A High-Diversity Transceiver Design for MISO Broadcast Channels

Junyeong Seo, *Member, IEEE*, Youngchul Sung, *Senior Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

***Abstract*— In this paper, the outage behavior and diversity order of the mixture transceiver architecture for multiple-input single-output broadcast channels are analyzed. The mixture scheme groups users with closely-aligned channels and applies superposition coding and successive interference cancellation decoding to each group composed of users with closely-aligned channels, while applying zero-forcing beamforming across semi-orthogonal user groups. In order to enable such analysis, closed-form lower bounds on the achievable rates of a general multiple-input single-output broadcast channel with superposition coding and successive interference cancellation are newly derived. By employing channel-adaptive user grouping and proper power allocation, which ensures that the channel subspaces of user groups have an angle larger than a certain threshold, it is shown that the mixture transceiver architecture achieves full diversity order in multiple-input single-output broadcast channels and opportunistically increases the multiplexing gain while achieving full diversity order. Furthermore, the achieved full diversity order is the same as that of the single-user maximal ratio transmit beamforming. Hence, the mixture scheme can provide reliable communication under channel fading for ultra-reliable low latency communication. The numerical results validate our analysis and show the outage superiority of the mixture scheme over conventional transceiver designs for multiple-input single-output broadcast channels.**

***Index Terms*— Multiple-input single-output broadcast channels, outage probability, diversity order, successive interference cancellation, user grouping, mixture reception.**

I. Introduction

THE multiple-input single-output (MISO) broadcast channel (BC) model is an important channel model which captures modern cellular downlink communication in which a base station (BS) equipped with multiple transmit antennas simultaneously serves multiple receivers each equipped with a single receive antenna by using the spatial domain. Due to its importance it has been investigated extensively for more than a decade and major current wireless communication standards support MISO BC downlink communication [2]–[5]. It is known that the capacity region of a MISO BC can be achieved by dirty paper coding (DPC) [2]. However, because of the unavailability of practical dirty paper codes, simple linear downlink beamforming such as zero-forcing (ZF) beamforming is widely considered and used in practice [4], [6]. Although such simple linear beamforming is not a capacity-achieving scheme, it can yield good performance when it is combined with multi-user diversity and user scheduling [3], [4], [7]–[9]. That is, when the number of users in the cell is sufficiently large as compared to the number N of transmit antennas, the BS can select N users with nearly orthogonal channel vectors so that linear ZF downlink beamforming is sufficient. However, such orthogonality-based user scheduling for linear downlink beamforming may not be appropriate in certain cases. One example is the case in which the number of transmit antennas is large under rich scattering environments since it is difficult to simultaneously select multiple users with roughly orthogonal channels in this case [8]–[10]. Another emerging important example is *ultra-reliable low-latency communication (URLLC)* for fast machine-type communication in 5G. In the case of URLLC, such orthogonality-based user scheduling induces extra delay in communication since the users requiring immediate data transmission may not have channel vectors nearly orthogonal to each other or to other on-going overlapping data users under spatial multiplexing. In both examples, the channel vectors of the scheduled users are not guaranteed to be nearly orthogonal and the performance of linear ZF beamforming can be severely degraded.

Recently, inspired by the usefulness of superposition coding and successive interference cancellation (SIC) decoding in non-orthogonal multiple access (NOMA) [11], [12], a mixture (or hybrid) transceiver architecture was considered for MISO BCs to overcome the drawback of the fully linear ZF downlink beamforming based on *user grouping* and *mixture of linear and non-linear reception* [13], [14]. The basic idea of the mixture transceiver architecture is as follows. Under the assumption of independent and identically distributed (i.i.d.) realization of K channel vectors in K-user MISO downlink, if the channel vectors of some users are closely aligned, the performance of ZF beamforming is severely degraded. However, if we group the closely-aligned users and apply superposition coding and non-linear SIC decoding for each closely-aligned user group while applying ZF beamforming

Manuscript received June 29, 2018; revised October 24, 2018 and January 22, 2019; accepted March 10, 2019. Date of publication March 25, 2019; date of current version May 8, 2019. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2013R1A1A2A10060852, and in part by the NSF Award under Grant CCF-1526780. This work is a part of [1]. The associate editor coordinating the review of this paper and approving it for publication was J. Yuan. *(Corresponding author: Youngchul Sung.)*

J. Seo is with Samsung Electronics, Hwasung 18448, South Korea (e-mail: j89.seo@samsung.com).

Y. Sung is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: ysung@ee.kaist.ac.kr).

H. Jafarkhani is with the Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA 92697 USA (e-mail: hamidj@uci.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2019.2905609

across roughly-orthogonal user-groups, the performance degradation by the full ZF beamforming can be alleviated. Preliminary study on such user grouping and mixture transreception was performed on the two-user grouping case, where intra-group rate analysis is rather simple [13], [14]. In [13], Pareto-optimal beam design is considered for the two-user grouping case, the beam vectors and corresponding rates are numerically obtained, and the performance of the mixture scheme is compared with the full ZF beamforming numerically. In [14], under the assumption of two users in each group, closed-form beam vectors are obtained to minimize the transmit power under a signal-to-interference-plus-noise ratio (SINR) constraint for each user based on quasi-degradation, and it was shown that such a mixture architecture based on two-user grouping increases the diversity order by one as compared to the conventional ZF downlink beamforming. Although such two-user grouping for the mixture transceiver architecture is tractable, it has limitation in diversity order improvement. (The related idea of hierarchical coding and user grouping was discussed in the dual scenario of multiple access channel in [15], and the idea of user grouping and inter-group zero forcing was also considered in [16] using the intra-group processing of a classical spatial multiplexing from a capacity perspective.)

In this paper, we fully generalize the mixture transceiver architecture for general MISO BCs. The contributions of the paper are summarized as follows:

- In order to enable analysis of the outage probability and diversity order of the mixture transceiver architecture, we derive a new lower bound on the achievable rate of each user in closed form in terms of each user's channel norm for a MISO BC with superposition coding and SIC decoding with an arbitrary number of users.
- We propose a channel-adaptive user grouping method which ensures a condition for the channel subspace angle property for the constructed user groups and a power allocation method necessary for achievability of full diversity order.
- Combining the newly derived achievable rate result and the property of the proposed adaptive user grouping method, we derive the diversity order of the mixture transceiver architecture, and show that *the mixture transceiver architecture achieves full diversity order in MISO BCs, which is the same as that of the single-user maximal ratio transmit (MRT) beamforming*, and furthermore it opportunistically increases multiplexing gain.
- We further investigate the related issues such as diversity-and-multiplexing trade-off associated with the mixture scheme, impact of imperfect channel state information (CSI), etc.

Notations: Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^T$ and $\mathrm{Tr}(\mathbf{A})$ indicate the complex conjugate, conjugate transpose, transpose and trace of $\mathbf{A}$, respectively, and $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}^{\perp}(\mathbf{A})$ denotes the linear subspace spanned by the columns of $\mathbf{A}$ and its orthogonal complement, respectively. $\Pi_{\mathbf{A}}$ and $\Pi_{\mathbf{A}}^{\perp}$ are the projection matrices to $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}^{\perp}(\mathbf{A})$, respectively. $[\mathbf{a}_1, \cdots, \mathbf{a}_n]$ denotes the matrix composed of column vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$. $||\mathbf{a}||$ represents the 2-norm of vector $\mathbf{a}$. $\mathbf{I}_n$ denotes the identity matrix of size n (the subscript is omitted when unnecessary). $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \Sigma)$ means that random vector $\mathbf{x}$ is circularly-symmetric complex Gaussian distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

II. The Channel Model and Preliminaries

A. The Channel Model

In this paper, we consider a Gaussian MISO BC composed of a transmitter with N transmit antennas and K single-antenna users (i.e., receivers), where the number of users is less than or equal to the number of transmit antennas, i.e., $K \leq N$. The received signal y_k at the k-th user is given by

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \quad k = 1, 2, \cdots, K, \tag{1}$$

where $\mathbf{x}$ is the $N \times 1$ transmit signal vector at the transmitter with the total transmit power $P_t = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$, n_k is the additive white Gaussian noise (AWGN) at the k-th user, i.e., $n_k \sim \mathcal{CN}(0, \sigma^2)$ with σ^2 set to 1 for simplicity, and $\mathbf{h}_k$ is the $N \times 1$ (conjugated) channel vector from the transmitter to the k-th user following independent Rayleigh fading, i.e.,

$$\mathbf{h}_k = [h_{k1}, h_{k2}, \cdots, h_{kN}]^T \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0}, 2\mathbf{I}). \tag{2}$$

Here, we set $2\mathbf{I}$ as the covariance matrix for convenience so that both real and imaginary components of each element of $\mathbf{h}_k$ have variance one and thus $||\mathbf{h}_k||^2$ has the chi-square distribution of degrees of freedom $2N$. Different scaling can be absorbed into the transmit power. Concatenating all the received signals $y_1, \cdots, y_K$, we can write the matrix model for the received signals as

$$\mathbf{y} = \mathbf{H}^H \mathbf{x} + \mathbf{n}, \tag{3}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_K]^T$, $\mathbf{n} = [n_1, n_2, \cdots, n_K]^T$, and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]$. We assume that the CSI $\mathbf{H}$ is available at the transmitter. Due to the assumption of $K \leq N$, the $K \times N$ overall channel matrix $\mathbf{H}^H$ is a fat matrix and hence it is right-invertible so that conventional ZF transmit beamforming is feasible. Design of the signal vector $\mathbf{x}$ and receiver processing based on $\{y_1, y_2, \cdots, y_K\}$ will be explained in the subsequent sections.

B. Preliminaries: Reliability and Diversity Order

Channel fading is inherent in wireless communication, and communication reliability under channel fading is dependent on the diversity order of the communication channel. Consider the well-known single-user MRT beamforming with multiple transmit antennas. The corresponding channel model is given by the channel model (1) with only a single user, i.e., $K = 1$. For MRT beamforming, we have $\mathbf{x} = \frac{\mathbf{h}_1}{||\mathbf{h}_1||}\sqrt{p_1}s_1$ with $\mathbb{E}\{|s_1|^2\} = 1$. The resulting equivalent single-input single-output (SISO) channel and rate are respectively given by

$$y_1 = ||\mathbf{h}_1||\sqrt{p_1}s_1 + n_1 \quad \text{and} \quad R_1 = \log(1 + ||\mathbf{h}_1||^2 \mathrm{SNR}), \tag{4}$$

where $\mathrm{SNR} := \frac{p_1}{\sigma^2}$, and the probability density function (pdf) of $||\mathbf{h}_1||^2 = |h_{11}|^2 + \cdots + |h_{1N}|^2$ is given by the chi-square

distribution with degree of freedom $2N$ since it is the sum of the squares of $2N$ standard normal random variables:

$$f_{||\mathbf{h}_1||^2}(x) = \frac{1}{2^N(N-1)!}x^{N-1}e^{-x/2} = \frac{1}{2^N(N-1)!}x^{N-1}+o(x^{N-1}), \quad \text{as } x \to 0, \qquad (5)$$

where $o(\cdot)$ is the small o notation. Communication outage is defined as the event that the channel cannot support a given target rate R^{th}, and the corresponding outage probability is given by $P_{out} = \Pr\{R_1 < R^{th}\}$ [17]. Then, the diversity of order of the channel is defined as [17]

$$D := -\lim_{\text{SNR}\to\infty} \frac{\log P_{out}}{\log \text{SNR}}. \qquad (6)$$

In the single-user MRT beamforming case, the outage probability is given by $P_{out} = \Pr\left\{||\mathbf{h}_1||^2 \le \frac{2^{R^{th}}-1}{\text{SNR}}\right\} \approx \frac{(2^{R^{th}}-1)^N}{2^N N!\text{SNR}^N}$ [17], and hence the diversity order in this case is N. That is, the outage probability decays as SNR^{-N}, as SNR increases. Note that in the case of a Rayleigh-fading SISO channel with a single transmit antenna $N = 1$, the pdf (5) reduces to $f_{|h_{11}|^2}(x) = \frac{1}{2}e^{-x/2}$, and the diversity order reduces to one. Hence, MRT beamforming with N transmit antennas increases the diversity order by N times as compared to the SISO case.

Now, consider the general Gaussian MISO BC (1) with ZF downlink beamforming for $K = N$. In the ZF beamforming case, the overall transmit signal $\mathbf{x}$ is given by $\mathbf{x} = \mathbf{w}_1^{ZF}\sqrt{p_1}s_1 + \cdots + \mathbf{w}_K^{ZF}\sqrt{p_K}s_K$, where $\mathbf{w}_k^{ZF}$ and s_k are the ZF beam vector and data symbol for the k-th user with $||\mathbf{w}_k^{ZF}||^2 = 1$ and $\mathbb{E}\{|s_k|^2\} = 1$, respectively. Here, the ZF beam vector $\mathbf{w}_k^{ZF}$ lies in $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$ so that $\mathbf{h}_i^H\mathbf{w}_k^{ZF} = 0$ for all $i \neq k$. Then, the resulting SISO channel for the k-th user is given by

$$y_k = \mathbf{h}_k^H\mathbf{w}_k^{ZF}\sqrt{p_k}s_k + n_k. \qquad (7)$$

In the case of independent Rayleigh fading, the channel vector $\mathbf{h}_k$ and the remaining $\{\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K\}$ are independent. Hence, the one-dimensional subspace $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$ is also independent of $\mathbf{h}_k$, and hence $\mathbf{h}_k$ is circularly-symmetric Gaussian distributed over $\mathbb{C}^N$ with respect to a reference direction of $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$. Therefore, taking the inner product between $\mathbf{h}_k$ and the unit-norm vector $\mathbf{w}_k^{ZF} \in \mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$ is equivalent to taking only one component out of N complex Gaussian components, and thus $|\mathbf{h}_k^H\mathbf{w}_k^{ZF}|^2$ has the same pdf as $f_{|h_{11}|^2}(x) = \frac{1}{2}e^{-x/2}$. Hence, the corresponding diversity order for the k-th user is simply one for all k [14] as in the SISO Rayleigh fading channel. Thus, ZF downlink beamforming for MISO BCs loses the diversity gain possibly obtainable from multiple transmit antennas.

Note that if $\mathbf{h}_k$ is perfectly orthogonal to $\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K$, then $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$ is perfectly aligned with $\mathbf{h}_k$ and hence in this case we have $\mathbf{h}_k^H\mathbf{w}_k^{ZF} = ||\mathbf{h}_k||$. In this case, the resulting SISO channel for the k-th user is the same as that of the MRT beamforming single-user channel in (4). Furthermore, suppose that the angle between $\mathbf{h}_k$ and one-dimensional subspace $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$ is equal to or less than a certain fixed threshold α. Then, we have $|\mathbf{h}_k^H\mathbf{w}_k^{ZF}| \ge ||\mathbf{h}_k||\cos\alpha$. Since $\cos\alpha$ is a constant, the pdf of $|\mathbf{h}_k^H\mathbf{w}_k^{ZF}|^2$ is a certain scaled version of that of $||\mathbf{h}_k||^2$ (the meaning of this statement will become clear in later sections), and the outage behavior for the k-th user in this case should be the same as that of the MRT single-user case as SNR increases without bound. Reflecting this, one can recognize that the degradation of diversity order of ZF beamforming for a MISO BC with independent channel fading results from the uncontrolled and arbitrary angle between $\mathbf{h}_k$ and $\mathcal{C}^{\perp}([\mathbf{h}_1, \cdots, \mathbf{h}_{k-1}, \mathbf{h}_{k+1}, \cdots, \mathbf{h}_K])$.

III. The Mixture Transceiver Architecture

In this section, motivated by the discussion in the previous section, we consider the mixture transceiver architecture for Gaussian MISO BCs [13], [14] in order to overcome the diversity drawback of ZF downlink beamforming. The mixture architecture is based on user grouping and mixture of linear ZF and non-linear SIC reception. First, user grouping is performed to group users with closely-aligned channel vectors. Then, superposition coding and SIC are applied to the users with closely-aligned channel vectors in each group, whereas ZF beamforming is applied across groups. In order to fully enhance the diversity order of the resulting individual user channel, we generalize the mixture architecture by adopting adaptive user grouping, which yields channel-dependent groups and enforces the angle between the subspace of each group and the orthogonal complement of the union of all other groups' subspaces to be less than a certain threshold so that inter-group ZF beamforming does not harm the overall diversity order.

From here on, we explain the mixture transceiver architecture with the proposed user grouping method in detail. We consider the MISO BC explained in Section II-A as our channel model. We assume the following for our transceiver architecture:

A.1 (User Grouping)

First, we group the K users into N_g groups. The constructed groups are denoted by the sets $\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_{N_g}$ such that $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$ for $i \neq j$ and $\bigcup_{j=1}^{N_g}\mathcal{G}_j = \{1, 2, \cdots, K\}$. User grouping is adaptive in the sense that the number of groups can vary and the number of members in each group can vary from one to K, depending on the channels such that $\sum_{j=1}^{N_g}|\mathcal{G}_j| = K$. The constructed groups satisfy a certain subspace angle property in order to apply inter-group ZF beamforming without degrading the diversity order. The detailed method for user grouping will be presented in Section III-B.

A.2 (Inter-Group Beamforming)

With the constructed groups, in order to control inter-group interference, we apply ZF beamforming across the constructed

groups. With this inter-group ZF beamforming, the inter-group interference across the groups is zero.

A.3 (Intra-Group Processing: Superposition Coding and SIC)

With the constructed groups, for intra-group processing we apply superposition coding and SIC decoding to each and every group with more than one user.

Under the aforementioned transceiver architecture, the transmit signal $\mathbf{x}$ of the transmitter can be expressed as

$$\mathbf{x} = \sum_{j=1}^{N_g} \Pi^{(j)} \sum_{i \in \mathcal{G}_j} \sqrt{p_i^{(j)}} \mathbf{w}_i^{(j)} s_i^{(j)}, \tag{8}$$

where $s_i^{(j)}$ is the transmit symbol from $\mathcal{CN}(0,1)$ for User i in group $\mathcal{G}_j$, $\mathbf{w}_i^{(j)}$ is the $N \times 1$ intra-group beamforming vector for User i in group $\mathcal{G}_j$ out of the feasible set $\widetilde{\mathcal{W}} := \{\mathbf{w} \mid \|\Pi^{(j)}\mathbf{w}\|^2 \leq 1\}$, $p_i^{(j)}$ is the power assigned to User i in group $\mathcal{G}_j$, and $\Pi^{(j)}$ is the inter-group ZF projection matrix for group $\mathcal{G}_j$. We assume that the total transmit power P_t is divided such that $|\mathcal{G}_j| \times P_t/K$ is allocated to group $\mathcal{G}_j$. Then, from (1) the received signal at User i in group $\mathcal{G}_j$ can be written as

$$\begin{aligned} y_i^{(j)} &= \mathbf{h}_i^{(j)H} \left(\Pi^{(j)} \sum_{i \in \mathcal{G}_j} \sqrt{p_i^{(j)}} \mathbf{w}_i^{(j)} s_i^{(j)} \right) + n_i^{(j)} \\ &\overset{(a)}{=} \left(\Pi^{(j)} \mathbf{h}_i^{(j)} \right)^H \left(\sum_{i \in \mathcal{G}_j} \sqrt{p_i^{(j)}} \mathbf{w}_i^{(j)} s_i^{(j)} \right) + n_i^{(j)}, \end{aligned} \tag{9}$$

where $\mathbf{h}_i^{(j)}$ is the $N \times 1$ channel vector between the transmitter and User i in group $\mathcal{G}_j$, and $n_i^{(j)} \sim \mathcal{CN}(0,1)$ is the AWGN at User i in group $\mathcal{G}_j$ (here, the single user index k in (1) is properly mapped to the two indices: intra-group user index i and group index j). The inter-group ZF projection matrix $\Pi^{(j)}$ is given by $\Pi^{(j)} = \Pi^{\perp}_{\tilde{\mathbf{H}}_j}$, where $\tilde{\mathbf{H}}_j$ is the matrix composed of all channel vectors except the channel vectors of the users in group $\mathcal{G}_j$, i.e.,

$$\begin{aligned} \tilde{\mathbf{H}}_j := [&\mathbf{h}_1^{(1)} \cdots \mathbf{h}_{|\mathcal{G}_1|}^{(1)}, \cdots, \mathbf{h}_1^{(j-1)} \cdots \mathbf{h}_{|\mathcal{G}_{j-1}|}^{(j-1)}, \mathbf{h}_1^{(j+1)}, \\ &\cdots \mathbf{h}_{|\mathcal{G}_{j+1}|}^{(j+1)}, \cdots, \mathbf{h}_1^{(N_g)}, \cdots, \mathbf{h}_{|\mathcal{G}_{N_g}|}^{(N_g)}]. \end{aligned} \tag{10}$$

Due to the inter-group ZF beamforming, there is no inter-group interference in (9), and the property of an orthogonal projection matrix, $(\Pi^{\perp}_{\tilde{\mathbf{H}}_j})^H = \Pi^{\perp}_{\tilde{\mathbf{H}}_j}$, is used in Step (a) in (9).

A. Intra-Group Beam Design and the Corresponding Rates

In this subsection, we consider intra-group beam vector design for the mixture transceiver architecture and analyze the achievable rates of the intra-group processing. First, consider each group $\mathcal{G}_j$ with one user. In this case, the received signal (9) reduces to

$$y_1^{(j)} = \left(\Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_1^{(j)} \right)^H \sqrt{p_1^{(j)}} \mathbf{w}_1^{(j)} s_1^{(j)} + n_1^{(j)}, \quad |\mathcal{G}_j| = 1, \tag{11}$$

and the design of the intra-group beam vector $\mathbf{w}_1^{(j)}$ is simple. The optimal intra-group beam vector $\mathbf{w}_1^{(j)*}$ is the MRT beam matched to the projected effective channel vector $\Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_1^{(j)}$, i.e., $\sqrt{p_1^{(j)}} \mathbf{w}_1^{(j)*} = \sqrt{P_t/K}\, \Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_1^{(j)} / \| \Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_1^{(j)} \|$. In this case, the optimal beam vector is equivalent to the ZF-beamforming vector with power P_t/K.

Next, consider the intra-group beam design for each group with more than one user. As aforementioned, we apply superposition coding and SIC in this case. Suppose that group $\mathcal{G}_j$ consists of L users ($L > 1$). Then, with the group index (j) omitted for convenience, the received signal for User i, $i = 1, \cdots, L$, in group $\mathcal{G}_j$ is given by

$$y_i = \mathbf{g}_i^H \left(\sum_{i=1}^{L} \sqrt{p_i} \mathbf{w}_i s_i \right) + n_i, \quad i = 1 \cdots, L, \tag{12}$$

where $\sum_{i=1}^{L} p_i \leq P$ with P being the total group power allocated to group $\mathcal{G}_j$ (i.e., $P = L \times P_t/K$), and $\mathbf{g}_i$ is the projected effective channel of User i given by

$$\mathbf{g}_i = \Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_i^{(j)}, \quad i = 1, \cdots, L. \tag{13}$$

We assume that the intra-group beam vector $\mathbf{w}_i$ is designed based on the projected effective channels $\mathbf{g}_1, \cdots, \mathbf{g}_L$. Then, the feasible set for intra-group beam vector $\mathbf{w}_i$ is given by $\mathcal{W} := \{\mathbf{w} \mid \|\mathbf{w}\|^2 \leq 1\}$ from the fact that the beam design space for $\mathbf{w}_i$ is the linear subspace spanned by $\{\mathbf{g}_1, \cdots, \mathbf{g}_L\}$. (The beam component not in the subspace spanned by $\{\mathbf{g}_1, \cdots, \mathbf{g}_L\}$ does not affect the signal or the interference. Hence, it just wastes power.) Since $\mathbf{w}_i \in \mathcal{C}([\mathbf{g}_1, \cdots, \mathbf{g}_L])$, we have $\mathbf{w}_i \in \mathcal{C}^{\perp}(\tilde{\mathbf{H}}_j)$ by (13) and hence for the actual beam power constraint $\|\Pi^{(j)}\mathbf{w}_i\|^2 \leq 1$, we have $\|\Pi^{(j)}\mathbf{w}_i\|^2 = \|\Pi^{\perp}_{\tilde{\mathbf{H}}_j}\mathbf{w}_i\|^2 = \|\mathbf{w}_i\|^2 \leq 1$ in this case. So, we have the feasible set $\mathcal{W}$ for $\mathbf{w}_i$.

Note that with inter-group ZF beamforming, the intra-group signal model is separated from group to group based on the projected effective channels, and the system model (12) is a conventional MISO BC with L-user superposition coding beamforming. For superposition coding and SIC, we assume that the in-group users are ordered according to their channel norms as $\|\mathbf{g}_1\|^2 \geq \|\mathbf{g}_2\|^2 \geq \cdots \geq \|\mathbf{g}_L\|^2$. With this assumption, SIC at the in-group receivers is applied such that User i decodes and cancels the interference from Users $L, L-1, \cdots, i+1$ sequentially.[1] (Note that since User i has a better channel than Users $L, L-1, \cdots, i+1$, User i can decode the messages intended for Users $L, L-1, \cdots, i+1$.) Then, the rates of the in-group users can be expressed as

$$\begin{aligned} R_1 &= \log_2 \left(1 + p_1 |\mathbf{g}_1^H \mathbf{w}_1|^2 \right), \\ R_i &= \log_2 \left(1 + \min \left\{ \mathrm{SINR}_1^i, \cdots, \mathrm{SINR}_i^i \right\} \right), \quad i = 2, \cdots, L, \end{aligned} \tag{14}$$

[1]The considered decoding order may not be optimal if we consider the design of $\{\mathbf{w}_i\}$ but is sufficient for our purpose of analytic derivation of the diversity order of the mixture transceiver architecture.

where SINR_j^i is the SINR when User j decodes the message intended for User i, given by

$$\mathrm{SINR}_j^i = \frac{p_i |\mathbf{g}_j^H \mathbf{w}_i|^2}{\sum_{m=1}^{i-1} p_m |\mathbf{g}_j^H \mathbf{w}_m|^2 + 1} \tag{15}$$

The achievable rate region $\mathcal{R}$ of the MISO BC with superposition coding and SIC decoding is defined as the union of achievable rate-tuples:

$$\mathcal{R} := \bigcup_{\substack{(\mathbf{w}_1,\cdots,\mathbf{w}_L)\in\mathcal{W}^L \\ (p_1,\cdots,p_L)|p_i>0,\forall i,\ \sum_{i=1}^L p_i = P}} (R_1, R_2 \cdots, R_L), \tag{16}$$

where $(R_1, \cdots, R_L)$ is from (14). The Pareto boundary of the region $\mathcal{R}$ is the outer boundary of $\mathcal{R}$ and can be obtained by maximizing R_L for each feasible target rate-tuple $(R_1^*, \cdots, R_{L-1}^*)$. The maximization problem for given $(R_1^*, \cdots, R_{L-1}^*)$ can be solved by a convex programming approach based on reformulation [18] and the convex concave procedure (CCP) [19]. However, difficulty lies in knowing the feasible target rate-tuple set for the MISO BC with superposition coding and SIC since the rates depend on the beam vectors and channel vectors of all in-group users, although some induction approach for this was proposed in [20]. The difficulty to find the feasible rate tuple for Users $1, \cdots, L-1$ can be circumvented by formulating the problem as weighted sum rate maximization based on the rate-profile approach [21]. However, the existing algorithms for the Pareto-optimal design problem numerically provide rates based on numerically obtained beam vectors. Hence, these existing design algorithms do not provide closed-form rate expressions for general MISO BCs with superposition coding and SIC decoding which are necessary for our analytical derivation of the diversity order. In order to obtain desired closed-form expressions for the achievable rates of the MISO BC with superposition coding and SIC decoding, we consider beam design under the following constraint:

$$\begin{aligned} &\mathbf{w}_1 = \mathbf{w}_2 = \cdots = \mathbf{w}_L = \mathbf{w}, \quad ||\mathbf{w}||^2 \le 1, \ \text{i.e. } \mathbf{w} \in \mathcal{W} \\ &p_i = \delta_i P, \quad i = 1, 2, \cdots, L, \end{aligned} \tag{17}$$

where $(\delta_1, \cdots, \delta_L)$ is a power ratio-tuple out of the feasible power ratio-tuple set $\mathcal{D} := \{(\delta_1, \cdots, \delta_L) \mid \delta_i \ge 0 \ \forall i, \ \sum_{i=1}^L \delta_i = 1\}$. Here, δ_i is the ratio of the total group power P to the power allocated to User i, i.e., $p_i = \delta_i P$ is assigned to User i. Note that the constraint (17) satisfies the original beam design constraint in (16). Based on the restricted constraint (17), the following proposition provides simple closed-form lower bounds on the achievable rates for the MISO BC with superposition coding and SIC:

Proposition 1: In the MISO BC (12) with L users adopting superposition coding and SIC decoding with channel vectors $\mathbf{g}_1, \cdots, \mathbf{g}_L$ with ordering $\|\mathbf{g}_1\|^2 \ge \|\mathbf{g}_2\|^2 \ge \cdots \ge \|\mathbf{g}_L\|^2$ and total group power P, for an arbitrary given power ratio-tuple $(\delta_1, \cdots, \delta_L)$ out of the feasible power ratio-tuple set $\mathcal{D}$, the achievable rates $(R_1, R_2, \cdots, R_L)$ are lower bounded as

$$R_1 \ge \log_2 \left(1 + \frac{1}{c}\delta_1 \|\mathbf{g}_1\|^2 P\right) \tag{18}$$

$$R_i \ge \log_2 \left(1 + \frac{\delta_i}{\sum_{m=1}^{i-1} \delta_m} \frac{1}{1 + \left(\frac{1}{c}\|\mathbf{g}_i\|^2 \sum_{m=1}^{i-1} \delta_m P\right)^{-1}}\right) \quad i = 2, \cdots, L, \tag{19}$$

where the constant c is given by

$$c = \begin{cases} L & \text{if, } L \le 3, \\ 8L^2 & \text{if, } L > 3. \end{cases} \tag{20}$$

Proof: See Appendix A.

Note that the power ratio-tuple set $\mathcal{D}$ does not depend on the beam vectors and the channel vectors, and it is just a simplex. Thus, the rate lower bounds (18) and (19) with sweeping $(\delta_1, \cdots, \delta_L)$ within $\mathcal{D}$ yield an inner region of the achievable rate region $\mathcal{R}$ defined in (16). The key point in the derived lower bounds (18) and (19) on the achievable rates $(R_1, \cdots, R_L)$ is that *the lower bound on the rate R_i of User i in the superposition-and-SIC group is expressed only in terms of User i's channel norm square $||\mathbf{g}_i||^2$ and the power distribution factors* $(\delta_1, \cdots, \delta_L)$. This enables us to analyze the distribution of R_i via the distribution of $||\mathbf{g}_i||^2$ and to derive the diversity order of the mixture scheme in Section IV.

B. Adaptive User Grouping

Now, we consider user grouping, which should be done properly for good diversity performance of the mixture transceiver architecture. Since we apply inter-group ZF beamforming, a level of orthogonality across the constructed groups is required to guarantee high reliability, as discussed in Section II-B. Note that the channel orthogonality among the users within a group is not required since superposition coding and SIC are applied to the users in each group. There can exist many user grouping methods that guarantee certain orthogonality among the constructed groups. In this section, we provide one example for such user grouping. The main difference between our user grouping method and several previous user grouping methods proposed for NOMA [14], [20], [22] is that the number of groups and the number of members in each group are not predetermined and the angle between the channel subspaces of any two user groups is not less than a certain threshold in our user grouping method, whereas the number of groups and the number of members in each group are predetermined and fixed for the previous methods [14], [20], [22]. This angle property is necessary for derivation of the diversity order of the mixture architecture.

To measure the orthogonality across groups, we define a new subspace angle metric $\theta(\cdot,\cdot)$, which captures the angle between the subspaces $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ spanned by the columns of matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$\theta(\mathbf{A},\mathbf{B}) := \begin{cases} \max(\{\phi(\mathbf{A}, \mathbf{b}_i), \forall\ i\} \cup \{\phi(\mathbf{B}, \mathbf{a}_j), \forall\ j\}) \\ \quad \text{if } \mathbf{A} \text{ and } \mathbf{B} \text{ are non-empty matrices,} \\ 0 \quad \text{if } \mathbf{A} \text{ or } \mathbf{B} \text{ is an empty matrix,} \end{cases} \tag{21}$$

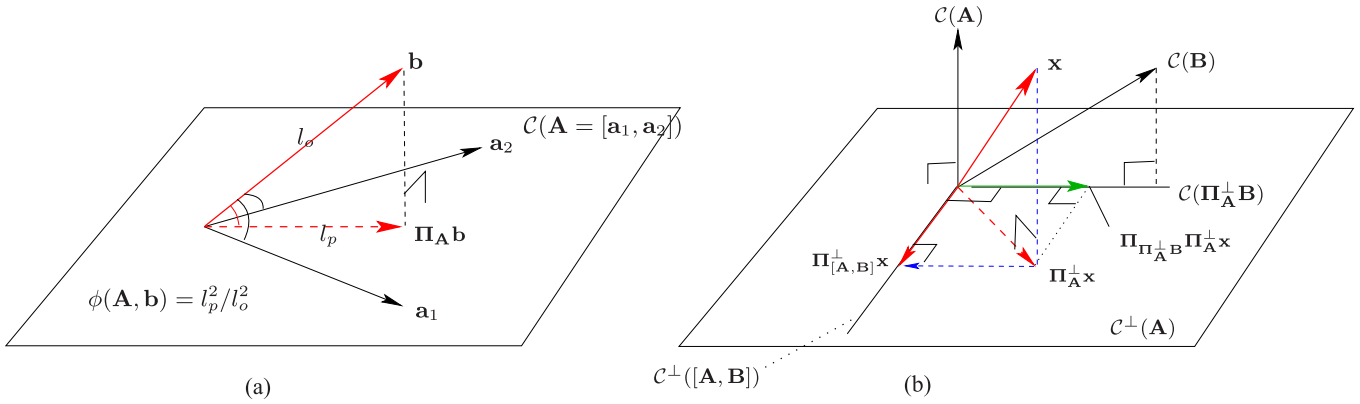


Fig. 1. (a) An illustration of $\phi(\mathbf{A},\mathbf{b})$ and (b) An illustration of sequential orthogonal projection.

where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$, $\mathbf{b}_j$ is the j-th column of $\mathbf{B}$, and $\phi(\cdot,\cdot)$ is another newly-defined angle metric which captures the angle between the vector $\mathbf{b}$ and the subspace $\mathcal{C}(\mathbf{A})$, defined as

$$\phi(\mathbf{A},\mathbf{b}) := \frac{\|\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{b}\|^2}{\|\mathbf{b}\|^2}. \tag{22}$$

In the case that $\mathbf{A}=[\mathbf{a}]$ is a vector, ϕ reduces to the square of the angle cosine of two vectors $\mathbf{a}$ and $\mathbf{b}$:

$$\phi(\mathbf{a},\mathbf{b}) = \frac{|\mathbf{a}^H\mathbf{b}|^2}{\|\mathbf{a}\|^2\|\mathbf{b}\|^2} = \cos^2\angle(\mathbf{a},\mathbf{b}) \in [0,1]. \tag{23}$$

When $\mathbf{B}=[\mathbf{b}]$ in (21) is a vector, $\theta(\mathbf{A},\mathbf{b})$ simply reduces to $\phi(\mathbf{A},\mathbf{b})$ because $\phi(\mathbf{A},\mathbf{b}) \geq \phi(\mathbf{b},\mathbf{a}_j)$ for all j, i.e., the angle between $\mathbf{b}$ and $\mathcal{C}(\mathbf{A})$ is smaller than or equal to the angle between $\mathbf{b}$ and individual column $\mathbf{a}_j$ of $\mathbf{A}$, as illustrated in Fig. 1(a). When $\theta=0$, two subspaces $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are mutually orthogonal. When $\theta=1$, on the other hand, there exists either at least a column of $\mathbf{A}$ contained in $\mathcal{C}(\mathbf{B})$ or at least a column of $\mathbf{B}$ contained in $\mathcal{C}(\mathbf{A})$, and the two subspaces $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are not separated.

The proposed user grouping algorithm based on $\theta(\cdot,\cdot)$ is presented in Algorithm 1. Before explaining the algorithm, we introduce a useful lemma regarding sequential orthogonal projection necessary to explain the algorithm.

Lemma 1: For a vector $\mathbf{x}$ and matrices $\mathbf{A}$ and $\mathbf{B}$ such that $[\mathbf{A},\mathbf{B}]$ is a tall matrix, the following equality holds: $\Pi_{[\mathbf{A},\mathbf{B}]}^{\perp}\mathbf{x} = (\mathbf{I}-\Pi_{\Pi_{\mathbf{A}}^{\perp}\mathbf{B}})\Pi_{\mathbf{A}}^{\perp}\mathbf{x} = \Pi_{\mathbf{A}}^{\perp}\mathbf{x} - \Pi_{\mathbf{A}}^{\perp}\mathbf{B}[(\Pi_{\mathbf{A}}^{\perp}\mathbf{B})^H\Pi_{\mathbf{A}}^{\perp}\mathbf{B}]^{-1}(\Pi_{\mathbf{A}}^{\perp}\mathbf{B})^H\Pi_{\mathbf{A}}^{\perp}\mathbf{x}$.

Proof: See Appendix B.

Lemma 1 states that the projection of $\mathbf{x}$ onto the orthogonal space of $\mathcal{C}([\mathbf{A},\mathbf{B}])$ can be accomplished in two steps first by projecting $\mathbf{x}$ onto the orthogonal space of $\mathcal{C}(\mathbf{A})$ and then by projecting this projected vector onto the orthogonal space of $\mathcal{C}(\Pi_{\mathbf{A}}^{\perp}\mathbf{B})$ (not $\mathcal{C}(\mathbf{B})$), as illustrated in Fig. 1(b). By successively applying Lemma 1, we can obtain $\Pi_{[\mathbf{A}_1,\mathbf{A}_2,\cdots,\mathbf{A}_n]}^{\perp}\mathbf{x}$ in a successive manner, where $[\mathbf{A}_1,\cdots,\mathbf{A}_n]$ is a tall matrix. That is, we first project $\mathbf{x}$ onto $\mathcal{C}^{\perp}(\mathbf{A}_1)$ to obtain $\Pi_{\mathbf{A}_1}^{\perp}\mathbf{x}$, and project the subspace matrices $\mathbf{A}_2,\mathbf{A}_3,\cdots,\mathbf{A}_n$ onto $\mathcal{C}^{\perp}(\mathbf{A}_1)$ to obtain $\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2, \Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_3,\cdots,\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_n$. Then, we project $\Pi_{\mathbf{A}_1}^{\perp}\mathbf{x}$ onto $\mathcal{C}^{\perp}(\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2)$ to obtain $(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{x}$, and also project $\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_3,\cdots,\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_n$ onto $\mathcal{C}^{\perp}(\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2)$ to obtain $(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_3,\cdots,(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_n$. Then, we project $(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{x}$ onto $\mathcal{C}^{\perp}((\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_3)$, and project the remaining subspace matrices $(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_4,\cdots,(\mathbf{I}-\Pi_{\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_2})\Pi_{\mathbf{A}_1}^{\perp}\mathbf{A}_n$ correspondingly. We continue this process until step n is reached. Then, this gives us $\Pi_{[\mathbf{A}_1,\mathbf{A}_2,\cdots,\mathbf{A}_n]}^{\perp}\mathbf{x}$.

Algorithm 1 tries to find single-user groups first (line 6). If the algorithm cannot find any single-user group further, it increases the number of users in group to two (lines 17 and 18), and tries to find two-user groups. It continues this process until n_g becomes K (line 9). Suppose that no group is found up to $n_g=K-1$. Then, at $n_g=K$, one argument in $\theta(\cdot,\cdot)$ in (24) becomes an empty matrix, θ becomes zero by the definition (21), and hence the condition (24) is satisfied. Thus, in this case the whole set $\{1,2,\cdots,K\}$ becomes a single group. Let us explain Algorithm 1 by using a specific example below:

Example 1: Suppose that initial $\mathcal{K}=\{1,2,3,4,5,6,7\}$ and suppose that initially user 1 satisfies $\theta([\mathbf{h}_2,\cdots,\mathbf{h}_7],\mathbf{h}_1) \leq \theta^{th}$. Then, we update $\mathcal{G}_1=\{1\}$ (line 13) and $\mathcal{K}=\{2,3,4,5,6,7\}$ (line 14), and project the channel vectors $\mathbf{h}_2,\cdots,\mathbf{h}_7$ onto $\mathcal{C}^{\perp}([\mathbf{h}_1])$ to obtain the projected channel vectors $\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2,\cdots,\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7$ (line 15). Next, suppose that $\theta([\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_3,\cdots,\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7],\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2) \leq \theta^{th}$. (Note that at this point we compute $\theta(\cdot,\cdot)$ using the *projected* channels (lines 15 and 16).) Then, we update $\mathcal{G}_2=\{2\}$ and $\mathcal{K}=\{3,4,5,6,7\}$ and project $\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_3,\cdots,\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7$ onto $\mathcal{C}^{\perp}(\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2)$ to obtain the further projected channels $(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_3,\cdots,(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7$. Now, suppose that we cannot find a single-user group further and that at $n_g=2$ only one pair of users $\{3,4\}$ satisfies $\theta([(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_5,\cdots,(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7],[(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_3,(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_4]) \leq \theta^{th}$. Then, we update $\mathcal{G}_3=\{3,4\}$ and $\mathcal{K}=\{5,6,7\}$, and the further projected channels for users $\{5,6,7\}$ are obtained by projecting $(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_5,\cdots,(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_7$ onto $\mathcal{C}^{\perp}([(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_3,(\mathbf{I}-\Pi_{\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_2})\Pi_{\mathbf{h}_1}^{\perp}\mathbf{h}_4])$. These final projected channels for users $\{5,6,7\}$ are the same as the

Algorithm 1 The Proposed User Grouping Algorithm

1: **Initialization**:
2: A threshold value $\theta^{th} \in (0, 1)$ is given.
3: Initially set $\mathbf{f}_1 \cdots, \mathbf{f}_K$ as the actual channel vectors $\mathbf{h}_1, \cdots, \mathbf{h}_K$ of the K users.
4: Set $\mathcal{K} \leftarrow \{1, \cdots, K\}$ (initial candidate set)
5: Set $i_g \leftarrow 0$ (group index)
6: Set $n_g \leftarrow 1$ (number of users in group)
7: Set $\mathbf{F}_{\mathcal{K}} \leftarrow [\mathbf{f}_1, \cdots, \mathbf{f}_K]$.
8: **Execution**:
9: **While** $n_g \leq K$
10: Find a group of users $\{u_1^*, \cdots, u_{n_g}^*\}$ with cardinality n_g such that $\mathcal{C}(\mathbf{F}_{\{u_1^*, \cdots, u_{n_g}^*\}})$ and $\mathcal{C}(\mathbf{F}_{\mathcal{K}\backslash\{u_1^*, \cdots, u_{n_g}^*\}})$ satisfy

$$\theta(\mathbf{F}_{\mathcal{K}\backslash\{u_1^*, \cdots, u_{n_g}^*\}}, \mathbf{F}_{\{u_1^*, \cdots, u_{n_g}^*\}}) \leq \theta^{th}. \tag{24}$$

11: **If** we find such a group of users $\{u_1^*, \cdots, u_{n_g}^*\}$,
12: $i_g \leftarrow i_g + 1$ (increase the group index by one).
13: $\mathcal{G}_{i_g} \leftarrow \{u_1^*, \cdots, u_{n_g}^*\}$ (construct one group).
14: $\mathcal{K} \leftarrow \mathcal{K}\backslash\{u_1^*, \cdots, u_{n_g}^*\}$ (update $\mathcal{K}$ by removing the selected users from the candidate set).
15: Update the vector $\mathbf{f}_u$ as $\mathbf{f}_u \leftarrow \left(\mathbf{I} - \mathbf{F}_{\mathcal{G}_{i_g}}(\mathbf{F}_{\mathcal{G}_{i_g}}^H \mathbf{F}_{\mathcal{G}_{i_g}})^{-1}\mathbf{F}_{\mathcal{G}_{i_g}}^H\right)\mathbf{f}_u$, $\forall u \in$ updated $\mathcal{K}$
16: Construct new $\mathbf{F}_{\mathcal{K}}$ with the updated $\mathbf{f}_u, \forall u \in$ updated $\mathcal{K}$.
17: **Else**
18: $n_g \leftarrow n_g + 1$
19: **Endif**
20: **Endwhile**
21: $N_g \leftarrow i_g$.
22: (Throughout the algorithm, $\mathbf{F}_{\mathcal{S}}$ means the submatrix of current $\mathbf{F}_{\mathcal{K}}$ composed of {current $\mathbf{f}_u, \forall u \in \mathcal{S} \subset \mathcal{K}$}.)

ZF projected channels $\Pi^{\perp}_{[\mathbf{h}_1,\cdots,\mathbf{h}_4]}\mathbf{h}_5, \cdots, \Pi^{\perp}_{[\mathbf{h}_1,\cdots,\mathbf{h}_4]}\mathbf{h}_7$ by Lemma 1. At the next iteration, n_g becomes 3 since we assumed that there is no further two-user group; one argument of $\theta(\cdot,\cdot)$ becomes an empty matrix since $\mathcal{K} = \{5, 6, 7\}$ and $n_g = 3$; hence $\mathcal{G}_4 = \{5, 6, 7\}$; no user is left in the candidate set $\mathcal{K}$ after update (line 14); no further channel projection in line 15 occurs since updated $\mathcal{K} = \emptyset$; and the algorithm stops.

Now, let us consider the norm property of the projected ZF channels associated with the constructed groups in the example, which is the key aspect of the proposed user grouping algorithm. Consider user 1 in firstly-constructed $\mathcal{G}_1$. Since $\theta([\mathbf{h}_2, \cdots, \mathbf{h}_7], \mathbf{h}_1) \leq \theta^{th}$, by the definition of $\theta(\cdot,\cdot)$ in (21), we have

$$\phi(\tilde{\mathbf{H}}_1 = [\mathbf{h}_2, \cdots, \mathbf{h}_7], \mathbf{h}_1) = \frac{\|\tilde{\mathbf{H}}_1(\tilde{\mathbf{H}}_1^H\tilde{\mathbf{H}}_1)^{-1}\tilde{\mathbf{H}}_1^H\mathbf{h}_1\|^2}{\|\mathbf{h}_1\|^2} \leq \theta^{th} \tag{25}$$

Hence, we have

$$\begin{aligned} ||\Pi^{\perp}_{\tilde{\mathbf{H}}_1}\mathbf{h}_1||^2 &= ||(\mathbf{I} - \tilde{\mathbf{H}}_1(\tilde{\mathbf{H}}_1^H\tilde{\mathbf{H}}_1)^{-1}\tilde{\mathbf{H}}_1^H)\mathbf{h}_1||^2 \\ &\overset{(a)}{=} (1 - \phi(\tilde{\mathbf{H}}_1, \mathbf{h}_1))||\mathbf{h}_1||^2 \\ &\geq (1 - \theta^{th})||\mathbf{h}_1||^2, \end{aligned}$$

where Step (a) is valid by the Pythagorean theorem. Next, consider the norm of the ZF effective channel for User 2 in $\mathcal{G}_2$. Due to the construction of $\mathcal{G}_1$ based on (25), $\mathbf{h}_1$ and $\mathbf{h}_2$ satisfy the following:

$$\begin{aligned} ||\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_2||^2 &= (1 - \phi(\mathbf{h}_1, \mathbf{h}_2))||\mathbf{h}_2||^2 \\ &\geq (1 - \phi(\tilde{\mathbf{H}}_1, \mathbf{h}_1))||\mathbf{h}_2||^2, \quad \text{since } \tilde{\mathbf{H}}_1 \text{ includes } \mathbf{h}_2 \\ &\geq (1 - \theta^{th})||\mathbf{h}_2||^2. \end{aligned} \tag{26}$$

By Lemma 1, $\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_3,\cdots,\mathbf{h}_7]}\mathbf{h}_2$ can be obtained by sequential orthogonal projection as

$$\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_3,\cdots,\mathbf{h}_7]}\mathbf{h}_2 = (\mathbf{I} - \Pi_{[\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_3,\cdots,\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_7]})\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_2,$$

but $\mathcal{G}_2$ was constructed such that $\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_2$ and $[\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_3, \cdots, \Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_7]$ satisfy the threshold θ^{th} requirement. Combining this fact and (26), we have

$$||\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_3,\cdots,\mathbf{h}_7]}\mathbf{h}_2||^2 \geq (1 - \theta^{th})^2||\mathbf{h}_2||^2.$$

Then, consider User 3 in $\mathcal{G}_3 = \{3, 4\}$. (The same applies to User 4 in $\mathcal{G}_3$.) By Lemma 1, we have

$$\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_3 = (\mathbf{I} - \Pi_{\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_2})\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_3 \tag{27}$$

$$\begin{aligned} \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_5,\mathbf{h}_6,\mathbf{h}_7]}\mathbf{h}_3 = (\mathbf{I} - \Pi_{[\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_5, \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_6, \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_7]}) \\ \cdot \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_3. \end{aligned} \tag{28}$$

In (27), $\mathcal{G}_1 = \{1\}$ was constructed such that $\mathbf{h}_1$ and $\mathbf{h}_3$ satisfy the angle constraint, and $\mathcal{G}_2 = \{2\}$ was constructed such that $\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_2$ and $\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_3$ satisfy the angle constraint. Hence, we have $||\Pi^{\perp}_{\mathbf{h}_1}\mathbf{h}_3||^2 \geq (1 - \theta^{th})^2||\mathbf{h}_3||^2$. Furthermore, in (28), $\mathcal{G}_3 = \{3, 4\}$ was constructed such that $[\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_3, \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_4]$ and the remaining $[\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_5, \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_6, \Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2]}\mathbf{h}_7]$ satisfy the angle constraint. Combining these facts, we have

$$||\Pi^{\perp}_{[\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_5,\mathbf{h}_6,\mathbf{h}_7]}\mathbf{h}_k||^2 \geq (1 - \theta^{th})^3||\mathbf{h}_k||^2, \quad k = 3, 4. \tag{29}$$

Finally, consider the norm of the ZF effective channels $\Pi^{\perp}_{[\mathbf{h}_1,\cdots,\mathbf{h}_4]}\mathbf{h}_5, \cdots, \Pi^{\perp}_{[\mathbf{h}_1,\cdots,\mathbf{h}_4]}\mathbf{h}_7$ of the last group $\mathcal{G}_4 = \{5, 6, 7\}$. These vectors are obtained by three sequential orthogonal projections based on Lemma 1, and at each projection stage the threshold θ^{th} was kept for group splitting. Hence, we have

$$||\Pi^{\perp}_{[\mathbf{h}_1,\cdots,\mathbf{h}_4]}\mathbf{h}_k||^2 \geq (1 - \theta^{th})^3||\mathbf{h}_k||^2, \quad k = 5, 6, 7.$$

Note that in general the proposed user grouping algorithm satisfies the following norm reduction property for the ZF effective channels:

$$||\mathbf{g}_i^{(j)}||^2 = ||\Pi^{\perp}_{\tilde{\mathbf{H}}_j}\mathbf{h}_i^{(j)}||^2 \geq (1 - \theta^{th})^{N_g - 1}||\mathbf{h}_i^{(j)}||^2, \tag{30}$$

where $\Pi^{\perp}_{\tilde{\mathbf{H}}_j}$ is the ZF projection matrix for group $\mathcal{G}_j$, $\mathbf{h}_i^{(j)}$ is the channel vector of User i in group $\mathcal{G}_j$, and N_g is the number of constructed groups, which is bounded by K. Since the number of antennas N and the number of users K ($\leq N$) are fixed in our MISO BC model with superposition coding and SIC, the lower bound $(1 - \theta^{th})^{K-1} \in (0, 1)$ of $(1 - \theta^{th})^{N_g - 1}$ is a constant.

Now, let us define a useful quantity for further exposition: We define the degrees of freedom of a fading channel $\mathbf{h}$ as

$$d := \lim_{x\to 0} \frac{\log \Pr(\|\mathbf{h}\|^2 \le x)}{\log x}. \tag{31}$$

This quantity captures the behavior of the tail probability of the random variable $||\mathbf{h}||^2$ in its lower tail, and the degrees of freedom d for $\mathbf{h}$ means that $\Pr(\|\mathbf{h}\|^2 \le x)$ behaves as $x^d + o(x^d)$, as $x \to 0$. This quantity is directly related to the diversity order of the SISO communication channel with the channel gain $||\mathbf{h}||$. For example, a Rayleigh fading channel $\mathbf{h} \sim \mathcal{C}(\mathbf{0}, 2\mathbf{I}_N)$ has the degrees of freedom N since

$$\Pr(||\mathbf{h}||^2 \le x) = \int_0^x f_{\|\mathbf{h}\|^2}(z)dz = \frac{1}{2^N N!}x^N + o(x^N) \tag{32}$$

as $x \to 0$, and $\lim_{x\to 0} \frac{\log \Pr(\|\mathbf{h}\|^2 \le x)}{\log x} = N$, where $f_{\|\mathbf{h}\|^2}(z)dz$ is given in (5). Finally, we provide the main statement of this subsection regarding the degrees of freedom of the ZF effective channels associated with the proposed grouping method in the following proposition:

Proposition 2: With the mixture transceiver architecture and the user grouping method in Algorithm 1, the projected effective channel $\mathbf{g}_j^{(i)} = \Pi^{\perp}_{\tilde{\mathbf{H}}_j} \mathbf{h}_i^{(j)}$ in (9) resulting from inter-group ZF beamforming has the same degrees of freedom as the original channel $\mathbf{h}_i^{(j)}$, i.e., $\lim_{x\to 0} \frac{\log \Pr(\|\mathbf{g}_i^{(j)}\|^2 \le x)}{\log x} = \lim_{x\to 0} \frac{\log \Pr(\|\mathbf{h}_i^{(j)}\|^2 \le x)}{\log x}, \forall\, i, j$.

Proof: See Appendix C.

1) Complexity of Algorithm 1: Note that in the worst case the number of group searches is given by $K + \binom{K}{2} + \binom{K}{3} + \cdots + \binom{K}{K}$, which scales as $K^{K/2}$. For each group search, we need to compute $\theta(\cdot,\cdot)$ in (24), which requires inversion of $K \times K$ matrices in the worst case (see (24) and the term $(\mathbf{A}^H\mathbf{A})^{-1}$ in (22)). Thus, Algorithm 1 is not scalable for large K. Nevertheless, the algorithm is devised to prove the diversity-order optimality of the mixture architecture in this paper. Invention of more efficient user grouping algorithms for the mixture architecture for MISO BCs is a future work. For one possible idea for polynomial complexity, the interested reader is referred to [23].

2) SIC Complexity: Since in the proposed adaptive user grouping, each group can have one to K members, it is required that each receiver be able to handle SIC of $K-1$ users in the worst case. SIC for a general number of users has been investigated extensively for code-division multiple access systems [24].

IV. Outage Analysis and Diversity Order of the Mixture Scheme

In this section, we present our main result regarding the diversity order of the mixture transceiver architecture for MISO BCs.

Theorem 1: For the Gaussian MISO BC with N transmit antennas and K single-antenna users with independent Rayleigh fading described in Section II-A, let the channels be ordered as $\|\mathbf{h}_1\|^2 \ge \|\mathbf{h}_2\|^2 \cdots \ge \|\mathbf{h}_K\|^2$ and let the k-th user be the user with the k-th largest channel norm. Then, the diversity order for the k-th user achievable by the mixture transceiver architecture with proper user grouping is given by $D_k = N \times (K-k+1)$. Here, the diversity order is defined as $D_k := \lim_{P_t\to\infty} -\frac{\log Pr\{R_k < R^{th}\}}{\log P_t}$, where R_k is the rate of the k-th user and R^{th} is a rate threshold. Note that $P_t \to \infty$ is equivalent to SNR $= P_t/\sigma^2 \to \infty$ since we set the noise variance $\sigma^2 = 1$ for simplicity.

Proof: For user grouping of the mixture architecture we adopt Algorithm 1. The diversity provided by such a grouping will provide an achievable bound for the diversity as claimed in the theorem. Proof is based on Propositions 1 and 2. In proof, we consider not only the distribution of the channel norm itself but also the order statistics resulting from the channel norm ordering. With the descending channel ordering $\|\mathbf{h}_1\|^2 \ge \|\mathbf{h}_2\|^2 \cdots \ge \|\mathbf{h}_K\|^2$, the pdf of the k-th channel norm square is given by order statistics as

$$f_{\|\mathbf{h}_k\|^2}(x) = \frac{K!}{(k-1)!(K-k)!}[F_{\|\mathbf{h}\|^2}(x)]^{K-k} \times [1 - F_{\|\mathbf{h}\|^2}(x)]^{k-1} f_{\|\mathbf{h}\|^2}(x), \tag{33}$$

where $f_{\|\mathbf{h}\|^2}(\cdot)$ and $F_{\|\mathbf{h}\|^2}$ are the pdf and cumulative distribution function (cdf) of chi-square distribution with $2N$ degree of freedom:

$$f_{\|\mathbf{h}\|^2}(x) = \frac{1}{2^N(N-1)!}x^{N-1}e^{-x/2} = \frac{1}{2^N N!}x^{N-1} + o(x^{N-1}), \quad \text{as } x \to 0 \tag{34}$$

$$F_{\|\mathbf{h}\|^2}(x) = \frac{1}{2^N N!}x^N + o(x^N), \quad \text{as } x \to 0. \tag{35}$$

Hence, we have for the k-th largest channel norm square $||\mathbf{h}_k||^2$

$$f_{||\mathbf{h}_k||^2}(x) = c_k x^{N(K-k+1)-1} + o(x^{N(K-k+1)-1}) \quad \text{as } x \to 0, \tag{36}$$

and thus

$$\lim_{x\to 0} \frac{\log \Pr(\|\mathbf{h}_k\|^2 \le x)}{\log x} = N(K-k+1). \tag{37}$$

The outage probability of the k-th user is expressed as

$$\Pr(R_k < R^{th}) = \sum_{j=1}^{N_g} \left[\Pr(k \in \mathcal{G}_j) \cdot \Pr\left(R_k < R^{th} \middle| k \in \mathcal{G}_j\right)\right] \tag{38}$$

$$= \sum_{j=1}^{N_g} \Big[\Pr(k \in \mathcal{G}_j) \cdot \Big\{\Pr\left(|\mathcal{G}_j| = 1 \mid k \in \mathcal{G}_j\right) \times \Pr\left(R_k < R^{th} \mid |\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right) + \Pr(|\mathcal{G}_j| \ne 1 \mid k \in \mathcal{G}_j) \times \Pr\left(R_k < R^{th} \mid |\mathcal{G}_j| \ne 1, k \in \mathcal{G}_j\right)\Big\}\Big]. \tag{39}$$

A. Lower Bound on the Outage Probability

We obtain a lower bound on the outage probability by considering only the event that the k-th user belongs to a group with cardinality one, i.e., the first term in the RHS of (39).

$$\begin{aligned}\Pr(R_k < R^{th}) &\geq \sum_{j=1}^{N_g} \Pr(k \in \mathcal{G}_j) \cdot \Pr\left(|\mathcal{G}_j| = 1 \mid k \in \mathcal{G}_j\right) \\ &\quad \cdot \Pr\left(R_k < R^{th} \;\middle|\; |\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right) \\ &= \sum_{j=1}^{N_g} \Pr\left(|\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right) \\ &\quad \cdot \Pr\left(R_k < R^{th} \;\middle|\; |\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right) \\ &= \sum_{j=1}^{N_g} \Pr\left(|\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right) \\ &\quad \cdot \Pr\left(\|\Pi^{(j)}\mathbf{h}_k\|^2 < K \cdot (2^{R^{th}} - 1) \cdot P_t^{-1}\right),\end{aligned} \tag{40}$$

where the last equality holds due to the rate $R_k = \log(1 + P_t||\Pi^{(j)}\mathbf{h}_k||^2/K)$ for a single-user group based on (11) and the corresponding optimal beam. Then, we can lower bound $-D_k$ as (41) - (44), shown at the top of the next page. Here, (44) is valid because $||\mathbf{h}_k||^2$ has the channel order $N(K-k+1)$ by (36) and (37); the projected effective channel $\|\Pi^{(j)}\mathbf{h}_k\|^2$ has the same channel order as $||\mathbf{h}_k||^2$ by Proposition 2; and the linear combination of terms with the same order has the same order as each term. Note that $\Pr\left(|\mathcal{G}_j| = 1, k \in \mathcal{G}_j\right)$ depends only on the joint distribution of $(\mathbf{h}_1, \cdots, \mathbf{h}_k)$ for the given user grouping algorithm not on the power P_t.

B. Upper Bound on the Outage Probability

For the upper bound, we include the second term in the RHS of (39) in addition to the first term in the RHS of (39) considered in the lower bound. The second term in the RHS of (39) is given by

$$\sum_{j=1}^{N_g} \Pr(k \in \mathcal{G}_j) \cdot \Pr\left(|\mathcal{G}_j| \neq 1 \mid k \in \mathcal{G}_j\right) \cdot \Pr\left(R_k < R^{th} \mid |\mathcal{G}_j| \neq 1, k \in \mathcal{G}_j\right) \tag{45}$$

$$= \sum_{j=1}^{N_g} \Pr\left(|\mathcal{G}_j| \neq 1, k \in \mathcal{G}_j\right) \cdot \Pr\left(R_k < R^{th} \mid |\mathcal{G}_j| \neq 1, k \in \mathcal{G}_j\right) \tag{46}$$

$$= \sum_{j=1}^{N_g} \sum_{\ell=2}^{K} \Pr\left(|\mathcal{G}_j| = \ell, k \in \mathcal{G}_j\right) \cdot \underbrace{\Pr\left(R_k < R^{th} \mid |\mathcal{G}_j| = \ell, k \in \mathcal{G}_j\right)}_{(a)}. \tag{47}$$

Define the following notations:

$$E_{k,j,i} := \text{Event that the } k\text{-th user is the } i\text{-th largest channel norm user in } \mathcal{G}_j \tag{48}$$

$$P_{k,j,i} := \Pr(E_{k,j,i}). \tag{49}$$

With these notations, the term (a) in (47) can be rewritten as

$$\Pr\left(R_k < R^{th} \;\middle|\; |\mathcal{G}_j| = \ell, k \in \mathcal{G}_j\right) = \sum_{i=1}^{\ell} P_{k,j,i} \cdot \Pr\left(R_k < R^{th} \;\middle|\; |\mathcal{G}_j| = \ell, k \in \mathcal{G}_j, E_{k,j,i}\right), \tag{50}$$

where R_k conditioned on the joint event $(|\mathcal{G}_j| = \ell, k \in \mathcal{G}_j, E_{k,j,i})$ is lower bounded by Proposition 1 as

$$R_k \geq \begin{cases} \log_2(1 + \frac{1}{c}\delta_1^{(j)}\|\Pi^{(j)}\mathbf{h}_k\|^2 \frac{\ell P_t}{K}) & \text{if } i = 1, \\ \log_2(1 + \frac{\delta_i^{(j)}}{\sum_{m=1}^{i-1}\delta_m^{(j)}} \times \frac{1}{1 + \left(\frac{1}{c}\|\Pi^{(j)}\mathbf{h}_k\|^2 \sum_{m=1}^{i-1}\delta_m^{(j)}\frac{\ell P_t}{K}\right)^{-1}}) & \text{if } i = 2 \cdots \ell. \end{cases} \tag{51}$$

where c is given in (20), and $(\delta_1^{(j)}, \delta_2^{(j)}, \cdots \delta_\ell^{(j)})$ is the power ratio-tuple in group $\mathcal{G}_j$, i.e., power $\delta_i^{(j)}\ell P_t/K$ is assigned to User i in group $\mathcal{G}_j$. ($\ell P_t/K$ is the total group power for group $\mathcal{G}_j$ with $|\mathcal{G}_j| = \ell$.) Therefore, the probability (50) is upper bounded as (53) - (55), shown at the top of the next page. In the derivation of (53) - (55), we used the fact that the event that R_k is less than R_{th} implies that its lower bound is less than R_{th}, and the threshold for $\|\Pi^{(j)}\mathbf{h}_k\|^2$ in the second term in (55) is obtained by manipulation of the second term in (54). By Lemma 4 in Appendix D, there always exists a collection of in-group power distribution factors $(\delta_1^{(j)}, \cdots, \delta_\ell^{(j)})$ such that $(\frac{\delta_i^{(j)}}{2^{R^{th}}-1} - \sum_{m=1}^{i-1}\delta_m^{(j)})$ in (55) is strictly positive for all $i = 2, \cdots, \ell$. Set $(\delta_1^{(j)}, \cdots, \delta_\ell^{(j)})$ as one of such collections. Then, each probability term in (55) behaves as $P_t^{-N(K-k+1)}$ as $P_t \to \infty$, since $\|\Pi^{(j)}\mathbf{h}_k\|^2$ has the same degrees of freedom of $N(K-k+1)$ in (36) and (37) as $||\mathbf{h}_k||^2$ by Proposition 2. Hence, their linear combination (53) behaves as $P_t^{-N(K-k+1)}$ as $P_t \to \infty$, and furthermore the term (47) as a linear combination of terms in (53) behaves as $P_t^{-N(K-k+1)}$ as $P_t \to \infty$. Now, by adding (47) and the term in (40), we have the exact outage probability. We already showed that the term in (40) behaves as $P_t^{-N(K-k+1)}$ as $P_t \to \infty$. Furthermore, the upper bound of (47) behaves as $P_t^{-N(K-k+1)}$ as $P_t \to \infty$. Hence, we have $-D_k = \lim_{P_t\to\infty} \frac{\log \Pr(R_k < R^{th})}{\log P_t} \leq -N(K-k+1)$. Combining this upper bound result with the lower bound result, we have

$$N(K-k+1) \leq D_k = -\lim_{P_t\to\infty} \frac{\log \Pr(R_k < R^{th})}{\log P_t} \leq N(K-k+1). \tag{52}$$

■

$$-D_k = \lim_{P_t\to\infty} \frac{\log \Pr(R_k < R^{th})}{\log P_t} \tag{41}$$

$$\geq \lim_{P_t\to\infty} \frac{\log\left(\sum_{j=1}^{N_g}\left[\Pr\left(|\mathcal{G}_j|=1, k\in\mathcal{G}_j\right)\cdot \Pr\left(\|\Pi^{(j)}\mathbf{h}_k\|^2 < K\cdot(2^{R^{th}}-1)\cdot P_t^{-1}\right)\right]\right)}{\log P_t} \tag{42}$$

$$= \lim_{P_t^{-1}\to 0} \frac{\log\left(\sum_{j=1}^{N_g}\left[\Pr\left(|\mathcal{G}_j|=1, k\in\mathcal{G}_j\right)\cdot \Pr\left(\|\Pi^{(j)}\mathbf{h}_k\|^2 < K\cdot(2^{R^{th}}-1)\cdot P_t^{-1}\right)\right]\right)}{-\log P_t^{-1}} \tag{43}$$

$$= -N(K-k+1). \tag{44}$$

$$\sum_{i=1}^{\ell}\left[P_{k,j,i}\cdot \Pr\left(R_k < R^{th} \;\middle|\; |\mathcal{G}_j|=\ell, k\in\mathcal{G}_j, E_{k,j,i}\right)\right] \tag{53}$$

$$\leq P_{k,j,1}\cdot \Pr\left(\log_2\left(1+\frac{1}{c}\delta_1^{(j)}\|\Pi^{(j)}\mathbf{h}_k\|^2\frac{\ell P_t}{K}\right) < R^{th}\right) + \sum_{i=2}^{\ell}\left[P_{k,j,i}\cdot \Pr\left(\log_2\left(1+\frac{\delta_i^{(j)}}{\sum_{m=1}^{i-1}\delta_m^{(j)}}\frac{1}{1+\left(\frac{1}{c}\|\Pi^{(j)}\mathbf{h}_k\|^2\sum_{m=1}^{i-1}\delta_m^{(j)}\ell P_t/K\right)^{-1}}\right) < R^{th}\right)\right] \tag{54}$$

$$= P_{k,j,1}\cdot \Pr\left(\|\Pi^{(j)}\mathbf{h}_k\|^2 < (2^{R^{th}}-1)\cdot\frac{c}{\delta_1^{(j)}}\cdot\frac{K}{\ell}P_t^{-1}\right) + \sum_{i=2}^{\ell}\left[P_{k,j,i}\cdot \Pr\left(\|\Pi^{(j)}\mathbf{h}_k\|^2 < c\left(\frac{\delta_i^{(j)}}{2^{R^{th}}-1}-\sum_{m=1}^{i-1}\delta_m^{(j)}\right)^{-1}\cdot\frac{K}{l}\cdot P_t^{-1}\right)\right]. \tag{55}$$

Corollary 1: For the Gaussian MISO BC with N transmit antennas and K single-antenna users with independent Rayleigh fading described in Section II-A, the diversity order of the overall system achievable by the mixture transceiver architecture with the proposed user grouping method is given by $D = N$.

Proof: The decay rate of the overall outage probability is dominated by the worst decay rate. The worst diversity order in Theorem 1 occurs when $k = K$, and is given by N. ■

Note that the diversity order of the full ZF downlink beamforming is given by [14]

$$D = N - K + 1. \tag{56}$$

Hence, a significant improvement in the diversity order is attained by the mixture scheme. Note that the possible maximum diversity order for User k with channel $\mathbf{h}_k \sim \mathcal{CN}(0, 2\mathbf{I})$ is simply N. Hence, the mixture transceiver architecture achieves the full diversity order N for MISO BCs.

C. Diversity and Multiplexing Trade-Off

With the cluster power factors $\{\delta_i > 0, i = 1, \cdots, L\}$ fixed, as the total cluster power P increases according to (17) without bound, in each group only the rate of the first user scales as log SNR but the rates of all other users saturate to constants: $\bar{R}_i^{(j)} = \log_2\left(1+\frac{\delta_i^{(j)}}{\sum_{m=1}^{i-1}\delta_m^{(j)}}\right)$, $i = 2, \cdots, L$, as seen in (19). (The target rates of User i, $i = 2, \cdots, L$, should be less than $\bar{R}_i^{(j)}$, but δ_i's can be designed for a common target rate R_{th} based on (84). Please see [23, Sec. V.C].) Hence, the multiplexing gain for one user group with superposition and SIC is one regardless of the number of users in the group. (A similar observation of multiplexing gain one per superposition-and-SIC user group was made in [25].) Thus, the overall multiplexing gain of the mixture scheme with the adaptive user grouping is the same as the number of user groups N_g which is less than or equal to $K(\leq N)$. Note that in the case of $K = N$, the multiplexing gain of the ZF beamforming is N, whereas its diversity order is one. Thus, diversity-order and multiplexing-gain trade-off known in single-user MIMO [26] occurs even in MISO BCs [27]. In fact, it can be shown by replacing L with K in Proposition 1 and going through the proof of Theorem 1 with $N_g = 1$ that the full diversity order N can be achieved by a single superposition-and-SIC group containing all K users without considering channel alignment and orthogonality at all. However, this single-group full superposition-and-SIC approach is not good since it yields multiplexing gain one regardless of channel realization. This scheme can be considered as an antipodal scheme of the ZF beamforming in terms of diversity and multiplexing trade-off: The diversity order and multiplexing gain of the full superposition-and-SIC approach versus full ZF beamforming are $(N, 1)$ versus $(1, N)$ for MISO BCs with $K = N$. On the other hand, the proposed user grouping method is adaptive and depends on the channel realization. The number of groups is not predetermined in the proposed user grouping method. The number of user groups can be K if all user channels are semi-orthogonal. The number of user groups can be one if all user channels are aligned. Hence, the number N_g of constructed user groups, i.e., the multiplexing gain of the mixture scheme, is adaptive to channel realization, while

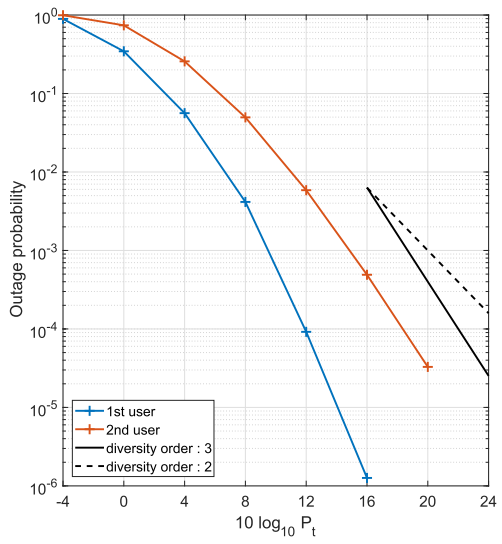
the full diversity order N is always achieved. So, we can view that the mixture scheme with such an adaptive user grouping method tries to opportunistically increase the multiplexing gain while achieving the full diversity order. An assessment of the multiplexing gain loss of our scheme compared to the ZF beamforming can be found in [23], where it is seen that the loss is not significant.

V. Numerical Results

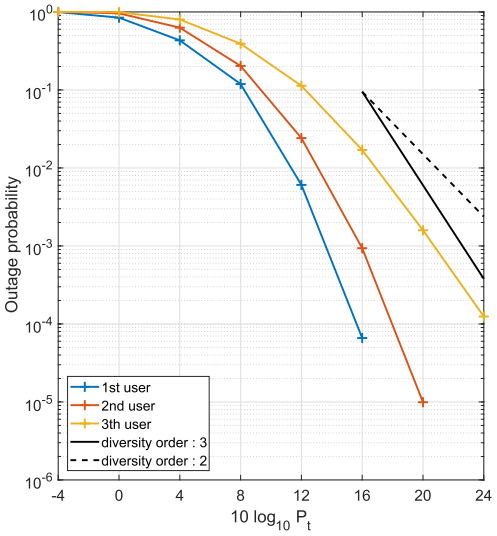
In this section, we provide some numerical results to validate our theoretical analysis in the previous sections. We considered the MISO BC described in Section II-A. In each simulation scenario, we generated the K channel vectors $\mathbf{h}_1, \cdots, \mathbf{h}_K$ of the system independently from the zero-mean complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, 2\mathbf{I})$ sufficiently many times to numerically compute outage probability. For each channel realization, we ran the user grouping algorithm (Algorithm 1) with $\theta^{th} = 0.9$. With the constructed groups, we applied inter-group ZF beamforming and designed the intra-group beam vectors according to the constraint (17), i.e., $\mathbf{w}_1 = \cdots = \mathbf{w}_L = \mathbf{w}^*$ with the solution $\mathbf{w}^*$ to the max-min problem (67) used in the proof of Proposition 1. The rate R_k of the k-th user is obtained based on the designed beam vectors. (Note that the beam vectors $\mathbf{w}_1 = \cdots = \mathbf{w}_L = \mathbf{w}^*$ designed in this way yield rates larger than or equal to the lower bounds in (18) and (19).) For the intra-group beam design, the power distribution factors are chosen to satisfy the condition in Lemma 4 in Appendix D. The used values for power distribution factors are $(0.2, 0.8)$ for every two-user group, $(0.05, 0.2, 0.75)$ for every three-user group in Figs. 2(a), 2(b) and 3(a), and are the solution of (84) with $C = 2$ in Fig. 3(b). For computation of the outage probability $\Pr(R_k \leq R_{th})$, we set the target rate threshold as $R_{th} = 1.5$ [bits/channel use] in all simulations.

First, we numerically evaluated the outage probability and diversity order of each user of the mixture transceiver architecture considering channel norm ordering. Fig. 2 shows the outage probability of the mixture transceiver architecture in two cases: (a) $N = 3$, $K = 2$ and (b) $N = 3$ and $K = 3$, where User k is defined as the user with the k-th largest channel norm (i.e. $\|\mathbf{h}_1\|^2 \geq \|\mathbf{h}_2\|^2 \geq \cdots \geq \|\mathbf{h}_K\|^2$). In the case (a) of $N = 3$, $K = 2$, Theorem 1 states that the diversity orders of Users 1 and 2 are 6 and 3, respectively. It is seen in Fig. 2(a) that the outage probability of User 2 has the slope corresponding to diversity order of 3, as SNR increases. It is also seen that the decay rate of User 1 is almost twice that of User 2. (In $\log_{10}$ y-scale, roughly User 1 has -4 and -5.9 and User 2 has -2.2 and -3.3 at $10\log P_t = 12$ and 16, respectively.) In the case (b) of $N = 3$, $K = 3$, Theorem 1 states that the diversity orders of Users 1, 2 and 3 are 9, 6, and 3, respectively. It is observed in Fig. 2(b) that the outage probability of User 3 has the slope corresponding to diversity order of 3, as SNR increases.

Then, we compared the mixture transceiver architecture with the full ZF downlink beamforming, based on the overall system diversity order. In order to see the overall diversity order, we computed overall outage probability. For this, we neglected channel norm ordering and computed the total



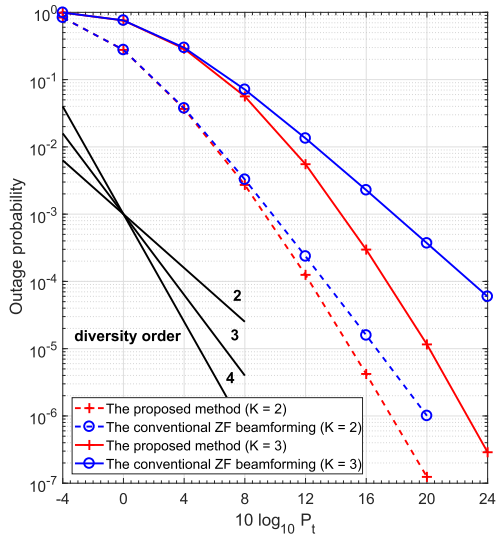
(a)



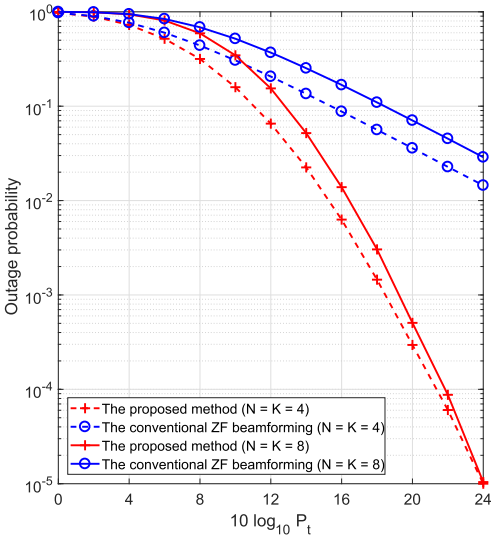
(b)

Fig. 2. Outage probability of the mixture transceiver architecture: (a) $N = 3$, $K = 2$ and (b) $N = 3$, $K = 3$.

number of outages occurred at all K users over all Monte Carlo runs. Fig. 3 shows the overall outage probability for the same channel statistics and the same rate threshold for the mixture scheme and the ZF downlink beamforming. We considered four cases: *i)* $N = 4$, $K = 2$ and *ii)* $N = 4$, $K = 3$ shown in Fig. 3(a) and *iii)* $N = K = 4$ and *iv)* $N = K = 8$ shown in Fig. 3(b). For the considered cases *i), ii), iii),* and *iv)*, the corresponding system diversity orders of the mixture scheme are 4, 4, 4 and 8 by Corollary 1, whereas the corresponding diversity orders of the ZF downlink beamforming are 3, 2, 1, and 1 by (56). It is seen in Fig. 3(a) that indeed the diversity orders of cases *i)* and *ii)* for the mixture scheme are the same as four. (The two red curves in Fig. 3(a) seem to have the same slope with some offset, as SNR increases.) On the other hand, it is seen that the diversity orders of the ZF downlink beamforming depends on K for the same N, as expected. The outage performance result for the cases with more transmit antennas $N = K = 4$ and $N = K = 8$ is shown in Fig. 3(b).

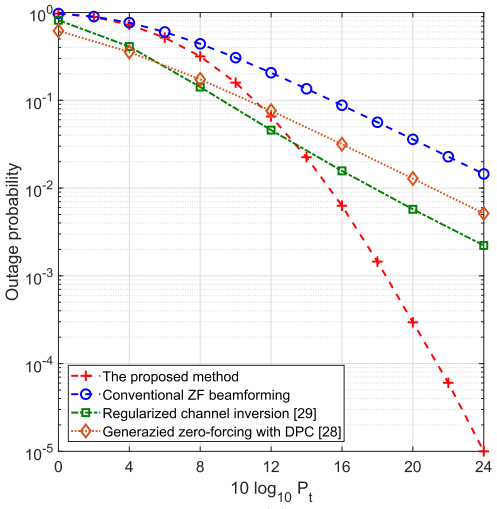


(a)

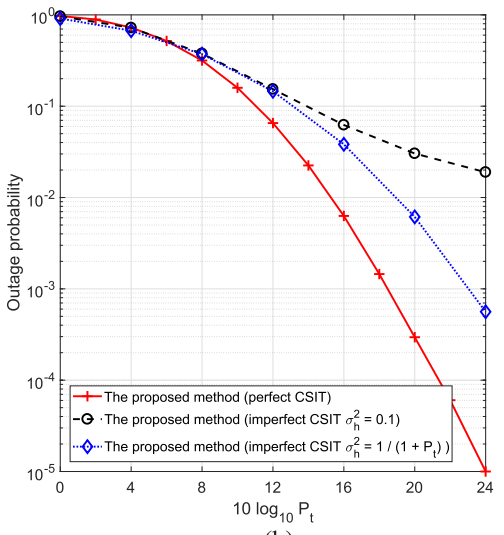


(b)

Fig. 3. Overall outage probability : (a) $N = 4$, $K = 2$ or 3 and (b) $N = K = 4$ and $N = K = 8$.



(a)



(b)

Fig. 4. Overall outage probability : (a) $N = 4$, $K = 2$ or 3 and (b) $N = K = 4$ and $N = K = 8$.

It is seen that the full ZF beamforming yields the same slope for the two cases $N = K = 4$ and $N = K = 8$, as expected, since it yields the diversity order of one in both cases by (56). On the other hand, it is seen that the diversity orders in the two cases $N = K = 4$ and $N = K = 8$ are different for the mixture scheme, as predicted by Corollary 1. Indeed, it is seen that the decay rate of the outage probability in the case of $N = K = 8$ is larger than that of the case of $N = K = 4$, although the outage probability of the case $N = K = 8$ is higher than that of the case $N = K = 4$ at low SNR. Note that the outage performance gain by the mixture scheme over the ZF beamforming is drastic in the case of $N = K = 4$ and $N = N = 8$ for the meaningful range where the outage probability is below 10^{-2}.

Next, we considered other advanced transceiver designs for MISO BCs, e.g., [28], [29], devised to improve the performance over ZF downlink beamforming, and compared the outage performance of these advanced designs with the mixture architecture. The result is shown in Fig. 4(a), where the setup is $N = K = 4$ and other parameter setting is the same as that in Fig. 3(b) with $N = K = 4$. It is seen that the advanced transceiver designs having the full multiplexing gain yield the same diversity order as the ZF beamforming although they yield better rates compared to the ZF beamforming. Thus, these advanced designs are at the multiplexing-gain-optimal side in terms of diversity-and-multiplexing trade-off. Although analysis of the outage performance under imperfect CSIT is beyond the scope of this paper, we briefly investigated the impact of imperfect CSIT through simulation. Again we considered the case of $N = K = 4$ with the same setting as that in Fig. 3(b). It is known that the number of CSI feedback bits per user should increase linearly with respect to SNR (or signal power for fixed noise variance) in log scale in order to achieve full multiplexing gain for MISO BCs [30]. For simulation the CSI error is assumed to be zero-mean

Gaussian with variance σ_e^2, where we set 1) σ_e^2 as a fixed constant of 0.1 and 2) $\sigma_e^2 = \frac{1}{1+P_t}$ to be consistent with the result in [30]. The result is shown in Fig. 4(b). It is seen that the fixed CSI quality with respect to SNR shows a floor for the outage probability as SNR increases. On the other hand, the CSI with quality $\sigma_e^2 = \frac{1}{1+P_t}$ does not show such a floor behavior. Indeed, it seems that the increasing CSI quality with respect to SNR is required to achieve the full diversity order although the exact increasing rate is not known yet. For numerical results on the rate distribution and the multiplexing gain loss, the interested reader is referred to [23], where it is seen that the loss is not significant.

VI. Conclusion

In this paper, we have considered the mixture transceiver architecture with channel-adaptive user grouping and mixture of linear and nonlinear SIC reception for MISO BCs, and have shown that the mixture transceiver architecture opportunistically increases the multiplexing gain while achieving full diversity order for MISO BCs. The mixture transceiver architecture can provide better outage performance compared to the widely-used conventional ZF downlink beamforming for MU-MISO BCs under channel fading environments. The gain in diversity order results from possible sacrifice of multiplexing gain through diversity-and-multiplexing trade-off, and thus the mixture scheme provides an alternative transceiver architecture for MISO BCs to applications such as emerging URLLC in which reliability is more important than data rate. Future research directions include optimization of angle threshold and power distribution, finding optimal diversity-and-multiplexing trade-off in MISO BCs, finding faster grouping algorithms scalable with the number of users for large systems, application of the mixture architecture to the uplink [15], and application of more advanced transmit signaling [31].

Appendix A
Proof of Proposition 1

For given $(\delta_1, \cdots, \delta_L)$, in order to obtain a lower bound on the achievable rate of each user, we simply set $\mathbf{w}_1 = \mathbf{w}_2 = \cdots = \mathbf{w}_L = \mathbf{w}$ with $||\mathbf{w}||^2 \leq 1$ as in the constraint (17). Then, the rates in (14) of the MISO BC with superposition coding and SIC can be rewritten as (57) - (58), shown at the top of next page. Using Lemma 2 below, we can bound the terms $|\mathbf{g}_1^H\mathbf{w}|^2$ in (57) and $\min\{|\mathbf{g}_1^H\mathbf{w}|^2, \cdots, |\mathbf{g}_i^H\mathbf{w}|^2\}$ in (58) as follows: Using the optimal solution $\mathbf{w}^*$ to the max-min problem (67), we have

$$|\mathbf{g}_1^H\mathbf{w}^*|^2 \geq \min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|}\right)^H\mathbf{w}^*\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_L}{\|\mathbf{g}_L\|}\right)^H\mathbf{w}^*\right|^2\right\}\|\mathbf{g}_1\|^2 \tag{59}$$

$$\geq \frac{1}{c}\|\mathbf{g}_1\|^2, \tag{60}$$

where (59) is valid since the minimum is taken over multiple terms including $|\mathbf{g}_1^H\mathbf{w}^*|^2$, and (60) is valid by Lemma 2 below. Next, we have

$$\min\{|\mathbf{g}_1^H\mathbf{w}^*|^2, \cdots, |\mathbf{g}_i^H\mathbf{w}^*|^2\} = \min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_i\|}\right)^H\mathbf{w}^*\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}\right)^H\mathbf{w}^*\right|^2\right\}\|\mathbf{g}_i\|^2 \tag{61}$$

$$\geq \min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|}\right)^H\mathbf{w}^*\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}\right)^H\mathbf{w}^*\right|^2\right\}\|\mathbf{g}_i\|^2 \tag{62}$$

$$\geq \min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|}\right)^H\mathbf{w}^*\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_L}{\|\mathbf{g}_L\|}\right)^H\mathbf{w}^*\right|^2\right\}\|\mathbf{g}_i\|^2 \tag{63}$$

$$\geq \frac{1}{c}\|\mathbf{g}_i\|^2, \tag{64}$$

where (62) is valid since $||\mathbf{g}_1|| \geq \cdots \geq ||\mathbf{g}_L||$, (63) is valid since we increased the number of terms in the minimization including the previous terms, and (64) holds by Lemma 2 below. Substituting (60) and (64) into (57) and (58), respectively, we have the rates that can be achieved by the optimal solution $\mathbf{w}^* = \mathbf{w}_1 = \cdots = \mathbf{w}_L$ to the max-min problem (67):

$$R_1 \geq \log_2\left(1 + \frac{1}{c}\delta_1\|\mathbf{g}_1\|^2 P\right) \tag{65}$$

$$R_i \geq \log_2\left(1 + \frac{\delta_i}{\sum_{m=1}^{i-1}\delta_m}\frac{1}{1+\left(\frac{1}{c}\|\mathbf{g}_i\|^2\sum_{m=1}^{i-1}\delta_m P\right)^{-1}}\right) \quad i = 2, \cdots, L. \tag{66}$$

The considered design here of $\mathbf{w}_1 = \cdots = \mathbf{w}_L = \mathbf{w}^*$ with $||\mathbf{w}^*||^2 \leq 1$ and $p_i = \delta_i P$ with $(\delta_1, \cdots, \delta_L) \in \mathcal{D}$, i.e., (17), satisfies the original beam design constraint $(\mathbf{w}_1, \cdots, \mathbf{w}_L) \in \mathcal{W}^L$ and $p_i > 0, \forall i$, $\sum_{i=1}^L p_i = P$ in (16). Hence, the rates achieved by $\mathbf{w}_1 = \cdots = \mathbf{w}_L = \mathbf{w}^*$ with $(\delta_1, \cdots, \delta_L)$ are lower bounds on the achievable rates. ■

Lemma 2: Consider the following max-min optimization problem:

$$\max\min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|}\right)^H\mathbf{w}\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_L}{\|\mathbf{g}_L\|}\right)^H\mathbf{w}\right|^2\right\} \quad \text{subject to } \|\mathbf{w}\|^2 \leq 1. \tag{67}$$

The optimal solution $\mathbf{w}^*$ to the problem (67) satisfies the following:

$$\min\left\{\left|\left(\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|}\right)^H\mathbf{w}^*\right|^2, \cdots, \left|\left(\frac{\mathbf{g}_L}{\|\mathbf{g}_L\|}\right)^H\mathbf{w}^*\right|^2\right\} \geq \frac{1}{c}, \tag{68}$$

where

$$c = \begin{cases} L & \text{if } L \leq 3, \\ 8L^2 & \text{if } L > 3. \end{cases} \tag{69}$$

$$R_1 = \log_2\left(1+\delta_1\ P|\mathbf{g}_1^H\mathbf{w}|^2\right) \tag{57}$$

$$\begin{aligned} R_i &= \log_2\left(1+\min\left\{\frac{\delta_i P|\mathbf{g}_1^H\mathbf{w}|^2}{\sum_{m=1}^{i-1}\delta_m P|\mathbf{g}_1^H\mathbf{w}|^2+1},\cdots,\frac{\delta_i P|\mathbf{g}_i^H\mathbf{w}|^2}{\sum_{m=1}^{i-1}\delta_m P|\mathbf{g}_i^H\mathbf{w}|^2+1}\right\}\right), \\ &= \log_2\left(1+\frac{\delta_i}{\sum_{m=1}^{i-1}\delta_m}\cdot\frac{1}{1+1/\left[\min\{|\mathbf{g}_1^H\mathbf{w}|^2,\cdots,|\mathbf{g}_i^H\mathbf{w}|^2\}(\sum_{m=1}^{i-1}\delta_m)P\right]}\right), \quad i=2,\cdots,L. \end{aligned} \tag{58}$$

$$\begin{aligned} &\Pi^{\perp}_{[\mathbf{A},\mathbf{B}]} \\ &= \mathbf{I}-[\mathbf{A}\ \mathbf{B}]\begin{bmatrix}\mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{B}\\ \mathbf{B}^H\mathbf{A} & \mathbf{B}^H\mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{A}^H\\ \mathbf{B}^H\end{bmatrix} \qquad (70)\\ &\overset{(a)}{=}\mathbf{I}-[\mathbf{A}\ \mathbf{B}]\begin{bmatrix}(\mathbf{A}^H\mathbf{A})^{-1}+(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}, & \\ -(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}, & \\ -(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1} & (\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{A}^H\\ \mathbf{B}^H\end{bmatrix} \\ &= \mathbf{I}-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H-\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \\ &\quad+\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H+\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H \\ &\quad-\mathbf{B}(\mathbf{B}^H\mathbf{B}-\mathbf{B}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{B})^{-1}\mathbf{B}^H \\ &= \mathbf{I}-\Pi_{\mathbf{A}}-\Pi_{\mathbf{A}}\mathbf{B}(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi_{\mathbf{A}}\mathbf{B})^H+\mathbf{B}(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi_{\mathbf{A}}\mathbf{B})^H+\Pi_{\mathbf{A}}\mathbf{B}(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}\mathbf{B}^H-\mathbf{B}(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}\mathbf{B}^H \\ &= \mathbf{I}-\Pi_{\mathbf{A}}-(\mathbf{B}-\Pi_{\mathbf{A}}\mathbf{B})(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\mathbf{B}-\Pi_{\mathbf{A}}\mathbf{B})^H \\ &= \Pi^{\perp}_{\mathbf{A}}-\Pi^{\perp}_{\mathbf{A}}\mathbf{B}(\mathbf{B}^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H \\ &\overset{(b)}{=}\Pi^{\perp}_{\mathbf{A}}-\Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H \qquad (71)\end{aligned}$$

Proof of Lemma 2

Define unit-norm $\mathbf{v}_i := \mathbf{g}_i/\|\mathbf{g}_i\|$ for $i=1,\cdots,L$. Then, (67) can be rewritten as

$$\begin{aligned} &\max\min\left\{\left|\mathbf{v}_1^H\mathbf{w}\right|^2,\cdots,\left|\mathbf{v}_L^H\mathbf{w}\right|^2\right\} \\ &\text{subject to } \|\mathbf{w}\|^2\le 1 \end{aligned} \tag{72}$$

The problem (72) can be reformulated as

$$\begin{aligned} \max\ &\frac{\min\left\{\left|\mathbf{v}_1^H\mathbf{w}\right|^2,\cdots,\left|\mathbf{v}_L^H\mathbf{w}\right|^2\right\}}{||\mathbf{w}||^2} \\ &= \min\ \frac{||\mathbf{w}||^2}{\min\left\{\left|\mathbf{v}_1^H\mathbf{w}\right|^2,\cdots,\left|\mathbf{v}_L^H\mathbf{w}\right|^2\right\}}, \end{aligned} \tag{73}$$

where inversion of the cost function is taken in the right-hand side (RHS) of (73). Thus, it is known that the optimal value of the problem (72) is equivalent to the inverse of the optimal value of the following quadratic programming (QP) [32]:

$$\begin{aligned} &\min\ \|\mathbf{w}\|^2 \\ &\text{subject to } |\mathbf{v}_i^H\mathbf{w}|^2\ge 1,\quad i=1,\cdots,L. \end{aligned} \tag{74}$$

The QP (74) can be solved by semi-definite relaxation of the rewritten form of (74) [32]:

$$\begin{aligned} &\min\ \mathrm{Tr}(\mathbf{W}) \\ &\text{subject to } \mathrm{Tr}(\mathbf{V}_i\mathbf{W})\ge 1,\quad i=1,\cdots,L \end{aligned} \tag{75}$$

where $\mathbf{W}:=\mathbf{w}\mathbf{w}^H$ and $\mathbf{V}_i:=\mathbf{v}_i\mathbf{v}_i^H$, $i=1\cdots,L$. Denote the optimal values of the optimization problems (74) and (75) by v^*_{qp} and v^*_{sdp}, respectively. Then, the relationship between v^*_{qp} and v^*_{sdp} is known as [33]

$$\begin{aligned} &v^*_{qp}=v^*_{sdp},\quad \text{if } L\le 3, \\ &v^*_{qp}\le 8L\cdot v^*_{sdp},\quad \text{if } L>3. \end{aligned} \tag{76}$$

Furthermore, note that $\mathbf{W}':=\sum_{i=1}^{L}\mathbf{V}_i$ is feasible for the problem (75) since $\mathrm{Tr}(\mathbf{V}_i\mathbf{W}')=\mathrm{Tr}(\mathbf{V}_i\sum_{i=1}^{L}\mathbf{V}_i))\ge\sum_{i=1}^{L}\mathrm{Tr}(\mathbf{V}_i\mathbf{V}_i))\ge\mathrm{Tr}(\mathbf{V}_i\mathbf{V}_i)=1$, and $\mathrm{Tr}(\mathbf{W}')=L$. Hence, we have

$$v^*_{sdp}\le L. \tag{77}$$

Hence, with the optimal solution $\mathbf{w}^*$ to (72), we have

$$\begin{aligned} \min\left\{\left|\mathbf{v}_1^H\mathbf{w}^*\right|^2,\cdots,\left|\mathbf{v}_1^H\mathbf{w}^*\right|^2\right\} &\overset{(a)}{=} 1/v^*_{qp} \\ &\overset{(b)}{\ge} L/c\cdot 1/v^*_{sdp} \\ &\overset{(c)}{\ge} 1/c, \end{aligned} \tag{78}$$

where c is given by (69). Here, Step (a) is valid due to the relationship between the original problem (72) and the QP (74); Step (b) is valid due to (76); and Step (c) is valid due to (77). ■

APPENDIX B
PROOF OF LEMMA 1

First, it is shown in (70) - (71) shown at the top of the previous page that $\Pi^{\perp}_{[\mathbf{A},\mathbf{B}]}\mathbf{x} = \Pi^{\perp}_{\mathbf{A}} - \Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H$, where $\Pi_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$, $\Pi^{\perp}_{\mathbf{A}} = \mathbf{I} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$. (In (70) - (71), the block matrix inversion formula is used in Step (a). In Step (b), we used $\Pi^{\perp H}_{\mathbf{A}}\Pi^{\perp}_{\mathbf{A}} = (\Pi^{\perp}_{\mathbf{A}})^2 = \Pi^{\perp}_{\mathbf{A}}$.) Therefore, we have

$$\begin{aligned}\Pi^{\perp}_{[\mathbf{A},\mathbf{B}]}\mathbf{x} &= \Pi^{\perp}_{\mathbf{A}}\mathbf{x} - \Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\mathbf{x}\\ &= \Pi^{\perp}_{\mathbf{A}}\mathbf{x} - \Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\\ &\quad\cdot(\Pi_{\mathbf{A}}\mathbf{x} + \Pi^{\perp}_{\mathbf{A}}\mathbf{x})\\ &\overset{(c)}{=} \Pi^{\perp}_{\mathbf{A}}\mathbf{x} - \Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{x}\\ &= (\mathbf{I} - \Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H)\Pi^{\perp}_{\mathbf{A}}\mathbf{x}\\ &= (\mathbf{I} - \Pi_{\Pi^{\perp}_{\mathbf{A}}\mathbf{B}})\Pi^{\perp}_{\mathbf{A}}\mathbf{x},\end{aligned}$$

where Step (c) holds because $\Pi^{\perp}_{\mathbf{A}}\mathbf{B}((\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^{-1}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})^H$ is the projection onto $\mathcal{C}(\Pi^{\perp}_{\mathbf{A}}\mathbf{B})$ which is a subspace contained in $\mathcal{C}^{\perp}(\mathbf{A})$. ■

APPENDIX C
PROOF OF PROPOSITION 2

Consider the effective channel $\mathbf{g}_i^{(j)} = \Pi^{\perp}_{\tilde{\mathbf{H}}_j}\mathbf{h}_i^{(j)}$, where $\mathbf{h}_i^{(j)}$ is the channel vector of User i in group $\mathcal{G}_j$, and $\tilde{\mathbf{H}}_j$ is defined in (10). By Lemma 1, $\mathbf{g}_i^{(j)} = \Pi^{\perp}_{\tilde{\mathbf{H}}_j}\mathbf{h}_i^{(j)}$ can be obtained from sequentially projecting $\mathbf{h}_i^{(j)}$ onto the sequential orthogonal spaces associated with the channel vectors of $\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_{j-1}, \mathcal{G}_{j+1}, \cdots, \mathcal{G}_{N_g}$, as discussed in Lemma 1 and Example 1, i.e., $\mathbf{g}_i^{(j)} = \mathcal{P}(\mathcal{G}_{N_g}|\mathcal{G}_{N_g-1}, \cdots, \mathcal{G}_{j+1}, \mathcal{G}_{j-1}, \cdots, \mathcal{G}_1)\cdots\mathcal{P}(\mathcal{G}_{j+1}|\mathcal{G}_{j-1}, \cdots, \mathcal{G}_1)\mathcal{P}(\mathcal{G}_{j-1}|\mathcal{G}_{j-2}, \cdots, \mathcal{G}_1)\cdots\mathcal{P}(\mathcal{G}_2|\mathcal{G}_1)\ \mathcal{P}(\mathcal{G}_1)\mathbf{h}_i^{(j)}$, where $\mathcal{P}(\mathcal{B}|\mathcal{A})$ denotes the sequential projection onto the orthogonal space of the projected subspace of $\mathcal{B}$ onto $\mathcal{C}^{\perp}(\mathcal{A})$. (Please see Lemma 1 and Example 1.) Here, we have $N_g - 1$ projection stages. At each projection stage, the proposed user grouping algorithm, Algorithm 1, guarantees that norm reduction is not beyond $(1-\theta^{th})$. The norm of the ZF effective channel can be written as (see Example 1)

$$||\mathbf{g}_i^{(j)}||^2 = Y||\mathbf{h}_i^{(j)}||^2, \tag{79}$$

where the reduction gain random variable Y depends on the channels, but $(1-\theta^{th})^{K-1} =: Y^{th} \le Y \le 1$ since $N_g \le K$. By (79) and Lemma 3 below, the claim follows. ■

Lemma 3: Let X be a random variable satisfying the condition, $\lim_{x\to 0}\frac{\log \Pr(X\le x)}{\log x} = d$, and let Y be a random variable satisfying the condition, $Y^{th} \le Y \le 1$, where Y^{th} is some constant $\in (0, 1]$ and d is some positive constant. Then, the product $Z := XY$ satisfies $\lim_{z\to 0}\frac{\log \Pr(Z\le z)}{\log z} = d$.

Proof of Lemma 3:

$$\begin{aligned}&\Pr(X \le z) \le \Pr(Z \le z) \le \Pr(Y^{th}X \le z)\\ &\Leftrightarrow \Pr(X \le z) \le \Pr(Z \le z) \le \Pr(X \le \frac{z}{Y^{th}})\\ &\Leftrightarrow \lim_{z\to 0}\frac{\log \Pr(X \le z)}{\log z} \le \lim_{z\to 0}\frac{\log \Pr(Z \le z)}{\log z}\\ &\le \lim_{z\to 0}\frac{\log \Pr(X \le \frac{z}{Y^{th}})}{\log z}\\ &\Leftrightarrow \lim_{z\to 0}\frac{\log \Pr(X \le z)}{\log z} \le \lim_{z\to 0}\frac{\log \Pr(Z \le z)}{\log z}\\ &\le \lim_{z\to 0}\frac{\log \Pr(X \le \frac{z}{Y^{th}})}{\log \frac{z}{Y^{th}}}\frac{\log \frac{z}{Y^{th}}}{\log z}\\ &\Leftrightarrow d \le \lim_{z\to 0}\frac{\log \Pr(Z \le z)}{\log z} \le d,\end{aligned} \tag{80}$$

where (80) holds because $Y^{th}X \le Z = YX \le X$ due to $Y \in (Y^{th}, 1)$. Therefore, the claim follows. ■

APPENDIX D
EXISTENCE OF POWER DISTRIBUTION FACTORS

Lemma 4: There always exists a collection of in-group power distribution factors $(\delta_1^{(j)}, \cdots, \delta_\ell^{(j)})$ for $\mathcal{G}_j$ with $|\mathcal{G}_j| = \ell$ such that $(\frac{\delta_i^{(j)}}{2^{R^{th}}-1} - \sum_{m=1}^{i-1}\delta_m^{(j)})$ in (55) is strictly positive for all $i = 2, \cdots, \ell$.

Proof: The condition is equivalent to the following:

$$\frac{\delta_i^{(j)}}{2^{R^{th}}-1} - \sum_{m=1}^{i-1}\delta_m^{(j)} > 0 \Leftrightarrow \frac{\delta_i^{(j)}}{\sum_{m=1}^{i-1}\delta_m^{(j)}} > 2^{R^{th}} - 1 \tag{81}$$

for $i = 2, \cdots, \ell$. Consider the following recursion

$$\delta_i^{(j)} = (2^{R^{th}} - 1 + C)(\delta_1^{(j)} + \cdots + \delta_{i-1}^{(j)}), \tag{82}$$

where $C > 0$ is an arbitrary positive constant. It is easy to see that any solution to (82) satisfies (81). Solving the recursion yields

$$\delta_i^{(j)} = \delta_1^{(j)}(2^{R^{th}} - 1 + C)(2^{R^{th}} + C)^{i-2}. \tag{83}$$

With normalization for $\sum_{i=1}^{\ell}\delta_i^{(j)} = 1$, we have

$$\delta_1^{(j)} = \frac{1}{(2^{R^{th}} + C)^{\ell-1}}, \quad \text{and } \delta_i^{(j)} = \frac{2^{R^{th}} - 1 + C}{(2^{R^{th}} + C)^{\ell-i+1}} \tag{84}$$

for $i = 2, \cdots, \ell$, and all $\delta_i^{(j)} \ge 0$. Hence, we have a collection of power distribution factors for the condition. ■

REFERENCES

[1] J. Seo, "Beamformer design based on non-linear reception in MIMO downlink system," Ph.D. dissertation, School Elect. Eng., KAIST, Daejeon, South Korea, Aug. 2018.

[2] 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, Sep. 2009.

[3] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[4] 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, Mar. 2006.

[5] *Evolved Universal Terrestrial Radio Access (E-UTRA): Downlink Multiple Input Multiple Output (MIMO) Enhancement for LTE-Advanced Release 11, V11.0.0,*, document TR 36.871, 3GPP, pp. 2011–2012.

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

[7] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.
[8] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1086–1100, Feb. 2016.
[9] G. Lee, Y. Sung, and M. Kountouris, "On the performance of random beamforming in sparse millimeter wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 560–575, Apr. 2016.
[10] H. Huh, A. M. Tulino, and G. Caire, "Network mimo with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
[11] 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 VTC*, Jun. 2013, pp. 1–5.
[12] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 133, Dec. 2018.
[13] J. Seo and Y. Sung, "A new transceiver architecture for multi-user MIMO communication based on mixture of linear and non-linear reception," in *Proc. SPAWC*, Sapporo, Japan, Jun. 2017, pp. 1–5.
[14] Z. Chen, Z. Ding, and X. Dai, "Beamforming for combating inter-cluster and intra-cluster interference in hybrid NOMA systems," *IEEE Access*, vol. 4, pp. 4452–4463, 2016.
[15] J. Kazemitabar and H. Jafarkhani, "Multiuser interference cancellation and detection for users with more than two transmit antennas," *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 574–583, Apr. 2008.
[16] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Oint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ Press, 2005.
[18] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.
[19] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.
[20] J. Seo and Y. Sung, "Beam design and user scheduling for nonorthogonal multiple access with multiple antennas based on Pareto optimality," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2876–2891, Jun. 2018.
[21] R. Zhang and S. Cui, "Cooperative interference management with miso beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.
[22] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.
[23] J. Seo, Y. Sung, and H. Jafarkhani. (Oct. 2018). "A high-diversity transceiver design for MISO broadcast channels." [Online]. Available: https://arxiv.org/abs/1807.00114
[24] J. Hou, J. E. Smee, H. D. Pfister, and S. Tomasin, "Implementing interference cancellation to increase the EV-DO rev a reverse link capacity," *IEEE Commun. Mag.*, vol. 44, no. 2, pp. 58–64, Feb. 2006.
[25] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.
[26] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[27] L. Mroueh, S. Rouquette-Leveil, G. R.-B. Othman, and J.-C. Belfiore, "Dmt of weighted parallel channels: Application to broadcast channels," in *Proc. ISIT*, pp. 2361–2365, Jul. 2008.
[28] S. Hu and F. Rusek, "A generalized zero-forcing precoder with successive dirty-paper coding in MISO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3632–3645, Jun. 2017.
[29] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[30] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
[31] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, Jun. 2013.
[32] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.
[33] Z.-Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang, "Approximation bounds for quadratic optimization with homogeneous quadratic constraints," *SIAM J. Optim.*, vol. 18, no. 1, pp. 1–28, 2007.

Junyeong Seo (S11–M'19) received the B.S., M.S., and Ph.D. degrees from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2011, 2013, and 2018, respectively, all in electrical engineering. He is currently with Samsung Electronics, and is involved in 5G research. His research interests include design and analysis of large-scale MIMO systems, new multiple access methods, and signal processing for 5G-and-beyond wireless communication systems.

Youngchul Sung (S'92–M'93–SM'09) received B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2005. Before joining the Ph.D. program, he was with LG Electronics Ltd., Seoul, from 1995 to 2000. From 2005 to 2007, he was a Senior Engineer with the Corporate Research and Development Center, Qualcomm, Inc., San Diego, CA, USA, involved in design of Qualcomm's 3GPP R6 WCDMA base station modem. Since 2007, he has been on the faculty with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

His research interests include signal processing for communications, statistical signal processing, and statistical inference and learning with applications to wireless communications and related areas. He is member of the Signal Processing and Communications Electronics Technical Committee of IEEE Communications Society, and the Signal and Information Processing Theory and the Methods Technical Committee of Asia-Pacific Signal and Information Processing Association. He is currently the Chair of the IEEE Communications Society Asia-Pacific Board and Chapters Coordination Committee. He is currently an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, from 2012 to 2014, and a Guest Editor of the 2012 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue Theories and Methods for Advanced Wireless Relays.

Hamid Jafarkhani (M'89–SM'01–F'06) was a Visiting Scholar with Harvard University in 2015 and a Visiting Professor with the California Institute of Technology in 2018. He is currently a Chancellor's Professor with the Department of Electrical Engineering and Computer Science, University of California Irvine, where he is also the Director of the Center for Pervasive Communications and Computing and the Conexant-Broadcom Endowed Chair. He was a recipient of the IEEE Marconi Prize Paper Award in Wireless Communications, the IEEE Communications Society Award for Advances in Communication, and the IEEE Eric E. Sumner Award.

He is listed as a highly cited researcher in http://www.isihighlycited.com. According to Thomson Scientific, he is one of the top 10 most-cited researchers in the field of computer science from 1997 to 2007. He was the 2017 Innovation Hall of Fame Inductee at the University of Maryland's School of Engineering. He is a Fellow of AAAS and the author of the book *Space-Time Coding: Theory and Practice.*