully converged to the minimum.

B. IMPACT ON ENERGY EFFICIENCY

Our results presented above demonstrate how the user pairing strategy affects the power consumption while maintaining the minimum required data rate, in 2-user NOMA downlink systems. Conversely,—as previously discussed—the authors in [9] presented the optimal pairing strategy that maximizes the sum rate with a fixed power budget. In this section, we study how the optimal user pairing strategy impacts the EE of the system in these two scenarios. The EE of transmission aims for the most efficient utilization of

power resources that achieves the data rate requirements. An optimization problem similar to the minimization problem in (9) with the same data rate constraints and maximum utilizable power budget, to maximize EE can be formulated which requires joint optimization of power control and user pairing. Authors in works such as [29], and [30] consider this to be an open problem. While the pursuit of the general solution to this problem is out of the scope of this paper, the insights obtained from the analysis and the corresponding simulation results we present here are intended as a preliminary contribution towards a potential general solution.

The EE of transmission is defined as [30]

$$\text{EE} = \frac{\text{Total sum rate at the receiver}}{\text{Total power consumption at the transmitter}} \text{ bits/J}. \tag{39}$$

We observe that the maximum EE of NOMA transmission that maximizes the sum rate (referred to as $\text{EE}_{\text{NOMA}}^{\text{SR}-\max}$) is achieved by adopting the NF pairing scheme. Also, since the power budget per RB is fixed for both scenarios, the OMA system collectively utilizes twice as much power to serve an equivalent number of users, compared to the 2-user NOMA system. Specifically, to accommodate $2K$ users, the OMA system requires $2K$ RBs, compared to the K RBs in the case of NOMA. Then, the gain in EE achieved by this approach compared to the EE of OMA (referred to as EE_{OMA}) can be calculated as

$$\begin{aligned}\frac{\text{EE}_{\text{NOMA}}^{\text{SR}-\max}}{\text{EE}_{\text{OMA}}} &= \frac{\text{NF pairing sum rate}}{\tilde{P}_{\max}} \div \frac{\text{OMA sum rate}}{2 \times \tilde{P}_{\max}} \\ &= \frac{2 \times \text{NF pairing sum rate}}{\text{OMA sum rate}},\end{aligned} \tag{40}$$

where $\tilde{P}_{\max}$ represents the total power required for NOMA transmission.

Similarly, the maximum EE of NOMA transmission that minimizes power utilization while maintaining the OMA-achievable data rate (referred to as $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$) is achieved by the pairing strategy that we derived in Section III-B. The gain in EE of this approach compared to EE_{OMA} can be computed as

$$\begin{aligned}\frac{\text{EE}_{\text{NOMA}}^{\text{PU}-\min}}{\text{EE}_{\text{OMA}}} &= \frac{\text{OMA sum rate}}{\tilde{P}^*} \div \frac{\text{OMA sum rate}}{2 \times \tilde{P}_{\max}} \\ &= \frac{2 \times \tilde{P}_{\max}}{\tilde{P}^*},\end{aligned} \tag{41}$$

where $\tilde{P}^*$ is the total power consumed by the optimal pairing scheme, and $\tilde{P}_{\max}$ is the total NOMA power budget for K RBs, twice of which is consumed in the OMA scenario, similar to the case in (41). Due to this reason, the expressions (41) and (42) are modulated by a factor of 2. Note that in these calculations only the power required for transmission, P_T is considered, and additional power consumption from the circuitry present at the transmitter, P_C is omitted for simplicity. Thus our analysis holds for $P_T \gg P_C$.

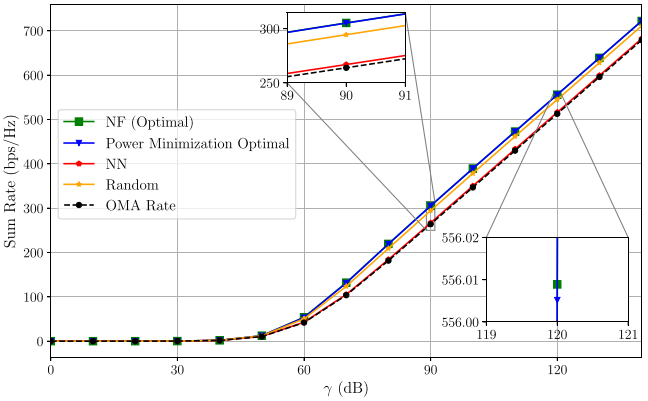

FIGURE 8. Achievable sum rates of various user pairing strategies under a fixed power budget, for the system parameters as listed in Table 1 with $\omega = 2$ and the K−factor fixed at 0. Analytical methods adopted from [9].

To aid our discussion, we have reproduced the results of [9] as illustrated in Fig. 8. The simulation was carried out in the same simulation setting and channel distributions as in Fig. 4. In addition to what is presented in [9] we also report the pairing strategy $u_{k,k+K}$ which we derived as the optimal pairing for the power minimization case that we studied, as a benchmark for this approach, which interestingly turns out to virtually overlap its optimal pairing strategy (NF pairing). In other words, the gain in the sum rate achieved through the NF pairing compared to the optimal pairing strategy for power minimization is marginal. It is observed from Fig. 4 that significant gains in sum rate with NF pairing are observed in a similar region of γ as power savings are observed in Fig. 4, namely for $\gamma > 60$ dB.

Fig. 9 offers a comparison of the gain in EE resulting from both of the above approaches relative to EE_{OMA}. Both $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ and $\text{EE}_{\text{NOMA}}^{\text{SR}-\max}$ observe twice as much EE compared to EE_{OMA} at the lower end of γ where they attain the OMA-achievable data rates (as seen on Figs. 4 and 8) by utilizing half the power budget. Furthermore, $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ presents itself with a shape that is a vertically inverted version of the power utilization curves demonstrated in Fig. 4. This is easily understood by observing (42), which is the precise reciprocal of the power utilization ratio corresponding to the optimal pairing strategy illustrated in Fig. 4, multiplied by 2. Therefore, the quantity $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ as $\gamma \to \infty$ can be inferred to be $2 \times \frac{|h_n|}{|h_m|}$ for any given pair of NOMA users, from the result of Theorem 4. Also, the maximum point of the $\text{EE}_{\text{NOMA}}^{\text{SR}-\max}$ curve corresponds to the inflection point of the sum rate achieved by its optimal user pairing scheme as seen in Fig. 8, which occurs at $\gamma \approx 60$ dB. $\text{EE}_{\text{NOMA}}^{\text{SR}-\max}$ exhibits marginal gains compared to $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ below $\gamma = 60$ dB. By $\gamma = 80$ dB however, $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ can attain an EE value around 4.5× that of OMA, while $\text{EE}_{\text{NOMA}}^{\text{SR}-\max}$ is on the decline with an EE around 2.4× higher than EE_{OMA}, on a trajectory towards its baseline value of 2. At the tail end of γ in Fig. 9, $\text{EE}_{\text{NOMA}}^{\text{PU}-\min}$ has converged

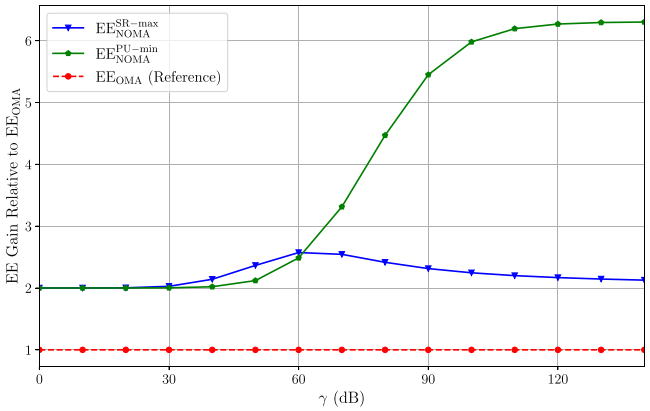

FIGURE 9. The maximum achievable $\frac{EE_{NOMA}}{EE_{OMA}}$ corresponding to the sum rate maximization approach and the power minimization approach with their respective optimal user pairing strategies for the simulation settings in Table 1 with $\omega = 2$ and the $K-$factor fixed at 0.

to a value of more than $6\times$ EE_{OMA}, demonstrating the fact that at the operational region of the γ, aiming for minimum power consumption while maintaining the OMA data rates is vastly more energy efficient compared to maximizing the sum rate by utilizing the total power budget.

V. CONCLUSION

In this paper, we investigated the impact of user pairing on the power consumption in the 2-user NOMA downlink scenario. The optimal solution we obtained has linear time complexity, which is a significant improvement over the computationally expensive methods usually found in the literature. Aided by our analysis, we demonstrated that the proposed optimal pairing strategy shows considerable performance gains compared to benchmark schemes through extensive simulation results. In particular, in a representative benchmark simulation setting, we can achieve the data rate requirements with this pairing by utilizing 37% of the power budget, saving 63%. This is approximately 10% less power utilized compared to the popular NF pairing scheme found in the literature and approximately 17% less compared to random pairing. Additionally, the effect of the propagation environment on the power consumption was studied. We demonstrated that the highest portion of power is saved with the Rayleigh fading channel, where the end users experience the highest variations in the channel coefficients among themselves. We also established the robustness of the optimal pairing scheme in environments characterized by varying path loss dynamics, along with the utility of adopting the optimal pairing scheme with a variable number of users. The performance gains in power savings attained are consistently maintained even with a large number of users (reported up to $N = 120$ users) present in the system. Finally, we carried out an analysis of the EE of NOMA transmission resulting from our objective of power minimization, benchmarked that of NOMA transmission resulting from the sum rate maximization approach studied in prior works. With a maximum transmit SNR equivalent to 80 dB, the sum rate maximization approach with its optimal pairing scheme achieves a gain in EE by a factor of 2.4 relative the EE achieved by OMA. On the other hand, the power minimization approach with the optimal pairing scheme we derived in this work exceeds the EE of OMA by a factor of 4.5, demonstrating a substantial improvement in the EE achieved. The natural extension of our research for future work is to investigate whether a globally optimal strategy for joint power control and user pairing exists as a separate solution to the EE maximization objective. Another interesting avenue to explore is to study the performance of the proposed scheme within the context of cooperative NOMA systems.

APPENDIX
PROOF OF THEOREM 4

Proof: We prove Theorem 4 by evaluating the expression for $P_{\min}^{\{m,n\}}$ from (16) as $\gamma \to \infty$. The full expression is as follows:

$$\begin{aligned}\lim_{\gamma\to\infty} P_{\min}^{\{m,n\}} &= \Phi_n + \Phi_m + \Phi_n\phi_m \\ &= \lim_{\gamma\to\infty}\left[\frac{\sqrt{1+\gamma|h_n|^2}-1}{\gamma|h_n|^2} + \frac{\sqrt{1+\gamma|h_m|^2}-1}{\gamma|h_m|^2}\right. \\ &\quad \left. + \frac{\sqrt{1+\gamma|h_n|^2}-1}{\gamma|h_n|^2}\left(\sqrt{1+\gamma|h_m|^2}-1\right)\right] \quad (42)\end{aligned}$$

Since we are taking the limit in γ as the varying term, both $|h_m|^2$ and $|h_n|^2$ are treated as constants. If we let $x = \gamma$, $a = |h_m|^2$, and $b = |h_n|^2$, then, (42) can be rewritten as

$$\begin{aligned}\lim_{\gamma\to\infty} P_{\min}^{\{m,n\}} &= \lim_{x\to\infty}\left[\frac{\sqrt{1+bx}-1}{bx} + \frac{\sqrt{1+ax}-1}{ax}\right. \\ &\quad \left. + \frac{\sqrt{1+bx}-1}{bx}\left(\sqrt{1+ax}-1\right)\right] \\ &= \lim_{x\to\infty}\frac{\sqrt{1+bx}-1}{bx} + \lim_{x\to\infty}\frac{\sqrt{1+ax}-1}{ax} \\ &\quad + \lim_{x\to\infty}\frac{\sqrt{1+bx}-1}{bx}\left(\sqrt{1+ax}-1\right). \quad (43)\end{aligned}$$

Notice that the first two terms in (43) can be written in the form

$$\lim_{x\to\infty}\frac{\sqrt{1+cx}-1}{cx} = \lim_{x\to\infty}\frac{1}{\sqrt{1+cx}+1} = 0, \quad (44)$$

for $c \in \{a, b\}$, which tends to 0 as $x \to \infty$.

Therefore, we focus on evaluating the limit of the third term in (43), which can be broken down further as

$$\begin{aligned}\lim_{\gamma\to\infty} P_{\min}^{\{m,n\}} &= \lim_{x\to\infty}\left[\frac{\sqrt{1+bx}-1}{bx}\left(\sqrt{1+ax}-1\right)\right] \\ &= \lim_{x\to\infty}\frac{1}{bx} - \lim_{x\to\infty}\frac{\sqrt{1+ax}}{bx} \\ &\quad - \lim_{x\to\infty}\frac{\sqrt{1+bx}}{bx} + \lim_{x\to\infty}\frac{\sqrt{(1+ax)(1+bx)}}{bx}. \quad (45)\end{aligned}$$

It is easily seen that the first term tends to 0.

The second and third terms can be evaluated by applying L'Hôpital's rule [31]. These terms also tend to 0 in the limit. For completion, some steps of this procedure are shown below.

$$\begin{aligned}
&\lim_{x\to\infty}\left[\frac{\sqrt{1+ax}}{bx}+\frac{\sqrt{1+bx}}{bx}\right]\\
&=\lim_{x\to\infty}\left[\frac{\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{1+ax}}{\frac{\mathrm{d}}{\mathrm{d}x}bx}\right]+\lim_{x\to\infty}\left[\frac{\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{1+bx}}{\frac{\mathrm{d}}{\mathrm{d}x}bx}\right]\\
&=\lim_{x\to\infty}\left[\frac{a}{2b\sqrt{1+ax}}\right]+\lim_{x\to\infty}\left[\frac{1}{2\sqrt{1+bx}}\right]\\
&=0. \qquad (46)
\end{aligned}$$

The only term that remains to be evaluated is the fourth term of (45), namely the term $\lim_{x\to\infty}[\frac{\sqrt{(1+ax)(1+bx)}}{bx}]$. Some steps of the evaluation procedure are shown below.

$$\begin{aligned}
&\lim_{x\to\infty}\left[\frac{\sqrt{(1+ax)(1+bx)}}{bx}\right]\\
&=\lim_{x\to\infty}\left[\frac{\sqrt{x^{-2}+(a+b)x^{-1}+ab}}{b}\right]\\
&=\lim_{x\to\infty}\frac{\sqrt{ab}}{b}=\sqrt{\frac{a}{b}}. \qquad (47)
\end{aligned}$$

Consequently, according to our definitions, $x=\gamma$, $a=|h_m|^2$, and $b=|h_n|^2$, the above analysis yields the solution to the original limit expression given in (42) as

$$\lim_{\gamma\to\infty}P_{\min}^{\{m,n\}}=\frac{|h_m|}{|h_n|}. \qquad (48)$$

This completes the proof of Theorem 4. ■

REFERENCES

[1] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[2] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[3] 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.

[4] 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.

[5] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[6] S. Euler, X. Lin, E. Tejedor, and E. Obregon, "High-altitude platform stations as international mobile telecommunications base stations: A primer on HIBS," *IEEE Veh. Technol. Mag.*, vol. 17, no. 4, pp. 92–100, Dec. 2022.

[7] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[8] 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.

[9] 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.

[10] Y. Yin, T. Ohtsuki, G. Gui, C. Yuen, H.-C. Wu, and H. Sari, "Joint optimization of user pairing, power allocation and content server deployment in NOMA-assisted wireless caching networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16866–16870, Dec. 2023.

[11] A. Masaracchia, D. B. Da Costa, T. Q. Duong, M.-N. Nguyen, and M. T. Nguyen, "A PSO-based approach for user-pairing schemes in NOMA systems: Theory and applications," *IEEE Access*, vol. 7, pp. 90550–90564, 2019.

[12] F. Jiang, Z. Gu, C. Sun, and R. Ma, "Dynamic user pairing and power allocation for NOMA with deep reinforcement learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2021, pp. 1–6.

[13] B. Narottama, D. K. Hendraningrat, and S. Y. Shin, "Quantum-inspired evolutionary algorithms for NOMA user pairing," *ICT Exp.*, vol. 8, no. 1, pp. 11–17, Mar. 2022.

[14] Z. Duan, N. Okada, A. Li, M. Naruse, N. Chauvet, and M. Hasegawa, "High-speed optimization of user pairing in NOMA system using laser chaos based MAB algorithm," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2021, pp. 073–077.

[15] B. K. S. Lima, R. Dinis, D. B. da Costa, R. Oliveira, and M. Beko, "User pairing and power allocation for UAV-NOMA systems based on multi-armed bandit framework," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13017–13029, Dec. 2022.

[16] A. Köse, M. Koca, E. Anarim, M. Médard, and H. Gökcesu, "Graph-theoretical dynamic user pairing for downlink NOMA systems," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3234–3238, Oct. 2021.

[17] X.-T. Dang, M. T. P. Le, H. V. Nguyen, S. Chatzinotas, and O.-S. Shin, "Optimal user pairing approach for NOMA-based cell-free massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4751–4765, Apr. 2023.

[18] L. Lei, D. Yuan, and P. Värbrand, "On power minimization for non-orthogonal multiple access (NOMA)," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2458–2461, Dec. 2016.

[19] K. N. Doan, M. Vaezi, W. Shin, H. V. Poor, H. Shin, and T. Q. S. Quek, "Power allocation in cache-aided NOMA systems: Optimization and deep reinforcement learning approaches," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 630–644, Jan. 2020.

[20] M. F. Sohail, C. Y. Leow, and S. Won, "Energy-efficient non-orthogonal multiple access for UAV communication system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10834–10845, Nov. 2019.

[21] M. F. Sohail, C. Y. Leow, and S. Won, "A cat swarm optimization based transmission power minimization for an aerial NOMA communication system," *Veh. Commun.*, vol. 33, Jan. 2022, Art. no. 100426.

[22] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems*, 2nd ed. Chichester, U.K.: Wiley, 1999.

[23] M. F. Sohail, C. Y. Leow, and S. Won, "Non-orthogonal multiple access for unmanned aerial vehicle assisted communication," *IEEE Access*, vol. 6, pp. 22716–22727, 2018.

[24] J. Zou, C. W. Sung, and K. W. Shum, "High-dimensional superposition NOMA and its user pairing strategy," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3800–3814, Jun. 2021.

[25] A. Burchard. "A short course on rearrangement inequalities," 2009. [Online]. Available: https://www.math.utoronto.ca/almut/rearrange.pdf

[26] J. Zou, K. W. Shum, and C. W. Sung, "Spherical code superposition NOMA and its user pairing strategy," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[27] A. Shahini and N. Ansari, "NOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7183–7191, Aug. 2019.

[28] F. Palacio et al., "Radio environmental maps: Information models and reference model," Centre de Comunicacions Avançades de Banda Ampla, Barcelona, Spain, document D4.1," Apr. 2011.

[29] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. M. Leung, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 48–53, Apr. 2018.

[30] S. Rezvani, E. A. Jorswieck, R. Joda, and H. Yanikomeroglu, "Optimal power allocation in downlink multicarrier NOMA systems: Theory and fast algorithms," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1162–1189, Apr. 2022.

[31] A. E. Taylor, "L'hospital's rule," *Am. Math. Month.*, vol. 59, no. 1, pp. 20–24, 1952.