Beamforming for Combating Inter-cluster and Intra-cluster Interference in Hybrid NOMA Systems

Zhiyong Chen, Zhiguo Ding, and Xuchu Dai

***Abstract*—In this paper, downlink beamforming (BF) for hybrid non-orthogonal multiple access (NOMA) systems is considered, in order to combat inter and intra cluster interference. Firstly, to minimize the inter and intra cluster interference, the Projection Hybrid NOMA (PH-NOMA) beamforming algorithm is introduced, by combining conventional zero-forcing beamforming (ZFBF) and Hybrid NOMA (H-NOMA) precoding. Secondly, to further reduce the overall interference, two user pairing algorithms, termed Projection Based Pairing Algorithm (PBPA) and Inversion Based Pairing Algorithm (IBPA), are also proposed, by adopting the properties of quasi-degradation developed previously. Consequently, the proposed beamforming algorithm is obtained by combining PH-NOMA and PBPA/IBPA. Moreover, the system performance in terms of outage probability and diversity is analyzed for both the proposed beamforming algorithms and conventional ZFBF. Finally, computer simulations are conducted to demonstrate the efficiency of the proposed beamforming algorithms and to validate the correctness of the performance analysis.**

***Index Terms*—Non-orthogonal multiple access, beamforming, user pairing.**

I. Introduction

THE multiple-input multiple-output (MIMO) broadcast channels have been investigated extensively during the last two decades. By assuming perfect channel state information at transmitter (CSIT), the capacity region can be achieved by dirty-paper coding (DPC), in which all users share the same frequency-time resource [1]. However, DPC is only a theoretic approach and has prohibitive complexity in practical systems. On the other hand, low-complexity orthogonal access schemes, e.g., time/frequency division multiple access (TD/FDMA), can only achieve a small fraction of the capacity [2]. Fortunately, linear precoding techniques have been proposed to reduce the performance gap between DPC and orthogonal multiple access schemes [3] and [4]. It has been shown that ZFBF, can achieve the optimal asymptotic performance with much lower complexity [5]. However, ZFBF may suffer significant performance loss in practical wireless systems especially when the channels are ill-conditioned.

Manuscript received XXX XX, 2016. The associate editor coordinating the review of this paper and approving it for publication was XXX.

This work was supported in part by the National Natural Science Foundation of China (No. 61471334).

Z. Chen and X. Dai are with the Key Laboratory of Wireless-Optical Communications, Chinese Academy of Sciences, School of Information Science and Technology, University of Science and Technology of China, No. 96 Jinzhai Road, Hefei, Anhui Province, 230026, P. R. China. Z. Ding is with the School of Computing and Communications, Lancaster University, LA1 4YW, UK. (email: zhiyong@mail.ustc.edu.cn, z.ding@lancaster.ac.uk, daixc@ustc.edu.cn).

Recently, NOMA has attracted a lot of research interests due to its superior spectral efficiency [6], [7]. For example, NOMA has been proposed for downlink scenarios in 3rd generation partnership project long-term evolution (3GPP-LTE) systems [8]. In addition, NOMA has also been recognized as a promising multiple access candidate for 5G wireless systems [9] and [10].

To further enhance the system capacity, the NOMA-based multi-user beamforming has been introduced in [11] for multiple-input single-output (MISO) NOMA systems. Unlike conventional multi-user beamforming, NOMA-based multi-user beamforming involves designing a single beamforming vector to support multiple users. The number of supportable users as well as the sum capacity can be increased. In [12], a hybrid NOMA approach was proposed, where users were grouped into small-size clusters, NOMA was implemented within each cluster, and MIMO detection was used to cancel inter-cluster interference. Subsequently, it was recognized in [13] that the multi-user beamforming scheme proposed in [11] may become inefficient when taking user fairness into consideration. In [13], by adopting the idea of user grouping, the so-called two-stage beamforming strategy has been proposed. Minimum transmit power was considered within one group, and ZFBF was employed between different groups. A simplified greedy algorithm for user pairing was also proposed by replacing the cost function with user correlation. Lately, a novel concept termed quasi-degradation was introduced in [14], to show the performance gap between MISO-NOMA and optimal DPC. Consequently, a theoretical framework of quasi-degradation was fully characterized in [15], to obtain the closed-form H-NOMA precoding algorithm for two-user MISO-NOMA systems.

In this paper, the downlink beamforming algorithms for multi-user hybrid MISO-NOMA systems is considered. Firstly, the concept of quasi-degradation and the closed-form H-NOMA precoding algorithm are revisited concisely. Secondly, by combining ZFBF and the H-NOMA precoding algorithm, an intra-cluster beamforming algorithm, termed PH-NOMA, is introduced, to minimize both inter and intra cluster interference. Thirdly, by adopting the properties of quasi-degradation, two user pairing algorithms, termed PBPA and IBPA, are proposed, to handle the cases $N \geq K-1$ and $N \geq K$, respectively, where N is the number of transmit antennas at the base station (BS), and K is the number of users (assumed to be a even number). Consequently, the proposed low-complexity beamforming algorithms are obtained, by combining PH-NOMA and PBPA/IBPA. Moreover, the performance analysis is provided to show that the proposed beamforming algorithms are superior to conventional ZFBF in terms of outage performance as well as diversity. Finally, simulation results demonstrate the efficiency of the proposed beamforming algorithms and the correctness of the performance analysis. The contributions of this paper are listed as follows:

1): The differences from the existing work in [13] lie in the following three aspects. Firstly, closed-form PH-NOMA beamforming is used within each cluster, not only significantly reducing the computational complexity, but also enhancing the performance. Secondly, more sophisticated user pairing algorithms are provided, to further improve the overall performance. Finally, a more realistic system model is considered, where the mobile users are randomly deployed within one cell, while the near-far-user distribution has been assumed in [13].

2): The differences from the existing work in [15] are listed as follows. Firstly, PH-NOMA beamforming is proposed, not only minimizing the total power consumption intra cluster, but also perfectly removing the inter-cluster interference, whereas H-NOMA without consideration of inter-cluster interference was utilized in [15]. Secondly, to further enhance the performance of the proposed PH-NOMA beamforming, PBPA/IBPA for user pairing is proposed, to yield a low-complexity beamforming algorithm, whereas H-NOMA precoding and sequential user pairing algorithm (SUPA) were combined in [15]. Finally, performance analysis including diversity order and the expectation of total power consumption for multi-user downlink transmission is provided, whereas [15] only obtained the mathematical results for two-user counterpart.

3): Note that a greedy algorithm often performs poorly in such NOMA systems, since weak users that are not paired well may lead to a significant performance loss. The proposed user pairing algorithms apply the properties of quasi-degradation, and result in a significant performance improvement. Moreover, by utilizing the channel matrix inversion, IBPA is proposed, and significantly reduces the computational complexity without sacrificing any performance gains.

4): The provided performance analysis shows that proposed multi-user beamforming can collect a diversity gain of $N-K+2$, while conventional ZFBF can only realize a diversity gain of $N-K+1$.

The remainder of the paper is organized as follows. Section II briefly describes the system model and the problem formulation. In Section III, some existing transmission schemes are illustrated. Section IV revisits the concept of quasi-degradation and introduces the proposed beamforming algorithms. Performance analysis is given in Section V. Section VI illustrates the simulation results, and Section VI summarizes this paper.

II. System Model and Problem Formulation

In this section, we describe a MISO NOMA downlink system for multi-user beamforming, and formulate an optimization problem for satisfying QoS transmission requirements.

A. System Model

Consider a downlink communication system with one BS and K (assumed to be a even number) mobile users, where the BS is equipped with N transmit antennas and each user is equipped with a single antenna. The BS is located at the center of a disk with radius R, and K mobile users are randomly deployed within this disk. Let

$$\mathbf{H} = (\mathbf{h}_1^H, \mathbf{h}_2^H, ..., \mathbf{h}_K^H)^T \in \mathbb{C}^{K\times N}$$

represent the total channel matrix, where $\mathbf{h}_i^H$ denotes the channel vector between the BS and the i-th user. Let $\mathbf{s} = (s_1, s_2, ..., s_K)^T \in \mathbb{C}^{k\times 1}$ represent the signal (data) vector, where the power normalized signal s_i is intended to the i-th user. Then, the overall system model can be modelled as follows:

$$\mathbf{y} = \mathbf{HWs} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} = (y_1, y_2, ..., y_K)^T \in \mathbb{C}^{K\times 1}$ is the received signal with y_k representing the signal received at the i-th user, $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K) \in \mathbb{C}^{N\times K}$ is the precoding matrix, and $\mathbf{n} = (n_1, n_2, ..., n_K) \in \mathbb{C}^{K\times 1}$ is the additive noise vector with n_i representing the noise at the i-th receiver. For notional convenience, we denote $\mathbf{x} = \mathbf{Ws}$ as the transmit signal. Furthermore, the noise vector $\mathbf{n}$ is assumed to be independent and identically distributed (i.i.d) Gaussian noise, i.e., $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_K)$. The channel vector between the BS and the i-th user $\mathbf{h}_i$ is modelled as follows:

$$\begin{aligned} &\mathbf{h}_i \sim \mathcal{CN}(0, \sigma_i^2 \mathbf{I}_N), \\ &\sigma_i^2 = \min(d_i^{-\alpha}, d_0^{-\alpha}), \end{aligned} \tag{2}$$

where d_i is the distance between the BS and the i-th user, and α is the path loss exponent. The parameter d_0 is introduced to avoid singularity when d_i is small.

B. Problem Formulation

In this paper, we consider a QoS optimization problem to minimize the total power for supporting reliable communications. Given a set of target interference levels $r = [r_1, r_2, ..., r_K]$ with r_i denoting the target interference level of the i-th user, the optimization problem can be formulated as follows:

$$\begin{aligned} \min \quad & \mathrm{tr}(\mathbf{W}^H\mathbf{W}) \\ \text{s.t.} \quad & \mathrm{SINR}_i \geq r_i, \quad i = 1, 2, ..., K, \end{aligned} \tag{3}$$

where SINR_i denotes the signal to interference plus noise ratio (SINR) of user i.

III. Existing Transmission schemes

A. Conventional Multi-user Beamforming

We rewrite the system model in (1) by focusing on the received signal at user i:

$$y_i = \mathbf{h}_i^H \mathbf{w}_i s_i + \sum_{k=1, k\neq i}^{K} \mathbf{h}_i^H \mathbf{w}_k s_k + n_i, \tag{4}$$

where the second term is referred to as the co-channel interference (CCI). Consequently, if a linear receiver is adopted at each user side, the SINR of user i can be written as

$$\mathrm{SINR}_i = \frac{\mathbf{w}_i^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{w}_i}{1 + \sum_{k=1, k\neq i}^{K} \mathbf{w}_k^H \mathbf{h}_i \mathbf{h}_i^H \mathbf{w}_k}. \tag{5}$$

Generally, this expression results in a non-convex optimization problem with K coupled variables [16]. To avoid optimization with the coupled variables, the zero-forcing (ZF) scheme is utilized to decouple all the precoding vectors $\mathbf{w}_i$, by enforcing the following conditions [17] and [4]:

$$\begin{cases} \mathbf{h}_i^H \mathbf{w}_k = 0, \quad i \neq k \\ \mathbf{h}_i \mathbf{w}_i = \sqrt{r_i}, \quad , i = 1, 2, ...K \end{cases}. \tag{6}$$

Note that the ZF solution can be briefly written as

$$\mathbf{W}_{ZF} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{\Gamma}, \tag{7}$$

where

$$\mathbf{\Gamma} = \mathrm{diag}([\sqrt{r_1}, \sqrt{r_2}, ..., \sqrt{r_K}]).$$

However, the extra conditions in (6) lead to a suboptimal solution. Moreover, to make the solution exist, the number of transmit antennas needs to be at least as large as the number of all the receive antennas.

B. NOMA Based Multi-user Beamforming

Unlike conventional multi-user beamforming, NOMA based multi-user beamforming involves designing a single BF vector to support multiple users [11]. Consequently, the number of supportable users can be increased, thus enhancing the sum capacity.

In [11], it was suggested that the users have high correlation should be paired into one cluster, and the BF vector was generated by the channel of the strong user in each group. Specifically, if the near user $s_{n,k}$ and the far user $s_{f,k}$ paired within a group k, the transmitted signal can be written as:

$$\mathbf{x} = \sum_{k=1}^{K/2} \mathbf{c}_k (\sqrt{\alpha_k P_k} s_{n,k} + \sqrt{(1-\alpha_k) P_k} s_{f,k}), \tag{8}$$

where $\mathbf{c}_k$ lies in the null space of all the other channel vectors of near users, $\alpha_k P_k$ and $(1-\alpha_k) P_k$ are the power parameters for the near and far users, respectively. However, as pointed out in [13], this scheme can be rather inefficient when taking user fairness into consideration, since that α_k has to be close to 0 if the far user needs a data rate comparable to that of the near user.

In [13], the author made a near-far-user assumption, i.e., there are $K/2$ near users and $K/2$ far users in the broadcast system. The precoding vectors were obtained by employing so-called two-stage beamforming. To avoid the interference from other groups, zero-forcing beamforming is firstly employed. Within each group, minimum transmit power with successive interference cancellation (SIC) at the near user is considered. However, the optimal precoding vector within a group does not have a closed-form expression, thus have a considerable computational complexity. Moreover, the proposed simplified greedy pairing algorithm performs well only if the near-far-user assumption holds, and it will be shown in our simulations that this pairing algorithm results in a significant performance loss when mobile users are deployed randomly in a cell, which is more close to reality.

IV. Proposed Beamforming Algorithms

A. Quasi-degradation and H-NOMA Precoder Revisit

Consider a two-user MISO-NOMA downlink system with user m and user n. User m is assumed to be the strong user. Herein, the strong user is defined as the user employing SIC to decode its message. Conversely, the weak user is the user directly decoding its own messages. In NOMA systems, the SINR for user m and user n are SINR_m and $\min(\mathrm{SINR}_{n,m}, \mathrm{SINR}_{n,n})$, respectively, where

$$\begin{cases} \mathrm{SINR}_m = \mathbf{h}_m^H \mathbf{w}_m \mathbf{w}_m^H \mathbf{h}_m \\ \mathrm{SINR}_{n,m} = \dfrac{\mathbf{h}_m^H \mathbf{w}_n \mathbf{w}_n^H \mathbf{h}_m}{1 + \mathbf{h}_m^H \mathbf{w}_m \mathbf{w}_m^H \mathbf{h}_m} \\ \mathrm{SINR}_{n,n} = \dfrac{\mathbf{h}_n^H \mathbf{w}_n \mathbf{w}_n^H \mathbf{h}_n}{1 + \mathbf{h}_n^H \mathbf{w}_m \mathbf{w}_m^H \mathbf{h}_n} \end{cases}. \tag{9}$$

Hence, by assuming a fixed decoding order, the minimum power optimization problem for two users with channel vectors $\mathbf{h}_m, \mathbf{h}_n$ given target SINR levels r_m, r_n can be formulated as:

$$\begin{aligned} \min \quad & \|\mathbf{w}_m\|^2 + \|\mathbf{w}_n\|^2 \\ \text{s.t.} \quad & \mathrm{SINR}_m \geq r_m \\ & \min(\mathrm{SINR}_{n,m}, \mathrm{SINR}_{n,n}) \geq r_n. \end{aligned} \tag{10}$$

In [14], the definition of quasi-degradation was first introduced. In [15], the closed-form precoding vectors were obtained for quasi-degraded channels and the H-NOMA precoding algorithm was proposed with closed-form expressions by applying the properties of quasi-degradation. The definition of quasi-degradation can be briefly illustrated as follows. For more details, please refer to [15].

Definition 1 (Quasi-Degradation [15]). *The broadcast channels $\mathbf{h}_m$ and $\mathbf{h}_n$ with respect to r_m and r_n are called quasi-degraded if and only if the minimum transmission power of NOMA is equivalent to that of DPC.*

As studied in our previous work, the optimal closed-form precoding vectors and the explicit sufficient and necessary condition for quasi-degraded channels were introduced in Proposition 1 and Proposition 2 in [15], respectively. The channel correlation between $\mathbf{h}_m$ and $\mathbf{h}_n$ is defined as

$$u = \cos^2\theta = \frac{\mathbf{h}_n^H \mathbf{h}_m \mathbf{h}_m^H \mathbf{h}_n}{\|\mathbf{h}_m\|^2 \|\mathbf{h}_n\|^2}.$$

Proposition 1 (Closed-form Solution [15]). *The optimal beamforming vectors are*

$$\begin{cases} \mathbf{w}_m^{N*} = \alpha_m ((1 + r_n)\mathbf{e}_m - r_n \mathbf{e}_n^H \mathbf{e}_m \mathbf{e}_n) \\ \mathbf{w}_n^{N*} = \alpha_n \mathbf{e}_n \end{cases}, \tag{11}$$

if $\mathbf{h}_m$ and $\mathbf{h}_n$ with respect to r_m and r_n are quasi-degraded, where

$$\begin{cases} \mathbf{e}_m = \mathbf{h}_m / \|\mathbf{h}_m\|, \mathbf{e}_n = \mathbf{h}_n / \|\mathbf{h}_n\| \\ \alpha_m^2 = \dfrac{r_m}{\|\mathbf{h}_m\|^2} \dfrac{1}{(1 + r_n \sin^2\theta)^2} \\ \alpha_n^2 = \dfrac{r_n}{\|\mathbf{h}_n\|^2} + \dfrac{r_m}{\|\mathbf{h}_m\|^2} \dfrac{r_n \cos^2\theta}{(1 + r_n \sin^2\theta)^2} \end{cases}.$$

Proposition 2 (Necessary and Sufficient Condition [15]). *The broadcast channels $\mathbf{h}_m$ and $\mathbf{h}_n$ with respect to r_m and r_n are quasi-degraded, if and only if*

$$Q(u) \leq \frac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_n\|^2}, \tag{12}$$

where $Q(u) = \frac{1+r_m}{u} - \frac{r_m u}{(1+r_n(1-u))^2}$.

The proofs of Propositions 1 and 2 can be found in [15]. By using Propositions 1 and 2, the H-NOMA precoding algorithm was proposed. The key idea is that the NOMA signal is transmitted if the channels are quasi-degraded, and the ZFBF signal is transmitted otherwise [15]. The required transmission power of H-NOMA with parameters $(\mathbf{h}_m, \mathbf{h}_n, r_m, r_n)$ can be written as:

$$P_{m,n}^{\mathrm{H-NOMA}} = \begin{cases} P_1, & \text{if} \quad Q(u) \leq \dfrac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_n\|^2} \\ P_2, & \text{otherwise} \end{cases}, \tag{13}$$

where

$$\begin{cases} P_1 = \dfrac{r_n}{\|\mathbf{h}_n\|^2} + \dfrac{r_m}{\|\mathbf{h}_m\|^2} \dfrac{1+r_n}{1+r_n \sin^2\theta} \\ P_2 = \dfrac{1}{\sin^2\theta}(\dfrac{r_m}{\|\mathbf{h}_m\|^2} + \dfrac{r_n}{\|\mathbf{h}_n\|^2}) \end{cases}. \tag{14}$$

B. Projection H-NOMA beamforming (PH-NOMA)

This subsection provides an efficient intra-cluster beamforming algorithm, termed PH-NOMA, not only minimizing the total power consumption intra cluster, but also perfectly removing inter-cluster interference.

Assume that user m and user n are paired into one cluster. By taking inter-cluster interference into consideration, the effective SINR for user m and user n may be attenuated compared to that of the two-user MISO-NOMA systems given in (9). For example, the SINR for user m becomes:

$$\mathrm{SINR}_m = \frac{\mathbf{h}_m^H \mathbf{w}_m \mathbf{w}_m^H \mathbf{h}_m}{1 + \mathbf{h}_m^H (\sum_{k=1, k\neq m,n}^K \mathbf{w}_k \mathbf{w}_k^H) \mathbf{h}_m}. \tag{15}$$

The second term of the denominator in (15) is the interference from other clusters. To avoid this interference, a straightforward solution is to employ ZFBF technique between clusters, i.e., the beamforming vector for user m must satisfy:

$$\begin{cases} \mathbf{h}_k^H \mathbf{w}_m = 0, \quad k \neq m, n \\ \mathrm{SINR}_m \geq r_m \\ \min(\mathrm{SINR}_{n,m}, \mathrm{SINR}_{n,n}) \geq r_n \end{cases}. \tag{16}$$

Denote $\mathbf{H}_{m,n}^H$ the sub-matrix obtained by striking $\mathbf{h}_m$ and $\mathbf{h}_n$ out of $\mathbf{H}^H$, i.e.,

$$\mathbf{H}_{m,n}^H = (\mathbf{h}_1, ..., \mathbf{h}_{n-1}, \mathbf{h}_{n+1}, ..., \mathbf{h}_{m-1}, \mathbf{h}_{m+1}, ..., \mathbf{h}_K).$$

Note that $\mathbf{H}_{m,n}^H \in \mathbb{C}^{N\times(K-2)}$, we assume that $N > K-2$. Then the orthogonal projection $\mathcal{P}_{m,n}^\perp$ of $\mathbf{H}_{m,n}^H$ can be written as

$$\mathcal{P}_{m,n}^\perp = \mathbf{I}_N - \mathbf{H}_{m,n}^H (\mathbf{H}_{m,n}\mathbf{H}_{m,n}^H)^{-1} \mathbf{H}_{m,n}. \tag{17}$$

It is clear that

$$\mathcal{P}_{m,n}^\perp \mathbf{h}_k = 0, \quad k \neq m, n. \tag{18}$$

The superposition for transmitting s_m and s_n can be formulated as

$$\mathbf{x} = \mathcal{P}_{m,n}^\perp (\mathbf{w}_m s_m + \mathbf{w}_n s_n).$$

Note that $\mathbf{h}_k^H \mathbf{x} = 0, k \neq m, n$ since $\mathcal{P}_{m,n}^\perp$ is the orthogonal projection of $\mathbf{H}_{m,n}^H$. Therefore, the received signal of user m can be written as

$$\begin{aligned} y_m &= \mathbf{h}_m^H \mathcal{P}_{m,n}^\perp (\mathbf{w}_m s_m + \mathbf{w}_n s_n) + n_m \\ &= (\mathcal{P}_{m,n}^\perp \mathbf{h}_m)^H (\mathbf{w}_m s_m + \mathbf{w}_n s_n) + n_m. \end{aligned} \tag{19}$$

Similarly, the received signal of user n can be written as

$$\begin{aligned} y_n &= \mathbf{h}_n^H \mathcal{P}_{m,n}^\perp (\mathbf{w}_m s_m + \mathbf{w}_n s_n) + n_n \\ &= (\mathcal{P}_{m,n}^\perp \mathbf{h}_n)^H (\mathbf{w}_m s_m + \mathbf{w}_n s_n) + n_n. \end{aligned} \tag{20}$$

Thus, $\mathbf{w}_m$ and $\mathbf{w}_n$ can be obtained by applying the H-NOMA precoding algorithm with parameters $(\mathcal{P}_{m,n}^\perp \mathbf{h}_m, \mathcal{P}_{m,n}^\perp \mathbf{h}_n, r_m, r_n)$. By applying Propositions 1 and 2, the proposed PH-NOMA algorithm for intra-cluster beamforming is obtained. Specifically, if $\mathcal{P}_{m,n}^\perp \mathbf{h}_m$ and $\mathcal{P}_{m,n}^\perp \mathbf{h}_n$ are quasi-degraded, i.e.,

$$Q(u_P) \leq \frac{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_m\|^2}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_n\|^2},$$

where

$$u_P = \frac{\mathbf{h}_n^H \mathcal{P}_{m,n}^\perp \mathbf{h}_m \mathbf{h}_m^H \mathcal{P}_{m,n}^\perp \mathbf{h}_n}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_m\|^2 \|\mathcal{P}_{m,n}^\perp \mathbf{h}_n\|^2},$$

then, the closed-form beamforming vectors can be written as

$$\begin{cases} \mathbf{w}_m = \alpha_m((1+r_n)\mathbf{e}_m - r_n \mathbf{e}_n^H \mathbf{e}_m \mathbf{e}_n) \\ \mathbf{w}_n = \alpha_n \mathbf{e}_n \end{cases}, \tag{21}$$

where

$$\begin{cases} \mathbf{e}_m = \dfrac{\mathcal{P}_{m,n}^\perp \mathbf{h}_m}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_m\|}, \mathbf{e}_n = \dfrac{\mathcal{P}_{m,n}^\perp \mathbf{h}_n}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_n\|} \\ \alpha_m^2 = \dfrac{r_m}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_m\|^2} \dfrac{1}{(1+r_n \sin^2\theta)^2} \\ \alpha_n^2 = \dfrac{r_n}{\|\mathcal{P}_{m,n}^\perp \mathbf{h}_n\|^2} + \dfrac{r_m}{\|\mathbf{h}_m\|^2} \dfrac{r_n \cos^2\theta}{(1+r_n \sin^2\theta)^2} \end{cases}.$$

C. User Pairing Algorithms PBPA/IBPA

To better illustrate the optimization problem when taking user pairing into consideration, we first define some sets. The user index set is denoted by $\mathcal{K} = \{1, 2, 3, ..., K\}$. The permutation function in K is denoted by $\sigma : \mathcal{K} \to \mathcal{K}$. We define a set of permutation functions as

$$\mathcal{A} = \{\sigma \mid \sigma(i) \neq i, \sigma(\sigma(i)) = i, \forall i \in \mathcal{K}\}.$$

The set of strong users can be defined as

$$\mathcal{I} = \{i \mid \|\mathbf{h}_i\| \geq \|\mathbf{h}_{\sigma(i)}\|, i \in \mathcal{K}\}.$$

Then, the user pairing configuration can be defined as

$$\Pi = \{(i, \sigma(i)) \mid i \in \mathcal{I}\}.$$

It can be observed that there is a one-to-one mapping relationship between all the pairing configurations and the permutation functions $\sigma \in \mathcal{A}$. For example, we consider a simple user index set $\{1,2,3,4\}$. A mapping between Π and σ is given as follows:

$$\Pi = \{(1,2),(3,4)\} \longleftrightarrow \sigma : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}.$$

The goal of the user pairing algorithm considered in this paper is to minimize the total transmission power. Mathematically, the optimization problem can be rewritten as:

$$\min_{\sigma \in \mathcal{A}} \sum_{i \in \mathcal{I}} P^{\mathrm{PH-NOMA}}_{i,\sigma(i)}, \tag{22}$$

where

$$P^{\mathrm{PH-NOMA}}_{m,n} = \begin{cases} P_1, & \text{quasi-degraded} \\ P_2, & \text{otherwise} \end{cases}, \tag{23}$$

and

$$\begin{cases} P_1 = \dfrac{r_n}{\|\mathcal{P}^{\perp}_{m,n}\mathbf{h}_n\|^2} + \dfrac{r_m}{\|\mathcal{P}^{\perp}_{m,n}\mathbf{h}_m\|^2}\dfrac{1+r_n}{1+r_n\sin^2\theta} \\ P_2 = \dfrac{1}{\sin^2\theta}(\dfrac{r_m}{\|\mathcal{P}^{\perp}_{m,n}\mathbf{h}_m\|^2} + \dfrac{r_n}{\|\mathcal{P}^{\perp}_{m,n}\mathbf{h}_n\|^2}) \end{cases}. \tag{24}$$

By observing (23) and (24), we can conclude some properties of the proposed PH-NOMA beamforming and the basic principles for user pairing as listed following.

1) The power consumption of quasi-degraded channels is minimized due to the definition of quasi-degraded channels. Therefore, one user is tend to be paired with another one with quasi-degraded channels as much as possible.
2) The performance loss is dominated by those users that 1) cannot be paired to form quasi-degraded channels, 2) have channel vectors with small 2-norm. Therefore, weak users have higher priority to pair, i.e., users should be sorted by their channels' 2-norm before pairing.
3) This performance loss becomes even more significant as the users' correlation increased. Therefore, users who cannot be paired with quasi-degraded channels should be paired with the one with minimum correlation.

Consequently, we propose two algorithms for user pairing to solve this problem, termed Projection Based Pairing Algorithm (PBPA) and Inversion Based Pairing Algorithm (IBPA), which are illustrated in detail in Algorithm 1 and Algorithm 2, respectively.

The difference between PBPA and IBPA is that, IBPA provides a more intelligent method to compute the projection matrix, provided that $N \geq K$ holds, i.e., the left inverse matrix of $\mathbf{H}^H$ exists. Specifically, we denote the left inverse matrix of $\mathbf{H}^H$ by $\mathbf{V}$, i.e., $\mathbf{V}\mathbf{H}^H = \mathbf{I}_K$. The i-th row of $\mathbf{V}$ is denoted by $\mathbf{v}_i^H$. Then, we have

$$\begin{cases} \mathbf{v}_i^H\mathbf{h}_i = 1, & i = 1,2,...,K \\ \mathbf{v}_i^H\mathbf{h}_j = 0, & i \neq j \end{cases}. \tag{25}$$

Given a pair of users m and n, the projection matrix $\mathbf{Q}$ satisfies

$$\mathbf{Q}\mathbf{h}_k = 0, \quad k \neq m,n. \tag{26}$$

Note that $\mathbf{v}_m$ and $\mathbf{v}_n$ lie in the null space of $\mathbf{H}^H_{m,n}$. Hence, the projection matrix $\mathbf{Q}$ of the subspace spanned by $\mathbf{v}_m, \mathbf{v}_n$ satisfies the conditions in (26). Mathematically, $\mathbf{Q}$ can be alternatively obtained by performing the following procedures.

$$\begin{aligned} \mathbf{v}_m &= \mathbf{v}_m/\|\mathbf{v}_m\|, \\ \mathbf{v}_n &= \mathbf{v}_n - (\mathbf{v}_m^H\mathbf{v}_n)\mathbf{v}_m, \\ \mathbf{v}_n &= \mathbf{v}_n/\|\mathbf{v}_n\|, \\ \mathbf{Q} &= \mathbf{v}_m\mathbf{v}_m^H + \mathbf{v}_n\mathbf{v}_n^H. \end{aligned} \tag{27}$$

Consequently, the proposed beamforming algorithms are obtained, by combining PH-NOMA beamforming as intra-cluster beamforming and PBPA/IBPA as user pairing algorithms.

The difference between H-NOMA/SUPA [15] and the proposed beamforming algorithms can be concluded as follows.

1) Note that the H-NOMA precoding is originally proposed for two-user scenario, thus, it does not take inter-cluster interference into consideration. Moreover, the H-NOMA precoding for K-user ($K > 2$) counterpart is not specified in [15]. Mathematically, if users m and n are paired into one cluster ($\|\mathbf{h}_m\| \geq \|\mathbf{h}_n\|$), the H-NOMA precoding vectors $\mathbf{w}_i$ ($i = m,n$) should be obtained by solving the following optimization problem.

$$\begin{aligned} \min \quad & \mathbf{w}_m^H\mathbf{w}_m + \mathbf{w}_n^H\mathbf{w}_n \\ \text{s.t.} \quad & \frac{\mathbf{h}_m^H\mathbf{w}_m\mathbf{w}_m^H\mathbf{h}_m}{1+\mathbf{h}_m^H(\sum_{k=1,k\neq m,n}^K \mathbf{w}_k\mathbf{w}_k^H)\mathbf{h}_m} \geq r_m \\ & \frac{\mathbf{h}_n^H\mathbf{w}_n\mathbf{w}_n^H\mathbf{h}_n}{1+\mathbf{h}_n^H(\sum_{k=1,k\neq n}^K \mathbf{w}_k\mathbf{w}_k^H)\mathbf{h}_n} \geq r_n. \end{aligned}$$

However, this problem is not investigated in [15], and the optimal solution is also difficult to obtain.
2) The H-NOMA/SUPA scheme proposed in [15] is efficient only when $N < K$, whereas the proposed PH-NOMA/PBPA (PH-NOMA/IBPA) scheme can be applied only when $N \geq K-1$ ($N \geq K$) holds.

Proposition 3. *The complexities of PBPA scheme and IBPA are $\mathcal{O}(K^2(K^3+K^2N+KN^2))$ and $\mathcal{O}(K^2(N+K))$, respectively. While the exhaustive search requires at least a complexity of $\mathcal{O}((K-1)!!N)$*

Proof: For PBPA, the number of "for" loops is

$$\sum_{t=1}^{K-1}(K-t) = \frac{K(K-1)}{2}.$$

In each "for" loop, the complexity is dominated by the computation of the projection matrix $\mathbf{P}$ in line No. 12 in the table for PBPA. Note that $\mathbf{H}^H_{t,v}$ is a $N \times (K-2)$ matrix, the complexity of computing $\mathbf{H}_{t,v}\mathbf{H}^H_{t,v}$ is $N(K-2)^2$, the complexity of computing $(\mathbf{H}_{t,v}\mathbf{H}^H_{t,v})^{-1}$ is $(K-2)^3$, the complexity of computing $\mathbf{H}^H_{t,v}(\mathbf{H}_{t,v}\mathbf{H}^H_{t,v})^{-1}$ is $N(K-2)^2$, and the complexity of computing $\mathbf{H}^H_{t,v}(\mathbf{H}_{t,v}\mathbf{H}^H_{t,v})^{-1}\mathbf{H}_{t,v}$ is $N^2(K-2)$. Therefore, the overall complexity for computing $\mathbf{P}$ in each loop can be written as $\mathcal{O}(K^3+K^2N+KN^2)$. Therefore, the overall complexity of PBPA is

$$\mathcal{O}(K^2(K^3+K^2N+KN^2)).$$

For IBPA, the complexity is dominated by the computing of the inverse matrix $\mathbf{V}$, and the "for" loop. Similarly, we can conclude that the complexity of computing $\mathbf{V}$ is $\mathcal{O}(K^3+2K^2N)$. Note that there is only a complexity of $\mathcal{O}(N)$ in each loop. Hence, the overall complexity of IBPA can be written as

$$\mathcal{O}(K^3+K^2N)=\mathcal{O}(K^2(K+N)).$$

For exhaustive search, note that the cardinal number of $\mathcal{A}$ is $|\mathcal{A}|=(K-1)!!$. Therefore, the complexity of exhaustive search is at least $\mathcal{O}((K-1)!!N)$, and the proof is finally completed. □

Note that a greedy algorithm can also be straightforwardly obtained, since the closed-form expressions of the cost function $P_{i,\sigma(i)}^{\mathrm{PH-NOMA}}$ can be acquired by (23) and (24). However, this greedy algorithm cannot obtain satisfactory performance since the weak users can lead to a significant performance loss, which can be seen in the simulation results. In [13], the author developed a simplified greedy algorithm by replacing the cost function with the correlation between the two channels in one group, also termed the correlation based user pairing algorithm in this paper for simplicity. It is suggested that one user pairs with another user with maximal correlation. Although the complexity is reduced, the performance is unsatisfactory.

Algorithm 1 Projection Based Pairing Algorithm (PBPA)

INPUT: $\mathbf{h}_1,\mathbf{h}_2,...,\mathbf{h}_K$ and $r_1,r_2,...,r_K$, $N\geq K-1$
OUTPUT: Π

1: $\sigma\leftarrow zeros(K,1)$
2: $\Pi\leftarrow\{\}$
3: Sort the users by the length of channel vectors, such that $\|\mathbf{h}_{i_1}\|\leq\|\mathbf{h}_{i_2}\|\leq...\leq\|\mathbf{h}_{i_K}\|$
4: **for** $t=1:K-1$ **do**
5:     **if** User i_t has been paired **then**
6:         Goto Line 4
7:     **end if**
8:     **for** $v=t+1:K$ **do**
9:         **if** User i_v has been paired **then**
10:             Goto Line 4
11:         **end if**
12:         $\mathbf{P}=\mathbf{I}_N-\mathbf{H}_{t,v}^H(\mathbf{H}_{t,v}\mathbf{H}_{t,v}^H)^{-1}\mathbf{H}_{t,v}$
13:         $\mathbf{h}_m=\mathbf{P}\mathbf{h}_{i_v},\mathbf{h}_n=\mathbf{P}\mathbf{h}_{i_t}$
14:         $u(t,v)=\frac{\mathbf{h}_m^H\mathbf{h}_n\mathbf{h}_n^H\mathbf{h}_m}{\|\mathbf{h}_n\|^2\|\mathbf{h}_m\|^2}$
15:         **if** $Q(u)\leq\frac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_n\|^2}$ **then**
16:             $\sigma(i_t)=i_v$,$\sigma(i_v)=i_t$
17:             $\Pi=\Pi\cup\{(i_v,i_t)\}$
18:             Goto Line 4
19:         **end if**
20:     **end for**
21:     $minv=\underset{v}{\mathrm{argmin}}\quad u(t,v)$
22:     $\sigma(i_t)=i_{minv}$,$\sigma(i_{minv})=i_t$
23:     $\Pi=\Pi\cup\{(i_{minv},i_t)\}$
24: **end for**

V. Performance Analysis

In this section, the performance comparison between the conventional ZFBF and the proposed beamforming algorithms

Algorithm 2 Inversion Based Pairing Algorithm (IBPA)

INPUT: $\mathbf{h}_1,\mathbf{h}_2,...,\mathbf{h}_K$ and $r_1,r_2,...,r_K$, $N\geq K$
OUTPUT: Π

1: $\sigma\leftarrow zeros(K,1)$
2: $\Pi\leftarrow\{\}$
3: Sort the users by the length of channel vectors, such that $\|\mathbf{h}_{i_1}\|\leq\|\mathbf{h}_{i_2}\|\leq...\leq\|\mathbf{h}_{i_K}\|$
4: Calculate $\mathbf{V}=(\mathbf{H}^H)^\dagger=(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{H}$
5: **for** $t=1:K-1$ **do**
6:     **if** User i_t has been paired **then**
7:         Goto Line 5
8:     **end if**
9:     **for** $v=t+1:K$ **do**
10:         **if** User i_v has been paired **then**
11:             Goto Line 5
12:         **end if**
13:         Calculate $\mathbf{Q}$ by (27)
14:         $\mathbf{h}_m=\mathbf{Q}\mathbf{h}_{i_v},\mathbf{h}_n=\mathbf{Q}\mathbf{h}_{i_t}$
15:         $u(t,v)=\frac{\mathbf{h}_m^H\mathbf{h}_n\mathbf{h}_n^H\mathbf{h}_m}{\|\mathbf{h}_n\|^2\|\mathbf{h}_m\|^2}$
16:         **if** $Q(u)\leq\frac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_n\|^2}$ **then**
17:             $\sigma(i_t)=i_v$,$\sigma(i_v)=i_t$
18:             $\Pi=\Pi\cup\{(i_v,i_t)\}$
19:             Goto Line 5
20:         **end if**
21:     **end for**
22:     $minv=\underset{v}{\mathrm{argmin}}\quad u(t,v)$
23:     $\sigma(i_t)=i_{minv}$,$\sigma(i_{minv})=i_t$
24:     $\Pi=\Pi\cup\{(i_{minv},i_t)\}$
25: **end for**

is analyzed. In order to derive the diversity order, the following Lemmas are introduced.

Lemma 1. *For a chi-squared distributed random variable x with $2n$ degrees of freedom, i.e., $x\sim\mathcal{X}_{2n}^2$,*

$$\lim_{\rho\to+\infty}-\frac{\log Pr\{x^{-1}\geq\frac{\rho}{c}\}}{\log\rho}=n,$$

where c is a positive constant.

Proof: The probability density function (pdf) of x can be written as

$$f(x)=\frac{1}{2^n(n-1)!}x^{n-1}e^{-x/2}.$$

Then, we have

$$\begin{aligned}&Pr\{x^{-1}\geq\frac{\rho}{c}\}=Pr\{x\leq c\rho^{-1}\}\\&=\int_0^{c\rho^{-1}}f(x)dx\\&=1-e^{-\frac{1}{2}c\rho^{-1}}\sum_{r=0}^{n-1}\frac{(\frac{1}{2}c\rho^{-1})^r}{r!}\\&=e^{-\frac{1}{2}c\rho^{-1}}\sum_{r=n}^{+\infty}\frac{(\frac{1}{2}c\rho^{-1})^r}{r!}\end{aligned}\tag{28}$$

Hence,

$$\lim_{\rho\to+\infty}-\frac{\log Pr\{x^{-1}\geq\frac{\rho}{c}\}}{\log\rho}=n,$$

and the proof is completed. □

Lemma 2. *For two i.i.d chi-squared distributed random variables x_1, x_2 with $2n$ degrees of freedom, i.e., $x_1, x_2 \sim \mathcal{X}_{2n}^2$,*

$$\lim_{\rho\to+\infty} -\frac{\log Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}}{\log\rho}=n,$$

where c_1 and c_2 are some positive constants.

Proof: We focus on the probability $Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}$. Firstly, it is straightforward to see that

$$Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}\geq Pr\{x_1^{-1}\geq\frac{\rho}{c_1}\}. \quad (29)$$

Secondly, we have

$$\begin{aligned}&Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}\\ &\leq Pr\{\big(x_1^{-1}\geq\frac{\beta\rho}{c_1}\big)\cup\big(x_2^{-1}\geq\frac{(1-\beta)\rho}{c_2}\big)\}\\ &\leq Pr\{x_1^{-1}\geq\frac{\beta\rho}{c_1}\}+Pr\{x_2^{-1}\geq\frac{(1-\beta)\rho}{c_2}\},\end{aligned} \quad (30)$$

where $\beta\in[0,1]$. Fix $\beta=\frac{c_1}{c_1+c_2}$, we have

$$Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}\leq 2Pr\{x_1\geq\frac{\rho}{c_1+c_2}\} \quad (31)$$

By combining (29) and (31), we can bound this probability by

$$\begin{cases}Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}\geq Pr\{x_1^{-1}\geq\frac{\rho}{c_1}\}\\ Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}\leq 2Pr\{x_1^{-1}\geq\frac{\rho}{c_1+c_2}\}\end{cases}. \quad (32)$$

Hence, by using Lemma 1, we have

$$\lim_{\rho\to+\infty} -\frac{\log Pr\{c_1x_1^{-1}+c_2x_2^{-1}\geq\rho\}}{\log\rho}=n,$$

and the proof is completed. □

By taking advantage of Lemmas 1 and 2, we have the following theorem for obtaining the diversity order achieved by the conventional ZFBF and the proposed beamforming algorithms, respectively.

Theorem 1. *The diversity order achieved by conventional ZFBF for each user is*

$$d_{ZF}(i)=N-K+1,\quad i=1,2,...,K,$$

while the diversity order achieved by the proposed beamforming algorithms is

$$d_{Pro}(i)=N-K+2,$$

if user i $(i=1,2,...,K)$ is perfectly paired[1].

[1]User i is called perfectly paired if it is paired with another user j with quasi-degraded channels, i.e., $\mathbf{h}_i$ and $\mathbf{h}_j$ are quasi-degraded.

Proof: We first derive the diversity order for conventional ZFBF. From (7), it is clear that the power allocated to the i-th $(i=1,2,...,K)$ user is

$$\begin{aligned}P_i&=[\mathbf{W}_{ZF}^H\mathbf{W}_{ZF}]_{ii}\\ &=[\mathbf{\Gamma}^H(\mathbf{HH}^H)^{-H}\mathbf{HH}^H(\mathbf{HH}^H)^{-1}\mathbf{\Gamma}]_{ii}\\ &=r_i[(\mathbf{HH}^H)^{-1}]_{ii}\\ &=\frac{r_i}{\mathbf{h}_i^H\mathbf{H}_i^H(\mathbf{H}_i\mathbf{H}_i^H)^{-1}\mathbf{H}_i\mathbf{h}_i}\\ &=\frac{r_i}{\mathbf{h}_i^H(\mathbf{I}-\mathbf{H}_i^H(\mathbf{H}_i\mathbf{H}_i^H)^{-1}\mathbf{H}_i)\mathbf{h}_i}\\ &=\frac{r_i}{\|\mathcal{P}_i^{\perp}\mathbf{h}_i\|^2},\end{aligned} \quad (33)$$

where $[(\mathbf{HH}^H)^{-1}]_{ii}$ denotes the i-th diagonal element of matrix $(\mathbf{HH}^H)^{-1}$, $\mathbf{H}_i^H$ represents the sub-matrix of $\mathbf{H}^H$ by striking $\mathbf{h}_i$ out of $\mathbf{H}^H$, i.e.,

$$\mathbf{H}_i^H=[\mathbf{h}_1,...,\mathbf{h}_{i-1},\mathbf{h}_{i+1},...,\mathbf{h}_K],$$

and

$$\mathcal{P}_i^{\perp}=\mathbf{I}-\mathbf{H}_i^H(\mathbf{H}_i\mathbf{H}_i^H)^{-1}\mathbf{H}_i$$

is the orthogonal projection of the subspace spanned by $\mathbf{H}_i^H$. Since $\mathbf{h}_i$ is i.i.d Rayleigh fading as described in (2), we have $\|\mathbf{h}_i\|^2/(\frac{\sigma_i^2}{2})\sim\mathcal{X}_{2N}^2$. Moreover, due to the properties of projection matrix, we also have

$$\|\mathcal{P}_i^{\perp}\mathbf{h}_i\|^2/(\frac{\sigma_i^2}{2})\sim\mathcal{X}_{2(N-K+1)}^2.$$

Then, by using Lemma 1, we can calculate the diversity order of user i as follows

$$\begin{aligned}d_{ZF}(i)&=\lim_{\rho\to+\infty}-\frac{\log Pr\{P_i\geq\rho\}}{\log\rho}\\ &=\lim_{\rho\to+\infty}-\frac{\log Pr\{\frac{r_i}{\|\mathcal{P}_i^{\perp}\mathbf{h}_i\|^2}\geq\rho\}}{\log\rho}\\ &=N-K+1.\end{aligned} \quad (34)$$

Secondly, we focus on the diversity order achieved by the proposed precoding algorithm. If the strong user m and the weak user n are perfectly paired, by applying (21), we can calculate the individual power consumption as follows.

Firstly, the power consumption of the strong user m can be written as

$$\begin{aligned}P_m&=\|\mathbf{w}_m\|^2\\ &=\|\alpha_m((1+r_n)\mathbf{e}_m-r_n\mathbf{e}_n^H\mathbf{e}_m\mathbf{e}_n)\|^2\\ &=\frac{r_m}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2}\frac{(1+r_n\sin^2\theta)(1+r_n)-r_n\cos^2\theta}{(1+r_n\sin^2\theta)^2},\end{aligned} \quad (35)$$

Hence, we can bound P_m by

$$\frac{r_m}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2}\leq P_m\leq\frac{r_m(1+r_n)}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2}, \quad (36)$$

Note that

$$\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2/(\frac{\sigma_m^2}{2})\sim X_{2(N-K+2)}^2,$$

by taking advantage of Lemma 1, we have

$$\begin{aligned} d_{Pro}(m) &= \lim_{\rho\to+\infty} -\frac{\log Pr\{P_m \geq \rho\}}{\log \rho} \\ &= N-K+2. \end{aligned} \tag{37}$$

Secondly, the power consumption of the weak user n can be written as

$$\begin{aligned} P_n &= \|\mathbf{w}_n\|^2 = \alpha_n^2 \\ &= \frac{r_n}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_n\|^2} + \frac{r_m}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2}\frac{r_n\cos^2\theta}{(1+r_n\sin^2\theta)^2}. \end{aligned} \tag{38}$$

Hence, we can bound P_n by

$$\frac{r_n}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_n\|^2} \leq P_n \leq \frac{r_n}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_n\|^2} + \frac{r_n r_m}{\|\mathcal{P}_{m,n}^{\perp}\mathbf{h}_m\|^2} \tag{39}$$

By taking advantage of Lemmas 1 and 2, we have

$$\begin{aligned} d_{Pro}(n) &= \lim_{\rho\to+\infty} -\frac{\log Pr\{P_n \geq \rho\}}{\log \rho} \\ &= N-K+2. \end{aligned} \tag{40}$$

Finally, by combining (37) and (40), we can obtain

$$d_{Pro}(i) = N-K+2, \tag{41}$$

if user i $(i=1,2,...,K)$ is perfectly paired, and the proof is completed. □

Remark 1: It has been revealed in both [14] and [15] that the quasi-degradation probability is high for i.i.d Rayleigh distributed channels, especially in a heterogenous environment. Moreover, by adopting the proposed user pairing algorithms, this probability can be further increased. In other words, most of the users can be paired perfectly after performing the proposed user pairing algorithm. Therefore, the overall diversity achieved by the proposed beamforming algorithms is close to $N-K+2$, which can be validated in our simulation results. On the other hand, the overall diversity is related to the number of user pairs having quasi-degraded channels. That means, to obtain a good diversity performance, an efficient user pairing algorithm should maximize the number of user pairs having quasi-degraded channels, which is also the goal of the proposed PBPA/IBPA.

To derive the total power consumption for the proposed beamforming algorithms (PH-NOMA in combination with PBPA/IBPA) and conventional ZFBF, we first define some user sets. Note that the total user set is

$$\mathcal{K} = \{1,2,...,K\}.$$

The strong user set is defined as

$$\mathcal{K}_s = \big\{m \big| \text{users m and n are paired with quasi-degraded channels, and } \sigma_m^2 \geq \sigma_n^2\big\}.$$

Similarly, the weak user set is defined as

$$\mathcal{K}_w = \big\{n \big| \text{users m and n are paired with quasi-degraded channels, and } \sigma_m^2 \geq \sigma_n^2\big\}.$$

Then, the set of the users who cannot be paired with quasi-degraded channels is $\mathcal{K}/\mathcal{K}_s/\mathcal{K}_w$.

Theorem 2. *The expectation of the total consumption for conventional ZFBF is*

$$P_{ZF} = \frac{1}{N-K}\sum_{k=1}^{K}\frac{r_i}{\sigma_i^2},$$

while the expectation of the total consumption for the proposed beamforming algorithms can be written as

$$\begin{aligned} P_{Pro} =& \frac{1}{N-K+1}\Big(\sum_{i\in\mathcal{K}_w}\frac{r_i}{\sigma_i^2} + \sum_{j\in\mathcal{K}_s}\frac{r_j}{\sigma_j^2}A_i\Big) \\ &+ \frac{1}{N-K}\sum_{l\in\mathcal{K}/\mathcal{K}_s/\mathcal{K}_w}\frac{r_l}{\sigma_l^2}, \end{aligned} \tag{42}$$

where

$$A_i = (1+r_i)(N-K+1)\sum_{k=0}^{\infty}(-1)^k\frac{r_i^k}{k+N-K+1}.$$

Proof: We first derive the expectation of the total power consumption for conventional ZFBF. As is derived in (33), the total power consumption can be calculated as

$$\begin{aligned} P_{ZF} &= \mathbb{E}\Big\{\sum_{i=1}^{K}\frac{r_i}{\|\mathcal{P}_i^{\perp}\mathbf{h}_i\|^2}\Big\} \\ &= \sum_{i=1}^{K}\frac{r_i}{\frac{1}{2}\sigma_i^2}\mathop{\mathbb{E}}_{x\sim X^2_{2(N-K+1)}}\Big\{\frac{1}{x}\Big\} \\ &= \sum_{i=1}^{K}\frac{r_i}{\frac{1}{2}\sigma_i^2}\frac{1}{2(N-K+1-1)} \\ &= \sum_{i=1}^{K}\frac{r_i}{\sigma_i^2}\frac{1}{N-K}. \end{aligned} \tag{43}$$

Then, we derive the expectation of the total power consumption for the proposed beamforming algorithms. Note that if users j and i are paired with quasi-degraded channels ($\sigma_j^2 \geq \sigma_i^2$), according to (35) and (38), the sum power consumption can be written as

$$P_i + P_j = \frac{r_i}{\|\mathcal{P}_{i,j}^{\perp}\mathbf{h}_i\|^2} + \frac{r_j}{\|\mathcal{P}_{i,j}^{\perp}\mathbf{h}_j\|^2}\frac{1+r_i}{1+r_i\sin^2\theta},$$

where θ is the angle between $\mathcal{P}_{i,j}^{\perp}\mathbf{h}_j$ and $\mathcal{P}_{i,j}^{\perp}\mathbf{h}_i$. According to Lemma 3 in [15], $u=\cos^2\theta$ follows Beta distribution with parameter $(1, N-K+1)$, i.e., the pdf of u is

$$f_u(x) = (N-K)(1-x)^{N-K-1}.$$

Consequently, the expectation of total power consumption of users j and i can be written as

$$\mathcal{E}\Big\{P_i+P_j\Big\} = \frac{1}{N-K+1}\Big(\frac{r_i}{\sigma_i^2} + \frac{r_j}{\sigma_j^2}A\Big), \tag{44}$$

where A can be calculated as

$$\begin{aligned} A_i &= \int_0^1 \frac{f_u(x)(1+r_i)}{1+r_i(1-x)}dx \\ &= (1+r_i)(N-k+1)\sum_{t=0}^{\infty}(-1)^t\frac{r_i^t}{t+N-K+1}. \end{aligned} \tag{45}$$

On the other hand, if user l is paired with non-quasi-degraded channels, according to the concept of PH-NOMA precoding algorithm, ZFBF signal is transmitted, and the expectation of user l is the same as that of conventional ZFBF. Mathematically, the expectation of power consumption of user l is

$$\mathcal{E}\Big\{P_l\Big\} = \frac{1}{N-K}\frac{r_l}{\sigma_l^2}. \tag{46}$$

By combining (44) and (46), the expectation of the total power consumption can be straightforwardly obtain, and the proof is completed. □

Remark 2: Note that the power consumption of a paired users with quasi-degraded channels is minimized, i.e.,

$$\frac{1}{N-K+1}\Big(\frac{r_i}{\sigma_i^2}+\frac{r_j}{\sigma_j^2}A_i\Big) \leq \frac{1}{N-K}\Big(\frac{r_i}{\sigma_i^2}+\frac{r_j}{\sigma_j^2}\Big).$$

Therefore, the total transmission power for the proposed beamforming algorithms can be minimized if all the users are perfectly paired. By observing the equation in (42), we can conclude that, the more the number of user pairs with quasi-degraded channels is, the smaller P_{Pro} is. Therefore, to obtain a minimal total power consumption, an efficient user pairing algorithm should maximize the number of user pairs with quasi-degraded channels, which is also the goal of PBPA/IBPA.

VI. Simulation Results

In this section, the computer simulation results are given to validate the efficiency of the proposed beamforming algorithms: PH-NOMA in combination with PBPA/IBPA. The performance of the conventional ZFBF [4] is simulated as benchmarking. Several user pairing algorithms are also simulated as comparisons, including the Greedy Algorithm (GA), Correlation based user pairing algorithm (Corr) [13], Random user pairing algorithm (Ran), and the Sequential User Pairing Algorithm (SUPA) [15]. Throughout our simulations, the system model described in II(A) is used. The path loss exponent is set as $\alpha = 3$, the parameter R, d_0 are set to be $R = 10, d_0 = 1$, respectively. The individual rate is optimized according to the max-min problem, for fixed total transmission power.

In Figs. 1 and 2, the total power consumption and sum rate performance versus the number of users K are illustrated for different user pairing algorithms, respectively. Firstly, it can be observed that PH-NOMA in combination with various user pairing algorithms can enhance the performance in comparison of that for conventional multi-user beamforming ZFBF. Secondly, different user pairing algorithms can result in different performance. Among these user pairing algorithms, it is observed that a greedy algorithm cannot work efficiently and only has a similar performance as that of the random pairing algorithm, since the performance loss caused by weak users can be deteriorated in a greedy algorithm. The correlation based user pairing algorithm [13] does not obtain satisfactory performance either. Moreover, SUPA for time/frequency division system can not obtain good performance here, since the effect of beamforming is not taken into consideration. Finally, the proposed algorithms PBPA/IBPA realize significant performance enhancement compared to other user pairing algorithms. Note that although the complexity of IBPA is only a fraction of that of PBPA, the performance achieved by both algorithms is nearly the same.

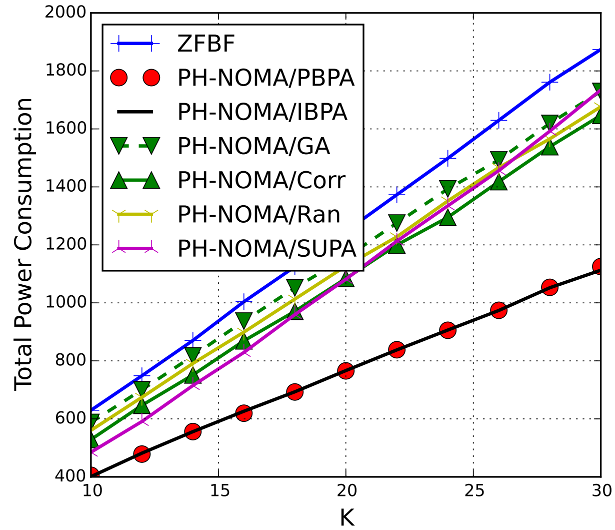

Fig. 1. Total power consumption versus the number of users K, with $N = K + 2$ and $r = [1, 1, ..., 1]$.

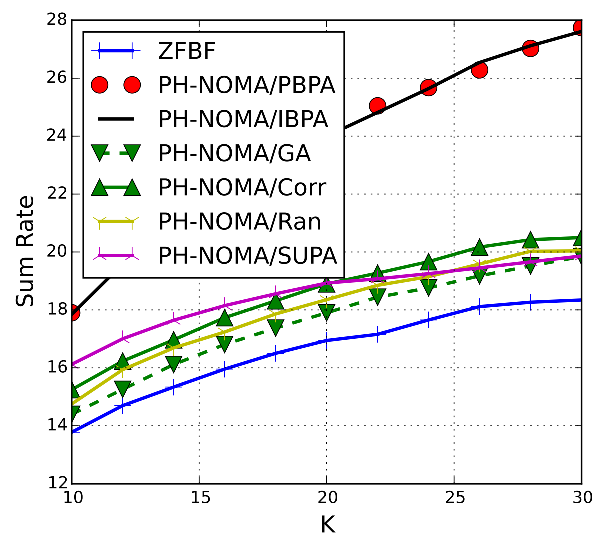

Fig. 2. Sum rate performance versus the number of users K, with $N = 32$.

In Fig. 3, the sum rate performance versus the number of antennas at the BS N is plotted. It is observed that PH-NOMA with PBPA/IBPA realizes the best performance. It is also worth noting that the performance gap between all the NOMA-based multi-user beamforming and conventional ZFBF algorithms diminishes as N becomes large. In Fig. 4, the sum rate performance versus SNR is shown. It is observed that the proposed PH-NOMA with PBPA/IBPA scheme has about 2.1dB gain compared to conventional ZFBF, while the PH-NOMA with other user pairing algorithms only have limited performance gain.

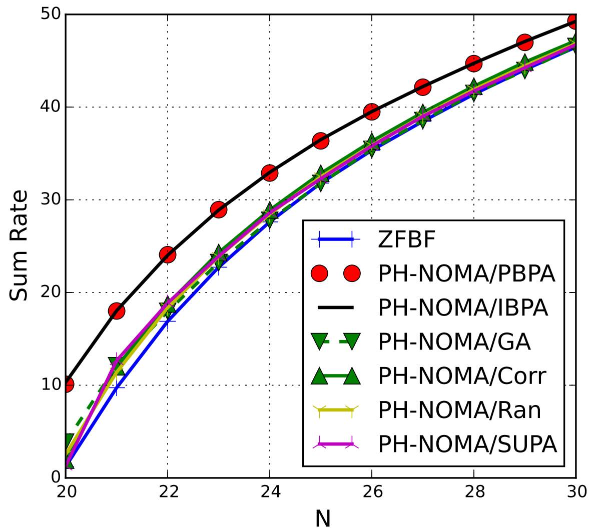

Fig. 3. Sum rate performance versus the number of antennas at the BS N, with $K = 20$.

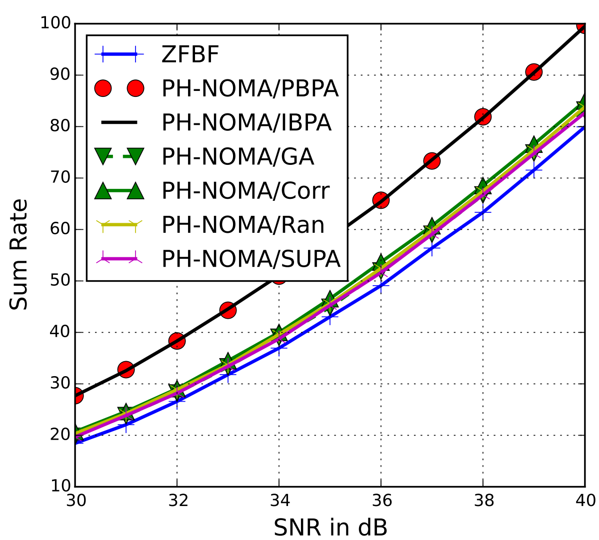

Fig. 4. Sum rate performance versus SNR, with $N = 32, K = 30$.

In Fig. 5, the outage performance is illustrated as a function of SNR for different user pairing algorithms. It can be observed that the proposed beamforming algorithms (PH-NOMA beamforming in combination with PBPA/IBPA) have a significant improvement compared to conventional ZFBF, while other user pairing algorithms only have limited improvement, which is consistent with the analysis results. More specifically, as revealed by Theorem 1, the overall diversity

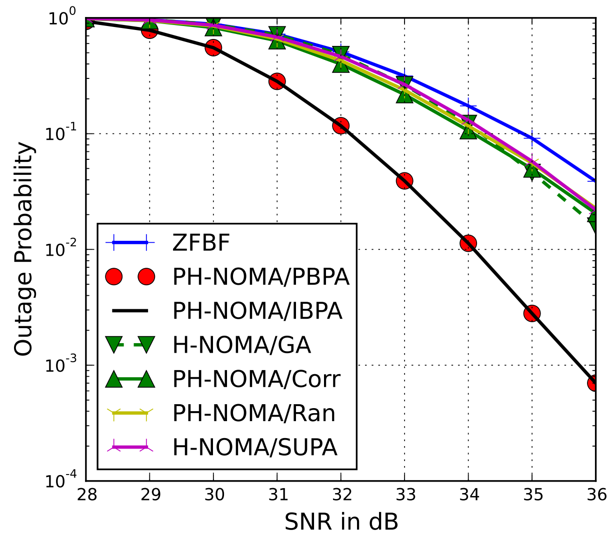

Fig. 5. Outage probability versus SNR, with $N = 32, K = 30$.

order is related to the number of user pairs with quasi-degraded channels. The proposed PBPA/IBPA maximizes the number of user pairs with quasi-degraded channels, thus has a significant performance enhancement compared to the other pairing algorithms.

VII. Conclusion

In this paper, we have considered the design of the downlink beamforming algorithms for multi-user hybrid NOMA systems. Firstly, we have briefly revisited the concept of quasi-degradation and the H-NOMA precoding algorithm. Secondly, we have introduced PH-NOMA beamforming for intra cluster beamforming, to minimize both inter and intra cluster interference. Thirdly, we have proposed two algorithms PBPA and IBPA for user pairing, by applying the properties of quasi-degradation. Consequently, low-complexity beamforming algorithms have been obtained, by combining PH-NOMA beamforming and PBPA/IBPA. Moreover, performance analysis has also been given, to show the superiority of the proposed beamforming scheme. Finally, simulation results have been conducted to demonstrate the efficiency of the proposed beamforming algorithms and the correctness of the performance analysis.

References

[1] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[2] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge,UK: Cambridge university press, 2005.

[3] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, 2007.

[4] Q. H. Spencer, M. Haardt *et al.*, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, 2004.

[5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J.Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[6] 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, 2015.

[7] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, Jun. 2013.

[8] 3rd Generation Partnership Project (3GPP), *Study on downlink multiuser superposition transmission for LTE*, Mar. 2015.

[9] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, 2014.

[10] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, to be published, 2015.

[11] B. Kim, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Military Communications Conference (MILCOM)*, 2013, pp. 1278–1283.

[12] Z. Ding, F. Adachi, and H. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, 2016.

[13] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, 2015.

[14] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, June 2016.

[15] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," submitted to *IEEE Trans. Signal Process.*, available at http://home.ustc.edu.cn/%7ezhiyong/H-NOMA.pdf.

[16] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, 2004.

[17] A. Bourdoux and N. Khaled, "Joint TX-RX optimisation for MIMO-SDMA based on a-space constraint," in *Proc. IEEE 56th Vehicular Technology Conference (VTC)*, vol. 1, Sep. 2002, pp. 171–174.