



# Article Efficient Precoding and Power Allocation Techniques for Maximizing Spectral Efficiency in Beamspace MIMO-NOMA Systems

Yongfei Liu<sup>1</sup>, Lu Si<sup>1,2</sup>, Yuhuan Wang<sup>1,2</sup>, Bo Zhang<sup>3</sup> and Weizhang Xu<sup>1,2,\*</sup>

- <sup>1</sup> Engineering Research Center of Digital Audio & Video Ministry of Education, Communication University of China, Beijing 100024, China
- <sup>2</sup> State Key Laboratory of Media Convergence & Communication, Communication University of China, Beijing 100024, China
- <sup>3</sup> Datang Mobile Communications Equipment Co., Ltd., Beijing 100083, China
- Korrespondence: wzhxu@cuc.edu.cn

Abstract: Beamspace MIMO-NOMA is an effective way to improve spectral efficiency. This paper focuses on a downlink non-orthogonal multiple access (NOMA) transmission scheme for a beamspace multiple-input multiple-output (MIMO) system. To increase the sum rate, we jointly optimize precoding and power allocation, which presents a non-convex problem. To solve this difficulty, we employ an alternating algorithm to optimize the precoding and power allocation. Regarding the precoding subproblem, we demonstrate that the original optimization problem can be transformed into an unconstrained optimization problem. Drawing inspiration from fraction programming (FP), we reconstruct the problem and derive a closed-form expression of the optimization variable. In addition, we effectively reduce the complexity of precoding by utilizing Neumann series expansion (NSE). For the power allocation subproblem, we adopt a dynamic power allocation scheme that considers both the intra-beam power optimization and the inter-beam power optimization. Simulation results show that the energy efficiency of the proposed beamspace MIMO-NOMA is significantly better than other conventional schemes.

Keywords: beamspace MIMO-NOMA; sum rate; precoding; power allocation; FP

# 1. Introduction

With the coverage of mobile connections expanded, wireless communications systems are facing an escalating demand for data traffic, which poses challenges in terms of spectral efficiency and energy efficiency. Non-orthogonal multiple access (NOMA) technology has emerged as a key solution for improving spectral efficiency and supporting massive links, as it enables multiple users to share the same spectrum resource simultaneously. The application of NOMA in conventional terrestrial communication systems, benefiting from its superior spectral efficiency capability and capacity to accommodate massive connectivity, has been thoroughly investigated in many aspects [1]. Additionally, beamspace multiple-input multiple-output (beamspace MIMO) as another key technology also has several advantages. It leverages the abundant spectrum resources in the millimeter wave band, enabling terminal equipment to achieve high-rate data transmission. Furthermore, by employing large-scale MIMO, beamspace MIMO forms directional beams with high gain, effectively mitigating the challenge of substantial signal transmission path loss inherent in millimeter wave communication. Consequently, beamspace MIMO is recognized as a promising technology for future wireless communications [2]. Does this mean we can combine the NOMA and beamspace MIMO technologies to effectively leverage their advantages in the power and spatial domains, leading to improved spectral efficiency? The answer is affirmative. Specifically, considering the characteristics of these two technologies,



Citation: Liu, Y.; Si, L.; Wang, Y.; Zhang, B.; Xu, W. Efficient Precoding and Power Allocation Techniques for Maximizing Spectral Efficiency in Beamspace MIMO-NOMA Systems. *Sensors* 2023, 23, 7996. https:// doi.org/10.3390/s23187996

Academic Editor: Kit Yan Chan

Received: 19 August 2023 Revised: 9 September 2023 Accepted: 18 September 2023 Published: 20 September 2023



**Copyright:** © 2023 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/). beamspace MIMO requires a large number of radio frequency (RF) chains, which leads to high energy consumption and renders the all-digital structure unsuitable for direct application [3]. Moreover, in beamspace MIMO, the number of supported users cannot exceed the number of RF chains, thereby limiting the system's capacity to accommodate users. However, NOMA excels in increasing the number of system access users. Consequently, the integration of NOMA with mmWave massive MIMO, known as beamspace MIMO-NOMA, has emerged as a promising solution for significantly increasing the number of connections and further enhancing spectral efficiency. This approach has garnered growing research interest [4].

#### 1.1. Prior Works

Typically, the optimization of precoding and power allocation designs is considered a means to improve the performance of beamspace MIMO-NOMA systems. These problems have been investigated jointly or partially. However, the presence of inter-beam and intra-beam interference makes these problems non-convex and challenging to solve [5]. Fortunately, researchers have developed efficient algorithms to tackle these challenges.

Some works have focused on separately designing precoding or power allocation to improve performance. In [6], the authors propose a ZF precoding scheme to mitigate interference between users and employ the Karush-Kuhn-Tucker (KKT) conditions to investigate the power allocation problem for maximizing the sum rate. Furthermore, [7] explores energy efficiency maximization through power allocation and presents a two-layer iterative algorithm to tackle the non-convex optimization problem. The outer layer converts the original fractional objective function by using the Dinkelbach method, while the inner layer utilizes alternating optimization to solve the transformed problem. In [8], the authors introduce a low-complexity iterative algorithm called mean square error-based dynamic power allocation algorithm (MSE-DPA), which achieves near-perfect performance. Ref. [9] proposes a criterion based on correlation for user pairing to reduce inter-user interference, with ZF precoding applied to the paired users. The results demonstrate that the proposed scheme achieves higher spectral efficiency compared to the conventional scheme. In [10], the main objective is to design a low-complexity hybrid precoder (HP), where the authors propose a symmetric successive over-relaxation (SSOR) algorithm combined with complex regularized zero-forcing (CRZF) linear precoding.

In addition, a significant focus of the work is to design a joint optimization scheme for precoding and power allocation to enhance system performance. In [11], the authors adopt a ZF-based precoding scheme to mitigate inter-beam interference and propose a dynamic power allocation method based on minimum mean square error (MMSE) to maximize the achievable sum rate in beamspace MIMO-NOMA systems. Ref. [12] addresses the limitations of complicated successive interference cancellation (SIC) that were disregarded in [11]. Based on the ZF beamforming technique, the power allocation optimization problem is represented as a fractional programming (FP) problem, which was transformed into a convex optimization problem using sequential convex approximation (SCA) and secondorder cone (SOC) transformation. In [13], the authors formulate a joint hybrid beamforming and power allocation problem to maximize the sum rate. They employ the approximate ZF method to design the digital beamforming for minimizing inter-group interference and solve the analog beamforming problem with a constant-modulus constraint using a proposed boundary-compressed particle swarm optimization algorithm. In [14], the authors design ZF precoding matrices and evaluate power allocation coefficients based on optimal spectral efficiency to mitigate intra-beam interference. Additionally, they derive a tight closed-form formula for optimal spectral efficiency using KKT analysis. In [15], from the perspective of spectral efficiency, the authors propose a joint optimization framework and employ the quadratic transformation (QT) method to convert the non-convex power allocation problem into a convex problem. They also design an iterative approach to obtain optimal power allocation and digital beamforming. In [16], the authors propose a hybrid precoder that combines user channel alignment and the ZF algorithm to enhance the SINR. Furthermore, they address the non-convex optimization problem by transforming it into a convex optimization problem for inter-cluster power allocation, which can be solved by using the KKT conditions.

## 1.2. Motivations and Contributions

While the aforementioned research contributions have established a strong foundation for beamspace MIMO-NOMA, further investigation and improvements are still necessary to address practical considerations. Firstly, there is scope for enhancing the optimization of key performance indicators that impact spectral efficiency through various methodologies. Secondly, there is a need for research to focus on reducing computational complexity while improving spectral efficiency simultaneously, which remains an open area of exploration. These observations have inspired our primary research objectives in this study. In this work, our main goal is to maximize the sum rate of beamspace MIMO-NOMA in downlink communications and propose an optimal design scheme for joint precoding and power allocation, building upon the previous research. Against this backdrop, we emphasize the following four aspects that constitute the contributions of our paper:

- Firstly, we employ block optimization to optimize the joint problem of precoding and power allocation in beamspace MIMO-NOMA systems. In the precoding optimization part, we demonstrated that the original constrained problem can be transformed into an unconstrained problem. Moreover, we elucidated the quantitative relationship between the solutions of the original problem and the equivalent unconstrained problem. For the power allocation part, we adopted a dynamic power allocation method based on a joint power optimization problem, taking into account power optimization within and between beams.
- Secondly, we devised a precoding scheme based on FP to decouple the optimization variables, effectively transforming the unconstrained problem into three equivalent subproblems. Subsequently, we derived closed expressions for the optimization variables.
- Thirdly, as the number of antennas at the BS and the number of users accessing the system increase, the hardware and signal processing complexity also escalates. Since the precoding optimization algorithm involves complex matrix inversion operations, its calculation complexity is  $\mathcal{O}(N_{RF}^3)$ , which grows cubically with the increase in the number of RF connections. To mitigate this complexity, we utilized the Neumann series expansion (NSE) method to approximate the inverse of the precise matrix and expand the lower-order terms, thereby reducing the complexity of the matrix inversion operation to  $\mathcal{O}(N_{RF}^2)$ .
- Finally, we validated the performance of the proposed scheme through simulation. The results demonstrated that the algorithm significantly improves spectral efficiency. Furthermore, the simulation results confirmed that the proposed precoding and power allocation scheme outperforms the benchmark methods.

## 1.3. Organization and Notations

The remainder of the paper is organized as follows. Section 2 outlines the system model of beamspace MIMO-NOMA. Based on this model, Section 3 formulates the maximum sum rate problem, and an introduction to the proposed algorithm is provided. Section 4 presents the simulation results to evaluate the performance. Finally, Section 5 concludes the paper.

Notation: Denote  $\mathbb{C}$  as the set of complex numbers, and  $Re\{\bullet\}$  as the real part. We use the superscript to denote the Hermitian transpose of a matrix and overline the complex conjugate. The bold lower-case letter denotes a vector; the bold upper-case letter denotes a matrix; the calligraphic upper-case letter denotes a set.  $I_n$  denotes the identity matrix of dimension  $n.(\bullet)^T, (\bullet)^H, (\bullet)^{-1}$  and  $\|\cdot\|_F$  denote transpose, Hermitian transpose, inversion, and Frobenius norm operations, respectively.

## 2. System Model and Problem Formulation

In this section, we first review the beamspace MIMO system model, followed by a detailed description of the beamspace MIMO-NOMA system model.

## 2.1. System Model of Beamspace MIMO

As illustrated in Figure 1, the system depicted represents a single-cell downlink mmWave MIMO communication system. The BS is equipped with *N* antennas and  $N_{RF}$  RF chains, serving *K* randomly distributed single-antenna users simultaneously [17]. Employing the usual uniform linear array (ULA) structure, utilizing a well-designed lens antenna array at the BS. The received signal vector  $y = [y_1, y_2, \dots, y_K]^T$  is represented as:

$$y = H^H W P s + n, \tag{1}$$

where  $s = [s_1, s_2, \dots, s_K]^T \in C^{K \times 1}$  represents the transmitted signal vector for all *K* users satisfied with  $E(ss^H) = I_K$ ,  $P = diag(p_1, p_2, \dots, p_K)$  is the diagonal power allocation matrix,  $W = [w_1, w_2, \dots, w_K] \in C^{N \times K}$  is the precoding matrix, and  $H = [h_1, h_2, \dots, h_K] \in C^{N \times K}$  is the Rayleigh fading channel matrix, where  $h_k \in C^{N \times 1}$  denotes the channel vector between the BS and the *k*th user. In addition, *n* is the noise vector that follows the distribution  $CN(0, \sigma^2 I_K)$ . We consider the widely used Saleh–Valenzuela channel model for mmWave communications,  $h_k$  can be represented as

$$h_{k} = \beta_{k,0}a(\theta_{k,0}) + \sum_{l=1}^{L} \beta_{k,l}a(\theta_{k,l}),$$
(2)

where  $\beta_{k,0}$  denotes the LoS complex path gain, while  $a(\theta_{k,0})$  represents the array steering vector for the LoS path, similarly,  $\beta_{k,l}$  and  $a(\theta_{k,l})$  denote the complex gain and steering vector for the *l*th NLoS path, respectively. Furthermore, *L* denotes the number of NLoS paths.



Figure 1. The system model of beamspace MIMO architecture.

For the typical ULA, the expression of  $a(\varphi)$  is can be expressed as follows [18]:

$$a(\theta) = \frac{1}{\sqrt{N}} \left[ e^{-j2\pi\theta m} \right]_{m \in J(N)'}$$
(3)

where  $J(N) = \{i - (N-1)/2, i = 0, 1, \dots, N-1\}$  is a symmetric set of indices centered around zero. The spatial direction of the channel is defined as  $\theta = \frac{d}{\lambda} \sin(\varphi)$ ,  $\lambda$  represents the wavelength,  $d = \frac{\lambda}{2}$  denotes the antenna spacing, and  $\varphi$  denotes the physical direction of the corresponding path satisfying  $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$ . The lens antenna array serves as a discrete Fourier transformation matrix U, defined as

$$U = \left[a(\overline{\theta}_1), a(\overline{\theta}_2), \cdots, a(\overline{\theta}_N)\right]^H,\tag{4}$$

where  $\overline{\theta}_n = \frac{1}{N} \left( n - \frac{N+1}{2} \right)$  for  $n = 1, 2, \dots, N$  are the predefined spatial directions. Then, the received signal vector  $\overline{y}$  in the beamspace MIMO systems is given by

$$\overline{y} = H^H U^H W P s + n = \overline{H}^H W P s + n, \tag{5}$$

where  $\overline{H} = UH$  is the beamspace channel matrix. We employ the classic maximummagnitude-based beam selection method to choose a subset of the *N* orthogonal beams to serve all *K* users without obvious performance loss [19]. Consequently, the number of RF chains is reduced from *N* to  $N_{RF}$ . Thus, the received signal can be written as

$$\overline{y} = \overline{H}_r^H W_r P s + n, \tag{6}$$

where  $\overline{H}_r = \overline{H}(m,:), m \in \mathcal{M}$  is the dimension-reduced beamspace channel matrix with size  $|\mathcal{M}| \times K$ , and  $\mathcal{M}$  is the index set of selected beams. It is important to note that in this system, one RF chain generates one beam, resulting in the number of selected beams  $|\mathcal{M}|$  being equal to the number of RF chains  $N_{RF}$  [11]. In addition, the dimension-reduced digital precoding matrix  $W_r$ , has a size of  $|\mathcal{M}| \times K$ . Since  $W_r$  has a smaller row dimension compared to the original precoding matrix W, the number of required RF chains can be significantly reduced [20].

Notwithstanding, reducing the number of RF chains also presents a challenge of limited connections. To overcome this fundamental limit, a novel transmission scheme known as beamspace MIMO-NOMA, which combines the concept of NOMA with beamspace MIMO, has been proposed. By incorporating NOMA into beamspace MIMO systems, both spectral efficiency and connection density can be further enhanced [6].

#### 2.2. System Model of Beamspace MIMO-NOMA

As shown in Figure 2, this is a typical beamspace MIMO-NOMA wireless communication system. We consider that there are  $N_{RF}$  groups assigned to provide service, and we denote the set of users  $S_m$  served by the *m*th beam with  $S_m \cap S_n = \emptyset$  for  $m \neq n$  and  $\sum_{m=1}^{N_{RF}} |S_m| = K$ . The received signal  $\hat{y}_{m,n}$  of the *n*th user in the *m*th beam can be expressed as follows:

$$\hat{y}_{m,n} = \underbrace{h_{m,n}^H w_m \sqrt{p_{m,n}} s_{m,n}}_{desired \ signal} + \underbrace{\left(\sum_{k=1}^{n-1} h_{m,n}^H w_m \sqrt{p_{m,k}} s_{m,k} + \sum_{k=n+1}^{|S_m|} h_{m,n}^H w_m \sqrt{p_{m,k}} s_{m,k}\right)}_{intra-beam \ interferences} + \underbrace{\sum_{k=n+1}^{N_{m,n}} h_{m,n}^H w_m \sqrt{p_{m,k}} s_{m,k}}_{inter-beam \ interferences} + \underbrace{\sum_{k=n+$$

where  $s_{m,n}$  is the transmitted signal for the *n*th user in the *m*th beam with normalized power, and  $p_{m,n}$  is the corresponding transmitted power,  $w_m = W_r(:,m)$  represents the *m*th beam digital precoding vector, and  $v_{m,n} \sim C\mathcal{N}(0,\sigma^2)$  refers to the noise. Based on the principle of NOMA, intra-beam interference can be mitigated by utilizing SIC. Supposing that  $||h_{m,1}||^2 \ge ||h_{m,2}||^2 \ge \cdots \ge ||h_{m,|S_m|}||^2$  for  $m = 1, 2, \cdots, N_{RF}$ , within the same beam, the *i*th user can sequentially detect the *j*th user (for all j > i) and remove the detected signals from its received signals [21]. In the *m*th beam, after employing SIC to decode the *n*th user's signal, the remaining received signal can be expressed as follows:

$$y_{m,n} = \underbrace{h_{m,n}^H w_m \sqrt{p_{m,n}} s_{m,n}}_{desired \ signal} + \underbrace{\sum_{k=1}^{n-1} h_{m,n}^H w_m \sqrt{p_{m,k}} s_{m,k}}_{intra-beam \ interferences} + \underbrace{\sum_{l\neq m}^{N_{RF}} \sum_{k=1}^{|S_l|} h_{m,n}^H w_l \sqrt{p_{l,k}} s_{l,k}}_{inter-beam \ interferences} + \underbrace{v_{m,n}}_{noise}.$$
(8)



Figure 2. The system model of beamspace MIMO-NOMA architecture.

Therefore, the signal-to-interference-plus-noise ratio (SINR) at the *n*th user in the *m*th beam can be expressed as follows:

$$\gamma_{m,n} = \frac{\|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\xi_{m,n}},$$
(9)

where

$$\xi_{m,n} = \sum_{k=1}^{n-1} \left\| h_{m,n}^H w_m \right\|_2^2 p_{m,k} + \sum_{l \neq m}^{N_{RF}} \sum_{k=1}^{|S_l|} \left\| h_{m,n}^H w_l \right\|_2^2 p_{l,k} + \sigma^2.$$
(10)

Hence, the corresponding achievable rate can be expressed as follows:

$$R_{m,n} = \log_2(1 + \gamma_{m,n}).$$
(11)

Consequently, the overall achievable sum rate of the beamspace MIMO-NOMA scheme is:

$$R_{sum} = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} R_{m,n},$$
(12)

Indeed, precoding optimization helps mitigate inter-beam interference, but intra-beam interference endures within beamspace MIMO-NOMA systems. Power allocation effectively mitigates this inter-beam interference, thus enhancing overall system performance. It is noteworthy that expressions (9)–(11) illustrate the substantial influence of power allocation parameters  $\{p_{m,n}\}$  and precoding vectors  $\{w_m\}$  on maximizing the sum rate. Thus, the system performance can be further enhanced through the meticulous design of precoding and power allocation strategies. Jointly optimizing precoding and power allocation is pivotal for maximizing overall system performance. While this may add complexity, thoughtful design, and analysis allow for performance improvements without imposing significantly higher computational demands. In the following section, we will explore these ideas in greater detail.

## 3. Alternating Optimization of Beam-Specific Digital Precoding and Power Allocation

In this section, we begin by formulating the optimization problem. Next, we present an alternating optimization method to obtain the solution for beam-specific digital precoding. Finally, we maximize the achievable sum rate by solving the joint power optimization problem using a dynamic power allocation scheme.

## 7 of 22

#### 3.1. Problem Formulation

Our objective is to maximize the achievable sum rate problem by jointly optimizing the beam-specific digital precoding and power allocation, while adhering to the maximum transmit power constraint of the BS. The optimization problem can be formulated as follows:

$$P_{1}: \max_{\{p_{m,n}\}\{w_{m}\}} K_{sum}$$
  
s.t. $C_{1}: p_{m,n} \ge 0, \forall m, n,$   
$$C_{2}: \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_{m}|} p_{m,n} ||w_{m}||^{2} \le P_{max}.$$
 (13)

Obviously, three problems need to be addressed to optimize  $\mathcal{P}_1$ . As shown in (9), the presence of both intra-beam interference and inter-beam interference in the system results in the optimization variable  $\{p_{m,n}\}$  and  $\{w_m\}$  appears in both the nomination and denominator of  $\gamma_{m,n}$ . Consequently, the problem becomes a non-convex optimization problem that is difficult to solve directly. Furthermore, it is highly nonlinear. Additionally, the optimization of precoding  $\{w_m\}$  is performed at the beam level, while the optimization of power allocation  $\{p_{m,n}\}$  is carried out at the user level. This implies that both aspects are difficult to optimize simultaneously.

To tackle the complexity of the original problem  $\mathcal{P}_1$ , we decompose it into two subproblems:  $\widetilde{\mathcal{P}}_{beam}$  and  $\mathcal{P}_{power}$  for optimization. For the sub-problem  $\widetilde{\mathcal{P}}_{beam}$ , we first convert the constrained optimization problem into an unconstrained optimization problem. Then, we employ the FP algorithm to handle the NP-hard problem, leading to the derivation of a closed expression for precoding W. Additionally, we leverage the NSE to reduce the complexity of the precoding process. As for the sub-problem of power allocation, we utilize a dynamic power allocation scheme to obtain a closed-form expression for the power distribution, ensuring lower complexity.

#### 3.2. The Proposed Beam-Specific Digital Precoding Optimization

In this subsection, we focus on optimizing the beam-specific digital precoding vectors  $\{w_m\}$  for a given set of power allocation parameters  $\{p_{m,n}\}$ . To accomplish this, we transform the non-convex precoding optimization problem into an unconstrained optimization problem. To be specific, the precoding problem can be formulated as follows:

$$\mathcal{P}_{beam} : \max_{\{w_m\}} R_{sum}$$

$$s.t.C_1 : \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 \le P_{\max}.$$
(14)

Specifically, inspired by [22], we establish the following definition and proposition.

**Definition 1.** (*Trivial Stationary Point*): If a point X satisfying HX = 0, which results in a zero-sum rate, we say that it is a trivial stationary point of the original problem  $\mathcal{P}_1$ .

**Proposition 1.** Any nontrivial stationary point of problem  $\mathcal{P}_{beam}$  must satisfy the constraint  $C_1$  with equality.

**Proof.** See Appendix A.  $\Box$ 

According to Proposition 1, it can be inferred that:

$$\frac{\sum_{x=1}^{N_{RF}}\sum_{y=1}^{|S_x|} p_{x,y} \|w_x\|^2}{P_{\max}} = 1.$$
(15)

Hence, the problem  $\tilde{\mathcal{P}}_{beam}$  can be transformed into the following unconstrained form:

$$\overline{\mathcal{P}}_{beam} : \max_{\{w_m\}} \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \log_2 \left( 1 + \frac{\|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\sum_{k=1}^{n-1} \|h_{m,n}^H w_m\|_2^2 p_{m,k} + \sum_{l \neq mk=1}^{N_{RF}} \sum_{k=1}^{|S_l|} \|h_{m,n}^H w_l\|_2^2 p_{l,k} + \frac{\sum_{x=1}^{N_{RF}} \sum_{j=1}^{|S_x|} p_{x,y} \|w_x\|^2}{p_{max}} \sigma^2 \right).$$
(16)

We can express the objective function as follows:

$$f_{beam}(w) = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \log_2 \left( 1 + \frac{\|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\tilde{\xi}_{m,n}} \right), \tag{17}$$

where

$$\widetilde{\xi}_{m,n} = \sum_{k=1}^{n-1} \left\| h_{m,n}^H w_m \right\|_2^2 p_{m,k} + \sum_{l \neq m}^{N_{RF}} \sum_{k=1}^{|S_l|} \left\| h_{m,n}^H w_l \right\|_2^2 p_{l,k} + \frac{\sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} p_{x,y} \|w_x\|^2}{P_{\max}} \sigma^2.$$
(18)

The following proposition establishes the relationship between  $\widetilde{\mathcal{P}}_{beam}$  and  $\overline{\mathcal{P}}_{beam}$ .

**Proposition 2.** The following relationship exists between the optimal solution  $\tilde{w}^{\circ}$  of the problem  $\tilde{\mathcal{P}}_{beam}$  and the optimal solution  $\overline{w}^{\circ}$  of the new unconstrained optimization problem  $\overline{\mathcal{P}}_{beam}$ .

$$\widetilde{w}_m^o = \sqrt{\frac{P_{max}}{\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|\overline{w}_m^o\|^2}} \overline{w}_m^o.$$
(19)

**Proof.** See Appendix **B**.□

This implies that if we find solution  $\overline{w}^{o}$ , then solution  $\widetilde{w}^{o}$  can be obtained according to Proposition 2.

Obviously, the objective function  $f_{beam}(w)$  remains non-convex, making it difficult to solve in polynomial time. To address this, we employ the Lagrangian dual transform to reframe the unconstrained problem  $\overline{\mathcal{P}}_{beam}$ , as demonstrated below [23].

$$\overline{\mathcal{P}}_{beam} : \min_{\{w_m\}\{u_{m,n}\}} f_{beam}^{LD}(\boldsymbol{w}, \boldsymbol{u}),$$

$$s.t.C_1 : u_{m,n} \in \mathbb{R}, \forall m,$$
(20)

where *u* refers to a set of auxiliary variables  $\{u_{m,n}\}$ , and the objective function of problem  $\overline{\mathcal{P}}_{beam}$  is formulated as follows:

$$f_{beam}^{LD}(\boldsymbol{w},\boldsymbol{u}) = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} - \left( \log_2(1+u_{m,n}) - u_{m,n} + \frac{(1+u_{m,n}) \|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\|h_{m,n}^H w_m\|_2^2 p_{m,n} + \tilde{\xi}_{m,n}} \right).$$
(21)

When  $w_m$  is held fixed, the optimal  $u_{m,n}$  can be obtained by solving  $\frac{\partial f_{beam}^{LD}(w,u)}{\partial u_{m,n}} = 0$ , i.e.,

$$\mu_{m,n}^{o} = \frac{\|h_{m,n}^{H}w_{m}\|_{2}^{2}p_{m,n}}{\tilde{\xi}_{m,n}}.$$
(22)

Now, we incorporate  $u_{m,n}^o$  into (21) and obtain

ı

$$f_{beam}^{LD}(\boldsymbol{w},\boldsymbol{u}) = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} - \left[ const(\boldsymbol{u}) + \frac{(1+u_{m,n}) \|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\|h_{m,n}^H w_m\|_2^2 p_{m,n} + \tilde{\xi}_{m,n}} \right],$$
(23)

where  $const(u) = \log_2(1 + u_{m,n}) - u_{m,n}$  is a constant term. Applying the multidimensional quadratic transform further transforms (23) and leads to the following expression:

$$f_{beam}^{MQ}(\boldsymbol{w},\boldsymbol{u},\boldsymbol{v}) = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \left[ -const(\boldsymbol{u}) - 2\sqrt{(1+u_{m,n})p_{m,n}} Re\left\{h_{m,n}^H w_m v_{m,n}^*\right\} + |v_{m,n}|^2 \left(\left\|h_{m,n}^H w_m\right\|_2^2 p_{m,n} + \tilde{\xi}_{m,n}\right)\right].$$
(24)

where v is the collection  $\{v_{m,n}\}$ . With  $u_{m,n}$  fixed, the optimal  $v_{m,n}$  can also be determined by setting  $\frac{\partial f_{bcam}^{MQ}(w,u,v)}{\partial v_{m,n}} = 0$ , and the optimal value  $v_{m,n}^{0}$  can be expressed as follows:

$$v_{m,n}^{o} = \frac{\sqrt{(1+u_{m,n})}p_{m,n}h_{m,n}^{H}w_{m}}{\left\|h_{m,n}^{H}w_{m}\right\|_{2}^{2}p_{m,n} + \tilde{\xi}_{m,n}}.$$
(25)

Likewise, with the other variables fixed, the optimal  $w_m$  satisfies the expression  $\frac{\partial f_{beam}^{MQ}(w,u,v)}{\partial w_m} = 0$ , i.e.,

$$w_{m}^{o} = \left(\sum_{n=1}^{|S_{m}|} |v_{m,n}|^{2} h_{m,n} h_{m,n}^{H} \sum_{k=1}^{n} p_{m,k} + \sum_{l \neq mn=1}^{N_{RF}} \sum_{k=1}^{|S_{l}|} ||v_{l,n}||_{2}^{2} \sum_{k=1}^{|S_{m}|} h_{l,n} h_{l,n}^{H} p_{m,k} + \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_{x}|} |v_{x,y}|^{2} \frac{\sum_{l=1}^{|S_{m}|} p_{m,l}}{P_{max}} \sigma^{2} I_{N_{RF}} \right)^{-1} \sum_{n=1}^{|S_{m}|} \sqrt{(1+u_{m,n})p_{m,n}} h_{m,n} v_{m,n}.$$
(26)

The proposed algorithm is summarized in Algorithm 1. Unfortunately, although  $N_{RF}$  is much smaller than K, the matrix inversion in the expression of  $w_m^o$  still remains high-dimensional, resulting in computational complexity of  $\mathcal{O}(N_{RF}^3)$  in each iteration, which may result in significant processing delays. To address this issue, the NSE has been explored as an alternative for approximating matrix inversion [24], we leverage the NSE to simplify the matrix inversion of  $w_m^o$  as follows.

Letting

$$A = \sum_{n=1}^{|S_m|} |v_{m,n}|^2 h_{m,n} h_{m,n}^H \sum_{k=1}^n p_{m,k} + \sum_{l \neq m}^{N_{RF}} \sum_{n=1}^{|S_l|} ||v_{l,n}||_2^2 \sum_{k=1}^{|S_m|} h_{l,n} h_{l,n}^H p_{m,k} + \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} |v_{x,y}|^2 \frac{\sum_{i=1}^{|S_m|} p_{m,i}}{P_{max}} \sigma^2 I_{N_{RF}},$$
(27)

we can observe that the matrix *A* exhibits diagonal dominance. In such cases, the inversion of *A* can be equivalently expressed as follows [25]:

$$A^{-1} = \sum_{n=0}^{\infty} \left( P^{-1}(P-A) \right)^n P^{-1}.$$
 (28)

By decomposing the matrix A as A = D + E, where D is a diagonal matrix consisting of the main diagonal elements of A, and E is a hollow matrix consisting of the remaining elements. Replace P in (28) with D and rewrite it as follows

$$A^{-1} = \sum_{n=0}^{\infty} \left( D^{-1} (D-A) \right)^n D^{-1} = \sum_{n=0}^{\infty} \left( -D^{-1}E \right)^n D^{-1}.$$
 (29)

Due to the high complexity of the full NSE algorithm, the truncated NSE, which aims to retain only the first k orders (k + 1 terms) of the Neumann series, is a more commonly used approach. The specific formula can be expressed as follows:

$$A^{-1} \approx A_k^{-1} = \sum_{n=0}^k \left( -D^{-1}E \right)^n D^{-1}.$$
 (30)

It should be noted that as the unfolding order increases (denoted as 'k > 1'), the computational complexity of the proposed NSE-based algorithm may exceed the complexity of  $\mathcal{O}(N_{RF}^3)$ . Therefore, to strike a balance between closely approximating the original precoding while reducing complexity, we choose k = 1, then  $A^{-1} \approx D^{-1} - D^{-1}ED^{-1}$ , we have the following expression [26]

$$w_m^o \approx \left( D^{-1} - D^{-1} E D^{-1} \right) \sum_{n=1}^{|S_m|} \sqrt{(1 + u_{m,n}) p_{m,n}} h_{m,n} v_{m,n}.$$
(31)

Based on this estimation, the NSE-level approximation algorithm can reduce the computational complexity from  $\mathcal{O}(N_{RF}^3)$  to  $\mathcal{O}(N_{RF}^2)$ . By combining the aforementioned updates, Algorithm 1 provides a detailed description of the proposed precoding optimization algorithm.

Algorithm 1 Proposed Precoding Framework.
Input:
Beamspace channel vectors: $h_{m,n}$ for $\forall m, n$ ;
Power allocation parameters: $p_{m,n}$ for $\forall m, n$ ;
Noise variance: $\sigma^2$ ;
Maximum iteration times: $T_{max}$ .
Output:
Optimal precoding vectors: $w_m^o$ for $\forall m$ ;
1. $t = 0.$
2. while $t < T_{max}$ do
3. Obtain the optimal $\{u_{m,n}^{(t)}\}$ according to (22);
4. Obtain the optimal $\left\{ v_{m,n}^{(t)} \right\}$ according to (25);
5. Obtain the optimal $\left\{w_m^{(t)}\right\}$ according to (31);
6. $t = t + 1$
7. end while
8. <b>return</b> $w_m^o = \sqrt{\alpha} w_m^{(t)}$ for $\forall m$ .

In Algorithm 1, it can be demonstrated that the computational complexity is primarily determined by line 5. Within each iteration, the complexity of obtaining the optimal values  $\{u_{m,n}^{(t)}\}$  in (22) and  $\{v_{m,n}^{(t)}\}$  in (25) is linear in the number of RF chains, i.e.,  $\mathcal{O}(N_{RF})$ .

Additionally, the complexity of finding the optimal value  $\left\{w_m^{(t)}\right\}$  in (31) is  $\mathcal{O}(N_{RF}^2)$ , due to the utilization of NSE. Consequently, the computational complexity is significantly lower than the complexity of  $\mathcal{O}(N_{RF}^3)$  stated in (26).

## 3.3. The Adopted Optimization Power Allocation

 $\pi$ 

The initial optimization problem  $\mathcal{P}_1$  can be transformed into the following problem when  $\{w_m\}$  is fixed.

$$\mathcal{P}_{power} : \max_{\{p_{m,n}\}} R_{sum}$$
  
s.t.C<sub>1</sub> :  $p_{m,n} \ge 0, \forall m, n,$   
C<sub>2</sub> :  $\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 \le P_{max}.$  (32)

Note that the problem remains challenging. To address this difficulty, we introduce Lemma 1 to simplify problem  $\mathcal{P}_{power}$ .

**Lemma 1.** Let 
$$f(a) = -\frac{ab}{\ln 2} + \log_2 a + \frac{1}{\ln 2}$$
 and  $a \in \mathbb{R}^{1 \times 1}$  be a positive scalar, we have

$$-\log_2 b = \max_{a>0} f(a), \tag{33}$$

where the optimal solution of a is  $a^o = \frac{1}{h}$ .

**Proof.** Since f(a) is a convex function, and the optimal solution of f(a) can be obtained by setting  $\frac{\partial f(a)}{\partial a} = 0$ , we can derive that

$$-\frac{b}{\ln 2} + \frac{1}{a\ln 2} = 0 \Rightarrow a = \frac{1}{b},\tag{34}$$

where the maximum value of f(a) is  $-\log_2 b$ .  $\Box$ 

Moreover, if we use the minimum mean square error (MMSE) to estimate  $s_{m,n}$ , then have the following expression:

$$e_{m,n} = E\Big\{|s_{m,n} - c_{m,n}y_{m,n}|^2\Big\},$$
(35)

where  $c_{m,n} \in \mathbb{C}^{1 \times 1}$  denotes the channel equalization coefficient,  $y_{m,n}$  is defined previously in (8). Substituting into (35), we obtain:

$$e_{m,n} = \left| 1 - c_{m,n} \sqrt{p_{m,n}} h_{m,n}^H w_m \right|^2 + |c_{m,n}|^2 \xi_{m,n}.$$
(36)

According to [11], the optimal equalization coefficient  $c_{m,n}$  can be obtained by the following formula:

$$c_{m,n} = \arg\min_{c_{m,n}} e_{m,n}, \tag{37}$$

and  $c_{m,n}$  can be calculated by  $\frac{\partial e_{m,n}}{\partial c_{m,n}} = 0$ , then we have

$$c_{m,n}^{0} = \left(\sqrt{p_{m,n}}h_{m,n}^{H}w_{m}\right)^{*} \left(\left\|h_{m,n}^{H}w_{m}\right\|_{2}^{2}p_{m,n} + \xi_{m,n}\right)^{-1}.$$
(38)

Substituting (38) into (36), we can obtain the optimal MMSE expression as follows:

$$e_{m,n}^{o} = 1 - p_{m,n} \left\| h_{m,n}^{H} w_{m} \right\|_{2}^{2} \left( \left\| h_{m,n}^{H} w_{m} \right\|_{2}^{2} p_{m,n} + \xi_{m,n} \right)^{-1}.$$
(39)

According to the extension of the Sherman-Morrison-Woodbury formula [27],

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(I + CDA^{-1}B)^{-1}CDA^{-1}.$$
 (40)

Thus,  $(1 + \gamma_{m,n})^{-1}$  can be reformulated as

$$(1+\gamma_{m,n})^{-1} = 1 - p_{m,n} \left\| h_{m,n}^H w_m \right\|_2^2 \left( \left\| h_{m,n}^H w_m \right\|_2^2 p_{m,n} + \xi_{m,n} \right)^{-1}.$$
(41)

We observe that the expression (41) has the same form as the MMSE expression (39). i.e., we have

$$(1 + \gamma_{m,n})^{-1} = \min_{c_{m,n}} e_{m,n} = (e_{m,n})^{o}$$
(42)

Using Lemma 1, we can equivalently rewrite  $\mathcal{P}_{power}$  as

$$\hat{\mathcal{P}}_{power} : \max_{\{p_{m,n}\}} \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \max_{\{a_{m,n}\}} \max_{\{a_{m,n}\}} \left( -\frac{a_{m,n}e_{m,n}}{\ln 2} + \log_2 a_{m,n} + \frac{1}{\ln 2} \right)$$

$$s.t.C_1 : p_{m,n} \ge 0, \forall m, n,$$

$$C_2 : \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 \le P_{max}.$$
(43)

where  $a_{m,n} > 0$  is an introduced slack variable. We propose to iteratively optimize  $\{p_{m,n}\}$ ,  $\{c_{m,n}\}$  and  $\{a_{m,n}\}$  by using the alternating optimization algorithm. The optimal solution can be obtained by:

$$a_{m,n}^{o} = \frac{1}{e_{m,n}^{o}}.$$
(44)

After obtaining the optimal values  $c_{m,n}^o$  and  $a_{m,n}^o$  in the iteration, the optimal value  $p_{m,n}^o$  can be obtained by solving the following problem:

$$\widetilde{\mathcal{P}}_{power} : \max_{\{p_{m,n}\}} \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} a_{m,n} e_{m,n}$$

$$s.t.C_1 : p_{m,n} \ge 0, \forall m, n,$$

$$C_2 : \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 \le P_{max}.$$
(45)

We observe that  $\tilde{\mathcal{P}}_{power}$  is a convex optimization problem, which can be solved by using the following Lagrange function:

$$L(p,\lambda) = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} a_{m,n} \left( \left| 1 - c_{m,n} \sqrt{p_{m,n}} h_{m,n}^H w_m \right|^2 + |c_{m,n}|^2 \xi_{m,n} \right) + \lambda \left( \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 - P_{max} \right)$$
(46)

where  $\lambda \ge 0$ . Then, the Karush–Kuhn–Tucker (KKT) condition of problem  $\mathcal{P}_{power}$  can be obtained as follows.

$$\frac{\partial L(p,\lambda)}{\partial p_{m,n}} = -\frac{a_{m,n}}{\sqrt{p_{m,n}}} Re\left\{c_{m,n}h_{m,n}^H w_m\right\} + \sum_{y=n}^{|S_m|} a_{m,y}|c_{m,y}|^2 \left\|h_{m,y}^H w_m\right\|_2^2 + \sum_{x\neq m}^{N_{RF}} \sum_{k=1}^{|S_x|} a_{x,k}|c_{x,k}|^2 \left\|h_{x,k}^H w_m\right\|_2^2 + \lambda \|w_m\|^2 = 0, \tag{47}$$

$$\lambda \left( \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 - P_{max} \right) = 0.$$
(48)

Finally, the optimal solution  $p_{m,n}^o$  from (45) can be found as follows:

$$p_{m,n}^{o} = \left(\frac{a_{m,n}Re\{c_{m,n}h_{m,n}^{H}w_{m}\}}{\sum_{y=n}^{|S_{m}|}a_{m,y}|c_{m,y}|^{2}\left\|h_{m,y}^{H}w_{m}\right\|_{2}^{2} + \sum_{x\neq m}^{N_{RF}}\sum_{k=1}^{|S_{x}|}a_{x,k}|c_{x,k}|^{2}\left\|h_{x,k}^{H}w_{m}\right\|_{2}^{2} + \lambda\|w_{m}\|^{2}}\right)^{2}.$$
(49)

We can see that the values of  $\{c_{m,n}^o\}$ ,  $\{a_{m,n}^o\}$  and  $\{p_{m,n}^o\}$  obtained in each iteration are closed optimal solutions because (37), (33) and (45) are all convex after a sequence of transformations. The iterative update of  $\{c_{m,n}^o\}$ ,  $\{a_{m,n}^o\}$  and  $\{p_{m,n}^o\}$  will only increase or maintain the objective function in (43). A monotonically growing sequence of objective function values in (43) can be obtained through iterative updating. However, it has an upper bound because of the transmission power restriction. Therefore, the proposed iterative optimization algorithm for power allocation will converge to a stationary solution of problem  $\hat{\mathcal{P}}_{power}$ . The power allocation optimization technique is described in detail in Algorithm 2. We summarize the proposed algorithm in Algorithm 3.

Algorithm 2 Proposed Power Allocation Framework.

Input: Beamspace channel vectors:  $h_{m,n}$  for  $\forall m, n$ ; Precoding vectors:  $w_m$  for  $\forall m$ ; Noise variance:  $\sigma^2$ ; Maximum iteration times:  $T_{max}$ . Output: Optimal power allocation vectors:  $p_{m,n}^o$  for  $\forall m, n$ ; 1. t = 0. 2. while  $t < T_{max}$  do Obtain the optimal  $\left\{c_{m,n}^{(t)}\right\}$  according to (38); 3. Obtain the optimal  $\left\{a_{m,n}^{(t)}\right\}$  according to (44); 4. Obtain the optimal  $\left\{ p_{m,n}^{(t)} \right\}$  according to (49); 5. 6. t = t + 17. end while **return**  $p_{m,n}^o = p_{m,n}^{(t)}$  for  $\forall m, n$ . 8.

Algorithm 3 Proposed Joint Precoding and Power Allocation Framework.

## Input:

Maximum iteration times:  $T_{max}$ . **Initialize:** Power allocation  $p_{m,n}^0$  for  $\forall m, n$  **Output:** Optimal power allocation vectors and precoding vectors:  $w_m^o$ ,  $p_{m,n}^o$  for  $\forall m, n$ ; 1. t = 0. 2. **while**  $t < T_{max}$  **do** 3. Find the optimal beamforming vectors  $w_m^{(t)}$  given  $p_{m,n}^{(t-1)}$  by Algorithm 1; 4. Find the optimal allocation vectors  $p_{m,n}^{(t)}$  given  $w_m^{(t)}$  by Algorithm 2; 5. t = t + 16. **end while** 

7. **return**  $w_m^o = \sqrt{\alpha} w_m^{(t)}$ ,  $p_{m,n}^o = p_{m,n}^{(t)}$  for  $\forall m, n$ .

The computational complexity of the proposed algorithm mainly arises from the iteration part. We observe that in each iteration, the complexity of obtaining the optimal values  $\{c_{m,n}^o\}$  in (38) and  $\{a_{m,n}^o\}$  in (44) is linear with the number of users, i.e.,  $\mathcal{O}(K)$ .  $\lambda$  in (48) can be obtained by using the Newton or bisection methods, both of which have a complexity of  $\mathcal{O}(K^2 \log_2 \delta)$ , where  $\delta$  represents the desired accuracy. The overall complexity of the suggested power allocation algorithm can be calculated to be

 $O(T_{max}K^2 \log_2 \delta)$ , where  $T_{max}$  is the maximum number of repetitions. Therefore, the complexity of the proposed joint precoding design and power allocation optimization algorithm is  $O(T_{max}K^2 \log_2 \delta + T_{max}N_{RF}^3)$ . While the computational complexity of the algorithm without NSE processing is  $O(T_{max}K^2 \log_2 \delta + T_{max}N_{RF}^4)$ .

#### 4. Simulation Result

The performance of the proposed joint optimization algorithm for the mmWave beamspace MIMO-NOMA scheme is evaluated by using numerical simulations in this section.

#### 4.1. Simulation Setup

In this paper, we consider a typical single-cell downlink mmWave massive MIMO system. The BS is equipped with a ULA of N = 256 transmit antennas that communicate with *K* users simultaneously. The system bandwidth is assumed to be 1 Hz, and the total transmit power is set to P = 32mW (15 dBm) [11]. For all users' channels, we assume L = 1 LoS component and L = 2 NLoS components, where  $\beta_k^{(0)} \sim C\mathcal{N}(0,1), \beta_k^{(l)} \sim C\mathcal{N}(0,10^{-1})$  for  $1 \le l \le L, \theta_k^{(0)}$  and  $\theta_k^{(l)}$  follows a uniform distribution within  $\left[-\frac{1}{2}, \frac{1}{2}\right]$  for  $1 \le l \le L$ . The SNR is set as  $\frac{P}{\sigma^2}$ , the maximum number of iterations  $T_{max} = 50$ .

We consider the following four typical mmWave massive MIMO solutions for comparison, and we aim to use the same system configuration in these systems to conduct a fair comparison: "traditional fully digital MIMO" (FDM), "traditional beamspace MIMO" (BM), "traditional MIMO-OMA"(MO), in particular, we compared our approach with the reference [11], which is a particularly classic and highly effective method based on a "beamspace MIMO-NOMA" (BMN) system, as a benchmark.

We evaluated the performance in terms of energy efficiency and spectral efficiency of each of the four baseline systems mentioned above. According to [20], energy efficiency can be expressed as:

$$\varepsilon_{EE} = \frac{R_{sum}}{P_t + N_{RF}P_{RF} + N_{RF}P_{SW} + P_{BB}}$$
(50)

where  $P_t$  represents the total transmit power,  $P_{RF}$  represents the power consumed by each RF,  $P_{SW}$  represents the power consumed by each switch, and  $P_{BB}$  represents the power consumed at the baseband. For the parameters, we have adopted the following common values:  $R_{RF}$ = 300mW,  $P_{SW}$ = 5mW and  $P_{BB}$ = 200mW.

#### 4.2. Simulation Results

The performance evaluation of the proposed MIMO-NOMA system was carried out in three different cases: performance comparison at different SNRs, performance comparison at different numbers of users, and performance comparison at different numbers of antennas.

## Comparison of performance with different SNRs

Figure 3 depicts the comparison of spectral efficiency versus SNRs with K = 32 and K = 128. As the SNR increases, both sets of curves demonstrate an increase in spectral efficiency. The proposed optimization structures namely proposed 'BMN' and proposed 'beamspace MIMO-NOMA with NSE' (BMNN), exhibited very similar results in terms of spectral efficiency growth. This indicates that our precoding scheme, approximated by the NSE, not only reduced the complexity of the original algorithm but also achieved comparable performance. These findings highlight the effectiveness of the NSE approximation algorithm.



Figure 3. Spectrum efficiency comparison versus SNRs of the two schemes with different users.

Figure 4 presents a comparison of spectral efficiency versus SNRs for the proposed system and the baseline systems. In particular, we compared the spectral efficiency of the proposed algorithm in the beamspace MIMO-NOMA system with the classical BMN [11] algorithm, both for 128 users and 32 users. The results indicate that in both scenarios, BMNN outperformed BMN [11], with the advantage becoming more pronounced as the number of users increased. When there were 32 users, the proposed BMNN scheme outperformed the BMN [11], BM, and MO schemes in terms of spectral efficiency. Particularly, compared to BMN [11], the performance gain of BMNN came mainly from the optimization of precoding for different beams in the first stage. Moreover, the proposed BMNN exhibited significantly better performance than BM, benefiting from the integration of beamspace MIMO and NOMA technologies, which enabled simultaneous service to multiple users within each beam and effectively improved spectral efficiency. Since NOMA can achieve higher spectral efficiency than OMA, it is evident that the proposed BMNN also outperforms MO in terms of spectral efficiency.



Figure 4. Spectrum efficiency comparison versus SNRs with different users.

Figure 5 illustrates the comparison of energy efficiency versus SNRs with K = 32 and K = 128 users for the proposed system and the baseline systems. It can be clearly seen

that increasing SNR can lead to a substantial growth in energy efficiency, and within the same system, for both 32 users and 128 users, our algorithm outperformed BMN [11]. Furthermore, in different systems with 32 users, the energy efficiency of the proposed BMNN was higher than that of the other four baseline systems. Specifically, compared to BM, our proposed BMNN achieved higher energy efficiency, by integrating NOMA and beamformed MIMO, allowing each beam to serve multiple users.





<sup>(2)</sup> Comparison of performance with different users

The aforementioned results were obtained while considering varying SNR, however, in real communication systems, especially in massive MIMO systems, the number of accessed users plays a significant role. Therefore, we further investigated the spectrum efficiency performance of the two proposed solutions under different user scenarios.

Figure 6 depicts how spectrum efficiency varies with the number of users. Both curves exhibit an upward trend with increasing user count, and the spectrum efficiency growth curves of the two proposed optimization structures yield similar results.



Figure 6. Spectrum efficiency comparison versus users of the two schemes at SNR = 0 dB.

Figure 7 illustrates a comparison of the spectrum efficiency of the four schemes under different user scenarios at 0 dB. The BMNN scheme outperformed the BMN [11], BM, and



MO schemes. Moreover, compared to the traditional BM schemes, the BMNN optimization scheme proposed in this study further improved spectrum efficiency.

**Figure 7.** Spectrum efficiency comparison versus users at SNR = 0 dB.

Figure 8 displays the energy efficiency performance for all considered schemes as the number of users increases. It is obvious that the proposed algorithm remained superior among the five schemes, which proves the effectiveness of the proposed scheme. Another noteworthy observation is that the performance of our proposed BMNN algorithm surpassed that of BMN [11] in terms of energy efficiency. This is mainly attributed to the fact that BMN [11] utilizes the ZF algorithm commonly employed in many studies in the precoding part, whereas our proposed algorithm optimizes the precoding parameters, thereby validating the necessity of optimizing precoding design parameters in our algorithm.



Figure 8. Energy efficiency comparison versus users at SNR = 0 dB.

Figure 9 shows how spectral efficiency varies with an increasing number of users at SNR levels of -5 dB, 0 dB, and 5 dB. It is important to note that, across all these SNR conditions, the BMNN algorithm we propose consistently outperformed the other schemes, and its superiority becomes even more pronounced as the SNR increases.



Figure 9. Spectral efficiency comparison versus users with different SNRs.

③ Comparison of performance with different antennas

From Figure 10, it can be observed that, BMNN exhibited a clear advantage over other algorithms until the number of antennas increases to 200. Beyond this point, the spectrum efficiency of the FDM algorithm surpassed that of the others. This is primarily attributed to the increase in the number of antennas in the FDM algorithm. With more antennas, precise beamforming becomes possible, allowing for more accurate signal focusing. This allows signals to be aimed more accurately at the receivers, reducing signal scattering and interference, ultimately leading to improved spectral efficiency.



**Figure 10.** Spectral efficiency comparison versus antennas number at SNR = 0 dB.

However, it is worth noting that the FDM algorithm typically requires more hardware and signal processing resources, which can lead to higher power consumption. As Figure 11 corroborates, the energy efficiency of the FDM tends to be lower. Nevertheless, as seen in the graph, our proposed BMNN algorithm achieved the highest energy efficiency among all algorithms, highlighting its potential to enhance system performance with a clear advantage.



**Figure 11.** Energy efficiency comparison versus antennas number at SNR = 0 dB.

## 5. Conclusions

In this research, we addressed the joint optimization problem of precoding and power allocation in massive MIMO-NOMA networks, aiming to maximize the sum rate for all devices. To tackle this challenge, we transformed the original optimization problem into an unconstrained problem for the precoding subproblem. We employed the FP approach to handle the non-convex problem, resulting in three equivalent problems and a closed expression for precoding. For the power allocation subproblem, which remains nonconvex, we utilized the MMSE-based dynamic power allocation scheme to solve it. Simulation results demonstrated that the proposed beamspace MIMO-NOMA system outperforms the baseline in terms of both spectrum and energy efficiency. In future work, we intend to extend the proposed optimization framework for precoding from beam-based optimization to user-based optimization, aiming to further improve system performance.

**Author Contributions:** Conceptualization and model building, Y.L., B.Z. and W.X.; algorithm and simulation, L.S. and B.Z.; writing—original draft preparation, Y.L. and Y.W.; writing—review and editing, L.S. and W.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Fundamental Research Funds for the Central Universities. (CUC23GZ003, in Communication University of China, China).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

## Appendix A

When  $\{p_{m,n}\}$  is held constant, the optimization problem  $\mathcal{P}_1$  can be transformed into the following form:

$$\mathcal{P}_{beam} : \max_{\{w_m\}} \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \log_2 \left( 1 + \frac{\|h_{m,n}^H w_m\|_2^2 p_{m,n}}{\sum_{k=1}^{n-1} \|h_{m,n}^H w_m\|_2^2 p_{m,k} + \sum_{l \neq m}^{N_{RF}} \sum_{k=1}^{|S_l|} \|h_{m,n}^H w_l\|_2^2 p_{l,k} + \sigma^2} \right)$$

$$s.t. : C_1 : \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|w_m\|^2 \le P_{max}.$$
(A1)

Let  $w_m = aw'_m$ ,  $\forall m \in \{1, 2, \dots, N_{RF}\}$ ,  $\alpha \in \mathbb{R}^+$ . Then,  $\|h_{m,n}^H w_m\|_2^2 = a^2 \|h_{m,n}^H w'_m\|_2^2$ , and the objective function of  $\mathcal{P}_{beam}$  can be expressed as the following expression:

$$R_{sum} = \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} \log_2 \left( 1 + \frac{\|h_{m,n}^H w_m'\|_2^2 p_{m,n}}{\sum_{k=1}^{n-1} \|h_{m,n}^H w_m'\|_2^2 p_{m,k} + \sum_{l \neq m}^{N_{RF}} \sum_{k=1}^{|S_l|} \|h_{m,n}^H w_l'\|_2^2 p_{l,k} + \frac{\sigma^2}{a^2}} \right)$$
(A2)

From (A2), it is evident that when  $\{p_{m,n}\}$  and  $\{w_m\}$  are held constant,  $R_{sum}$  exhibits a monotonically increasing behavior in relation to the variable  $\alpha$ . In other words, as  $R_{sum}$ increases, resulting in an increase in  $||w_m||^2$ . In the case where  $\{p_{m,n}\}$  is constant, as  $||w_m||^2$ increases, it will eventually reach the upper bound  $P_{max}$  set by constraint  $C_1$ . Therefore, to attain the maximum value of  $R_{sum}$  under the constraint  $C_1$ , the equation holds, denoted as  $\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} ||w_m||^2 = P_{max}$ .

#### Appendix **B**

Sufficiency:

Suppose  $\overline{w}_m^0$  is an arbitrary stationary point of unconstrained problem  $\overline{\mathcal{P}}_{beam}$ , and let  $\widetilde{w}_m^0$  be defined as  $\widetilde{w}_m^0 = \sqrt{a}\overline{w}_m^0, \forall m$ , where  $a = \frac{P_{max}}{\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} ||\overline{w}_m^0||^2}$ . The sufficiency condition aims to demonstrate that  $\widetilde{w}_m^0$  is a nontrivial stationary point of the original problem  $\widetilde{\mathcal{P}}_{beam}$ .

Let  $\overline{R}_{m,n}$  represent the achievable rate of the unconstrained optimization, which can be expressed as follows:

$$\overline{R}_{m,n} = \log_2 \left( \left\| h_{m,n}^H w_m \right\|_2^2 p_{m,n} + \widetilde{\xi}_{m,n} \right) - \log_2 \widetilde{\xi}_{m,n}.$$
(A3)

Since  $\overline{w}_m^o$  is a stationary point of the unconstrained optimization problem  $\overline{\mathcal{P}}_{beam}$ , we can obtain the following result:

$$\nabla_{w_m} \left( \sum_{n=1}^{|S_m|} \overline{R}_{m,n}(\overline{w}_m^o) \right) + \nabla_{w_m} \left( \sum_{x \neq m}^{N_{RF}} \sum_{y=1}^{|S_x|} \overline{R}_{x,y}(\overline{w}_x^o) \right) = 0.$$
(A4)

Observing that  $\nabla_{w_m} \overline{R}_{sum}(tw) = \frac{1}{t} \nabla_{w_m} \overline{R}_{sum}(w)$  for any value t > 0, Equation (A4) can be transformed into the following expression:

$$\nabla_{w_m} \left( \sum_{n=1}^{|S_m|} \overline{R}_{m,n}(\widetilde{w}_m^o) \right) + \nabla_{w_m} \left( \sum_{x \neq m}^{N_{RF}} \sum_{y=1}^{|S_x|} \overline{R}_{x,y}(\widetilde{w}_x^o) \right) = 0.$$
(A5)

Furthermore, we have:

$$\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \left\| \sqrt{a} \overline{w}_m^o \right\|^2 = \frac{P_{max}}{\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \left\| \overline{w}_m^o \right\|^2} \sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \left\| \overline{w}_m^o \right\|^2 = P_{max}.$$
(A6)

That is:

$$\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} p_{m,n} \|\widetilde{w}_m^o\|^2 = P_{max}.$$
(A7)

Specifically, the gradient of the rate of the *n*th user in the *m*th beam with respect to the variable  $w_m$  can be expressed as:

$$\nabla_{w_m} \overline{R}_{m,n} = \frac{\left(h_{m,n}^H w_m\right)^* h_{m,n}^* \sum_{k=1}^n p_{m,k} + \frac{\sigma^2 \sum_{k=1}^{|S_m|} p_{m,n}(w_m)^*}{P_{max}}}{\|h_{m,n}^H w_m\|_2^2 p_{m,n} + \widetilde{\xi}_{m,n}} - \frac{\left(h_{m,n}^H w_m\right)^* h_{m,n}^* \sum_{k=1}^{n-1} p_{m,k} + \frac{\sigma^2 \sum_{k=1}^{|S_m|} p_{m,n}(w_m)^*}{P_{max}}}{\widetilde{\xi}_{m,n}}.$$
 (A8)

Similarly, the gradient of the rate of the *n*th user in the xth( $x \neq m$ ) beam with respect to variable  $w_m$  can be expressed as:

$$\nabla_{w_m} \overline{R}_{x,y} = \frac{\left(h_{x,y}^H w_m\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k} + \frac{\sigma^2 \sum_{n=1}^{|S_m|} p_{m,n}(w_m)^*}{\frac{\left\|h_{x,y}^H w_n\right\|_2^2 p_{x,y} + \tilde{\xi}_{x,y}}{2}} - \frac{\left(h_{x,y}^H w_m\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k} + \frac{\sigma^2 \sum_{n=1}^{|S_m|} p_{m,n}(w_m)^*}{\frac{P_{max}}{\tilde{\xi}_{x,y}}}.$$
 (A9)

Since  $\widetilde{w}_m^o = \sqrt{\frac{P_{max}}{\sum_{m=1}^{N_{RF}} \sum_{n=1}^{|S_m|} ||\overline{w}_m^o||^2}} \overline{w}_m^o, \forall m, (A8) \text{ and } (A9) \text{ can be simplified, the expression after bringing them into (A4) can be specifically expressed as:}$ 

 $\sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} \frac{\left(h_{x,y}^H \tilde{w}_m^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_x^o\right\|_2^2 \sum_{k=1}^{y} p_{x,k} + \sum_{l \neq x}^{N_{RF}} \sum_{k=1}^{|S_l|} \left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} \frac{\left(h_{x,y}^H \tilde{w}_m^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_x^o\right\|_2^2 \sum_{k=1}^{y-1} p_{x,k} + \sum_{l \neq x}^{|S_k|} \left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} \frac{\left(h_{x,y}^H \tilde{w}_m^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_x^o\right\|_2^2 \sum_{k=1}^{y-1} p_{x,k} + \sum_{l \neq x}^{|S_k|} \left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} \frac{\left(h_{x,y}^H \tilde{w}_m^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_x^o\right\|_2^2 \sum_{k=1}^{y-1} p_{x,k} + \sum_{l \neq x}^{|S_m|} \left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{N_{RF}} \sum_{y=1}^{|S_x|} \frac{\left(h_{x,y}^H \tilde{w}_m^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_k^o\right\|_2^2 \sum_{k=1}^{y-1} p_{x,k} + \sum_{l \neq x}^{|S_m|} \left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{|S_m|} \frac{\left(h_{x,y}^H \tilde{w}_k^o\right)^* h_{x,y}^* \sum_{k=1}^{|S_m|} p_{m,k}}{\left\|h_{x,y}^H \tilde{w}_k^o\right\|_2^2 \sum_{k=1}^{|S_m|} p_{x,k} + \sum_{l \neq x}^{|S_m|} \frac{h_{x,y}^H \tilde{w}_l^o}{\left\|h_{x,y}^H \tilde{w}_l^o\right\|_2^2 p_{l,k} + \sigma^2} - \sum_{k=1}^{|S_m|} \frac{h_{x,y}^H \tilde{w}_k^o}{\left\|h_{x,y}^H \tilde{w}_k^o\right\|_2^2 \sum_{k=1}^{|S_m|} \frac{h_{x,y}^H \tilde{w}_k^o}{\left\|h_{x,y}^H \tilde{w}_k^O \tilde{w}_k^o\right\|_2^2 \sum_{k=1}^{|S_m|} \frac{h_{x,y}^H \tilde{w}_k^o}{\left\|h_{x,y}^H \tilde{w}_k^O \tilde{w}_k^o\right\|_2^2 \sum_{k=1}^{|S_m|} \frac{h_{x,y}^H \tilde{w}_k^o}{\left\|h_{x,y}^H \tilde{w}_k^O \tilde{w}_k^O \tilde{w}_k^O \tilde{w}_k^O \tilde{w}_k^O}\right\|_2^2 \sum_{k$ 

$$-(\widetilde{w}_{m}^{o})^{*}\sum_{x=1}^{N_{RF}}\sum_{y=1}^{|S_{x}|}\left[\frac{\frac{\sigma^{2}\sum_{l=1}^{|S_{m}|}p_{m,n}}{\frac{-1}{p_{max}}}}{\|h_{x,y}^{H}\widetilde{w}_{x}^{o}\|_{2}^{2}\sum_{k=1}^{y-1}p_{x,k}+\sum_{l\neq x}^{N_{RF}}\sum_{k=1}^{|S_{l}|}\|h_{x,y}^{H}\widetilde{w}_{l}^{o}\|_{2}^{2}p_{l,k}+\sigma^{2}}-\frac{\frac{\sigma^{2}\sum_{l=1}^{|S_{m}|}p_{m,n}}{\frac{-1}{p_{max}}}}{\|h_{x,y}^{H}\widetilde{w}_{x}^{o}\|_{2}^{2}\sum_{k=1}^{y}p_{x,k}+\sum_{l\neq x}^{N_{RF}}\sum_{k=1}^{|S_{l}|}\|h_{x,y}^{H}\widetilde{w}_{l}^{o}\|_{2}^{2}p_{l,k}+\sigma^{2}}-\frac{\sigma^{2}\sum_{l=1}^{|S_{m}|}p_{m,n}}{\frac{-1}{p_{max}}}}{\|h_{x,y}^{H}\widetilde{w}_{x}^{o}\|_{2}^{2}\sum_{k=1}^{y}p_{x,k}+\sum_{l\neq x}^{N_{RF}}\sum_{k=1}^{|S_{l}|}\|h_{x,y}^{H}\widetilde{w}_{l}^{o}\|_{2}^{2}p_{l,k}+\sigma^{2}}-\frac{\sigma^{2}\sum_{l=1}^{|S_{m}|}p_{m,n}}{\frac{-1}{p_{max}}}}\right]=0,\forall m$$
(A10)

$$\Leftrightarrow \nabla_{w_m} R_{sum}(\widetilde{w}_m^o) - \lambda(\widetilde{w}_m^o)^* = 0$$

where

$$\sum_{x=1}^{N_{RF}}\sum_{y=1}^{|S_x|} \frac{\sigma^2 \sum_{n=1}^{|S_m|} p_{m,n}}{\|h_{x,y}^H \widetilde{w}_x^o\|_2^2 \sum_{k=1}^{y-1} p_{x,k} + \sum_{l\neq x}^{N_{RF}} \sum_{k=1}^{|S_l|} \|h_{x,y}^H \widetilde{w}_l^o\|_2^2 p_{l,k} + \sigma^2} - \sum_{x=1}^{N_{RF}}\sum_{y=1}^{|S_x|} \frac{\sigma^2 \sum_{l=n,n}^{|S_m|} p_{m,n}}{\|h_{x,y}^H \widetilde{w}_l^o\|_2^2 \sum_{k=1}^{|S_m|} p_{k,k} + \sum_{l\neq x}^{N_{RF}} \sum_{k=1}^{|S_l|} \|h_{x,y}^H \widetilde{w}_l^o\|_2^2 p_{l,k} + \sigma^2} = \widetilde{\lambda} > 0.$$
(A11)

This equation implies that  $\tilde{w}_m^o$  satisfies the first-order optimality condition with respect to precoding, where  $\tilde{\lambda}$  is the Lagrange multiplier. Moreover, since  $\tilde{w}_m^o$  satisfies the power constraint and also satisfies the complementary slackness condition, it further satisfies the KKT condition of the original problem  $\tilde{\mathcal{P}}_{beam}$ . Thus,  $\tilde{w}_m^o$  is a nontrivial stationary point of  $\tilde{\mathcal{P}}_{beam}$ . The sufficiency proof is finally complete.

The sufficiency of the proposition can be demonstrated by reversing the steps of sufficiency proof.

#### References

- 1. Ali, M.S.; Tabassum, H.; Hossain, E. Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems. *IEEE Access* 2016, *4*, 6325–6343. [CrossRef]
- Xiao, M.; Mumtaz, S.; Huang, Y.; Dai, L.; Li, Y.; Matthaiou, M.; Karagiannidis, G.K.; Bjornson, E.; Yang, K.; I, C.-L.; et al. Millimeter wave communications for future mobile networks. *IEEE J. Sel. Areas Commun.* 2017, 35, 1909–1935. [CrossRef]
- Xiao, Z.; Zhu, L.; Choi, J.; Xia, P.; Xia, X.-G. Joint Power Allocation and Beamforming for Non-Orthogonal Multiple Access (NOMA) in 5G Millimeter-Wave Communications. *IEEE Trans. Wirel. Commun.* 2018, 17, 2961–2974. [CrossRef]
- Jiao, R.; Dai, L. On the Max-Min Fairness of Beamspace MIMO-NOMA. *IEEE Trans. Signal Process.* 2020, 68, 4919–4932. [CrossRef]
   Lee, J.; So, J. Reinforcement Learning-Based Joint User Pairing and Power Allocation in MIMO-NOMA Systems. *Sensors* 2020, 20, 7094. [CrossRef]

- Shili, M.; Hajjaj, M.; Ammari, M.L. Power Allocation with QoS Satisfaction in MmWave Beamspace MIMO-NOMA. *IET Commun.* 2022, 16, 164–171. [CrossRef]
- Chen, L.; Hu, B.; Xu, G.; Chen, S. Energy-Efficient Power Allocation and Splitting for MmWave Beamspace MIMO-NOMA With SWIPT. *IEEE Sens. J.* 2021, 21, 16381–16394. [CrossRef]
- Majeed, A.A.; Hburi, I. Beamspace-MIMO-NOMA Enhanced Mm-Wave Wireless Communications: Performance Optimization. In Proceedings of the 2022 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq, 15 March 2022; pp. 144–150.
- 9. Lee, Y.-R.; Lee, W.-S.; Jung, J.-S.; Park, C.-Y.; You, Y.-H.; Song, H.-K. Hybrid Beamforming with Reduced RF Chain Based on PZF and PD-NOMA in MmWave Massive MIMO Systems. *IEEE Access* 2021, *9*, 60695–60703. [CrossRef]
- 10. Sur, S.N.; Kandar, D.; Silva, A.; Nguyen, N.D.; Nandi, S.; Do, D.-T. Hybrid Precoding Algorithm for Millimeter-Wave Massive MIMO-NOMA Systems. *Electronics* **2022**, *11*, 2198. [CrossRef]
- 11. Wang, B.; Dai, L.; Wang, Z.; Ge, N.; Zhou, S. Spectrum and Energy-Efficient Beamspace MIMO-NOMA for Millimeter-Wave Communications Using Lens Antenna Array. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 2370–2382. [CrossRef]
- Liu, P.; Li, Y.; Cheng, W.; Zhang, W.; Gao, X. Energy-Efficient Power Allocation for Millimeter Wave Beamspace MIMO-NOMA Systems. *IEEE Access* 2019, 7, 114582–114592. [CrossRef]
- 13. Zhu, L.; Zhang, J.; Xiao, Z.; Cao, X.; Wu, D.O.; Xia, X.-G. Millimeter-Wave NOMA With User Grouping, Power Allocation and Hybrid Beamforming. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5065–5079. [CrossRef]
- Salem, A.A.; Benaya, A.M.; El-Rabaie, S.; Shokair, M. Spectrum Efficiency Analysis of Signal Alignment-based Beamspace Millimetre Wave MIMO-NOMA Systems. *IET Commun.* 2020, 14, 818–829. [CrossRef]
- 15. Zhang, Y.; Deng, H.; He, J.; Zhu, Z.; Peng, C.; Xiao, H. Spectral Efficiency Optimization of Uplink Millimeter Wave MIMO-NOMA Systems. *Sensors* 2022, 22, 6466. [CrossRef] [PubMed]
- Zhu, Z.; Deng, H.; Xu, F.; Zhang, W.; Liu, G.; Zhang, Y. Hybrid Precoding-Based Millimeter Wave Massive MIMO-NOMA Systems. Symmetry 2022, 14, 412. [CrossRef]
- Gao, X.; Dai, L.; Chen, Z.; Wang, Z.; Zhang, Z. Near-Optimal Beam Selection for Beamspace MmWave Massive MIMO Systems. IEEE Commun. Lett. 2016, 20, 1054–1057. [CrossRef]
- 18. Balanis, C.A. Antenna Theory: Analysis and Design, 3rd ed.; John Wiley: Hoboken, NJ, USA, 2005; ISBN 978-0-471-66782-7.
- Sayeed, A.; Brady, J. Beamspace MIMO for High-Dimensional Multiuser Communication at Millimeter-Wave Frequencies. In Proceedings of the 2013 IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, USA, 9–13 December 2013; pp. 3679–3684.
- Kaur, J.; Singh, M.L. User Assisted Cooperative Relaying in Beamspace Massive MIMO NOMA Based Systems for Millimeter Wave Communications. *China Commun.* 2019, 16, 103–113. [CrossRef]
- Wang, J.; Wang, Y.; Yu, J. Joint Beam-Forming, User Clustering and Power Allocation for MIMO-NOMA Systems. Sensors 2022, 22, 1129. [CrossRef]
- Zhao, X.; Lu, S.; Shi, Q.; Luo, Z.-Q. Rethinking WMMSE: Can Its Complexity Scale Linearly with the Number of BS Antennas? IEEE Trans. Signal Process. 2023, 71, 433–446. [CrossRef]
- Shen, K.; Yu, W. Fractional Programming for Communication Systems—Part I: Power Control and Beamforming. *IEEE Trans.* Signal Process. 2018, 66, 2616–2630. [CrossRef]
- Zhu, D.; Li, B.; Liang, P. On the Matrix Inversion Approximation Based on Neumann Series in Massive MIMO Systems. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 1763–1769.
- 25. Zhang, X.; Zeng, H.; Ji, B.; Zhang, G. Low-Complexity Implicit Detection for Massive MIMO Using Neumann Series. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9044–9049. [CrossRef]
- Wang, F.; Zhang, C.; Liang, X.; Wu, Z.; Xu, S.; You, X. Efficient Iterative Soft Detection Based on Polynomial Approximation for Massive MIMO. In Proceedings of the 2015 International Conference on Wireless Communications & Signal Processing (WCSP), Nanjing, China, 11–16 October 2015; pp. 1–5.
- Zhang, Q.; Li, Q.; Qin, J. Robust Beamforming for Nonorthogonal Multiple-Access Systems in MISO Channels. *IEEE Trans. Veh. Technol.* 2016, 65, 10231–10236. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.