Multiuser Selection Criteria for MIMO-NOMA Systems With Different Detectors

Quan Yu, *Senior Member, IEEE*, Chao Han, *Student Member, IEEE*, Lin Bai, *Senior Member, IEEE*, Jingchao Wang, Jinho Choi, *Senior Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA) is an effective way to improve the spectrum efficiency of multiple-input multiple-output (MIMO) systems, where this advancement can be fully exploited by employing a proper user selection strategy in the multiuser MIMO system. However, for multiuser scenarios, information-theoretic throughput may not be properly used as the performance indicator since the constraints at the receiver (e.g., suboptimal detectors) are not taken into consideration. In this paper, we adopt the error probability (EP) as the performance metric, which is equivalent to the practical throughput efficiency and can take the critical constraints into consideration. We analyze the pair-wise error probability (PEP) performance of different detectors and propose two kinds of user selection methods to minimize the PEP. To facilitate the detection performance evaluation, we further investigate the multiuser diversity of the proposed selection methods and derive a tight lower PEP bound for the case with randomly selected users. From the analytical and numerical results, we demonstrate that the proposed multiuser selection methods can suppress the inter-user interference effectively and achieve better performance than that of the state-of-the-art method.**

***Index Terms*—Multiple-input multiple-output (MIMO), signal detection, nonorthogonal multiple access (NOMA), user selection, pair-wise error probability (PEP).**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) has been studied to improve the spectral efficiency of wireless systems by exploiting the spatial diversity. In a MIMO system, the method to detect signals is one of the most crucial technologies and there are several well-known detectors. For example, the maximum likelihood (ML) detectors [1], zero-forcing (ZF) detectors [2], minimum-mean-square-error (MMSE) detectors [3] and other detectors (such as ZF-SIC [4], MMSE-SIC [5] and LR-based detectors [6]).

Although the receiver diversity gain can be realized by various MIMO detectors, user selection criteria are also necessarily required to exploit the multiuser diversity for a given multiuser MIMO system. Since the channel conditions of different users in the system could be different (e.g., channel gains and cross-relation coefficients), it is reasonable to select certain users with desirable channel conditions to improve the throughput/detection performance. Various user selection strategies have been designed to exploit the multiuser diversity [7]–[10]. In [7], the multiuser diversity has been analyzed for the multiuser MIMO systems based on different user selection criteria and signal detectors. Performances of different multiuser selection criteria have been illustrated in [8], where the system throughout can be improved with the proposed user-selection strategy [9]. To optimize the pair-wise error probability (PEP) performance, the optimal user selection criterion for MIMO systems has been presented in [10], where the maximum distance (MDist)-based user criterion can exploit the full multiuser diversity. All these studies consider exploiting the spatial/multiuser diversity [11] in the system where the number of the transmitted data streams is no more than that of the antennas at the base station (BS).

On the other hand, to further improve the communication efficiency, non-orthogonal multiple access (NOMA) has been recognized as a crucial technology in the future communication systems. It is known that extra degrees of freedom can be provided by NOMA [12]. In particular, for the MIMO system, NOMA can be used to improve the spectral efficiency (i.e., the sum rate) through superposition coding [14]–[17]. In [18]–[20], MIMO-NOMA systems have been designed to guarantee the user fairness under a multiuser cell scenario. Specifically, in downlink MIMO-NOMA, the BS sends superposed signals ($\mathbf{s}_1$ and $\mathbf{s}_2$) to the selected users while the multi-antenna users transmit messages to the BS simultaneously in uplink transmissions. Then, the whole communication system can exploit the multiple-antenna diversity to make the spaces of signals of different user groups (two users in a group) project into different orthogonal signal spaces [20]. However, in [20], the number of the antennas equipped at the users N should satisfy a special requirement (i.e., $N \geq M/2$), where M denotes the number of antennas equipped at the BS. In addition, the beamforming

Manuscript received March 27, 2019; revised August 2, 2019 and October 7, 2019; accepted December 1, 2019. Date of publication December 18, 2019; date of current version February 12, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0503002 and in part by the National Natural Science Foundation under Grant 61922010. The review of this article was coordinated by Dr. J. Liu. *(Corresponding author: Jingchao Wang.)*

Q. Yu is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China, and also with Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: quanyu@ieee.org).

C. Han is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: chaohan@buaa.edu.cn).

L. Bai is with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China, and also with the Beijing Laboratory for General Aviation Technology, Beihang University, Beijing 100191, China (e-mail: l.bai@buaa.edu.cn).

J. Wang is with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: wangjc61s@163.com).

J. Choi is with the School of Information Technology, Deakin University, Geelong, VIC 3220, Australia (e-mail: jinho.choi@deakin.edu.au).

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2019.2960591

(which is one kind of maximal ratio combining methods) at the user side is needed to separate the signals for different user clusters. In other words, when $N < M/2$ or the beamforming is not available to users, the proposed methods in [18]–[20] are unsuitable because of the incorrect signal estimation. In [20], a complicated interaction process between the BS and certain users is needed for MIMO-NOMA beamforming, which might be impractical for the system with highly dynamic users or base stations (BSs). Since the number of data streams is larger than that of the antennas at the BS, the inter-user interference could be the major factor degrading the detection performance. In [21], a mode selection strategy was proposed to improve the system capacity under a multi-user scenario with inter-user interference. In [22], the impact of inter-user interference on the performance of body sensor networks was well analyzed. However, how to suppress the inter-user interference with a reasonable user selection criterion in the MIMO-NOMA system has not been well addressed. Therefore, a more reasonable method is required to guarantee the signal detection in the multiuser MIMO-NOMA system.

Although the signal-to-interference-noise ratio (SINR) or signal-to-noise ratio (SNR) has been employed to select the optimal users [23] for achieving higher throughput, with a receiver/BS constraint, such a throughput might not be used as the performance metric in practice [8], [24]–[29]. For example, although the similar channel capacity could be achieved by different user access schemes, the signal detection performance at the receiver might be different [7], [8], [10]. Suppose that the throughput for user k is $T_k = R_k(1 - P_{\mathrm{Err},k})$ where R_k and $P_{\mathrm{Err},k}$ are the transmit rate and signal error probability (SEP) of user k, respectively [8]. Then, we can see that $P_{\mathrm{Err},k}$ is equivalent to the throughput efficiency T_k/R_k. However, when practical (or non-ideal) issues at the receiver (e.g., suboptimal signal detectors and data-rate requirements) are taken into consideration, the error probability (EP) could be adopted as a more reasonable indicator since the SINR-based throughput might not reflect the detection performance. Therefore, in this work, the error probability (EP) is considered to be a promising candidate for multiuser selection since it is an ultimate performance index of a digital communication system. We consider utilizing the PEP (or SEP) as the performance indicator of the whole system, which will be illustrated in detail in the following parts. However, since the criteria mentioned above are usually targeted for the point-to-point system, it is challenging to exploit the current multiuser selection approaches for a complicated MIMO-NOMA system with the PEP (or SEP) as the performance metric.

In this paper, an uplink multiuser MIMO-NOMA communication system is considered. Different from the traditional MIMO-multiple-access-channel (MIMO-MAC) system [30], we focus on the MIMO-NOMA detection system and mainly study the multiuser selection criterion designation. For the performance indicator, since the PEP is more practical than the information-theoretic throughput, we analyze the system PEP performance of different detectors. Furthermore, as the inter-user interferences cannot be completely eliminated, the detection performance for each user can be affected by the channel conditions of the two selected users simultaneously (which corresponds to a mutual coupling problem analyzed in the following sections). Then, by simplifying the coupling problem, several multiuser selection criteria are proposed to improve the detection performance, and the interference can be suppressed effectively. To evaluate the detection performance, the multiuser diversity is also investigated for the proposed selection methods and a tight lower PEP bound is derived when users are selected randomly. Simulation results show that our proposed selection methods can provide a performance improvement by exploiting the multiuser diversity. Different from the systems illustrated in [8] and [10], the signal detection in our proposed system is more complicated and the user selection criteria are challenging to be designed. Moreover, in our work, we mainly consider designing the user selection criteria for the MIMO-NOMA system without user-side precoding. The contributions of this paper can be summarized as follows:

- We first analyze the PEP upper bound for the MIMO-NOMA system using different detectors, and study the impact of inter-user interference on the PEP performance.
- Based on the derived PEP expression, different reasonable user selection criteria are proposed to improve the detection performance. According to different user selection criteria, the multiuser diversity is also investigated in this paper. The proposed multiuser selection strategies can provide a performance improvement by exploiting the multiuser diversity. Compared with conventional user pairing methods, it is shown that the proposed selection methods are more attractive to the presented multiuser MIMO-NOMA system.
- To evaluate the detection performance, the PEP performance under high SNR is analyzed and a tight lower bound on the average PEP is also calculated under the circumstance where users are randomly selected to communicate with the BS. Through the derived PEP expression, we can analyze the detection performance instead of utilizing the time-consuming simulation.

The remainder of this paper is organized as follows. System model of the uplink multiuser MIMO-NOMA system is introduced in Section II. The PEP upper bound of different detectors is investigated and different user selection criteria are proposed in Section III. In Section IV, the multiuser diversity of the system is studied and the lower bound on the PEP for the system with randomly selected users is evaluated. Simulation results are presented in Section V. Finally, Section VI gives the conclusion of this paper.

Notations: Let $\mathbf{H}$ denote a matrix while $\mathbf{h}$ is a vector. $\mathbf{H}^{\mathrm{T}}$ represents the transpose of $\mathbf{H}$, and $\mathbf{H}^{\mathrm{H}}$ denotes the Hermitian transpose of $\mathbf{H}$. $\|\mathbf{A}\|$ denotes the Frobenius norm of $\mathbf{A}$, $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$, $\|\mathbf{a}\|$ is the Euclidean norm, and $|a|$ returns the absolute value of a. Denote $\mathrm{E}[\cdot]$ the mathematical expectation and $\mathrm{Var}(\cdot)$ the mathematics variance. $\mathbf{I}$ denotes an identity matrix. $P(x)$ denotes the probability of event x. For set $\mathcal{A}$, $|\mathcal{A}|$ represents the cardinality of $\mathcal{A}$. Let $F(x) = \Pr(X \leq x)$ be the cumulative distribution function (CDF) of a random variable x, and its probability density function (PDF) of x is denoted by $f(x) = [F(x)]'$. The main symbols in this paper are summarized in Table I.

TABLE I
SUMMARY OF MAIN SYMBOLS USED IN THE PAPER

Symbols	Descriptions
M (or N)	The number of antennas equipped at the BS (or each user)
$\mathcal{G}_1$ (or $\mathcal{G}_2$)	User set including near (or far) users
u_k (or u_p)	A single user from $\mathcal{G}_1$ (or $\mathcal{G}_2$)
$\mathbf{H}_k$ (or $\mathbf{H}_p$)	Channel matrix of u_k (or u_p)
$\mathbf{M}_k$ (or $\mathbf{M}_p$)	Rayleigh fading channel matrix for u_k (or u_p)
N_0	Variance of the noise
d_k (or d_p)	Distance between the BS and u_k (or u_p)
$L(d_k)$ (or $L(d_p)$)	Path loss between the BS and u_k (or u_p)
α; $\mathcal{S}$	Attenuation factor; Signal constellation table
$\mathbf{s}$; $\tilde{\mathbf{s}}_k$; $\mathbf{s}_k$ ($\mathbf{s}_p$)	Transmitted signals; Estimated signals; Signals transmitted from u_k (or u_p)
$\mathbf{n}; T$	Noise; Channel static time
σ_u^2	Variance of the variables in $\mathbf{M}_k$ and $\mathbf{M}_p$
σ_s^2	Power of the transmitted signal
$\mathbf{W}_k$, $\mathbf{W}_p$	Linear detectors for u_k and u_p
$\mathbf{d}_k$; $\mathbf{d}_p$	$\mathbf{s}_k - \hat{\mathbf{s}}_k$; $\mathbf{s}_p - \hat{\mathbf{s}}_p$
$\mathbf{t}$	$\mathbf{H}_p\mathbf{s}_p$
$\mathcal{D}$	The error space of $\mathbf{s}_k$ (or $\mathbf{s}_p$)
$\lambda(\mathbf{A})$	The eigenvalues of matrix $\mathbf{A}$
$\lambda_{\max}(\mathbf{H})$, $\lambda_{\min}(\mathbf{H})$	The maximum eigenvalue and minimum eigenvalue of $\mathbf{H}$
$\mathcal{A}$ (or $\mathcal{L}$)	The distribution area of eigenvalues of random gaussian matrices (or Wishart matrices)
X; Z	$L(d_k)\lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})$; $L(d_p)\lambda_{\max}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})$
U; V	$\min\{Z_{p=K+1}, \cdots, Z_{p=K+P}\}$; $\max\{X_1, \cdots, X_K\}$
$\mathcal{D}_k$	Equivalent multiuser diversity for u_k
$w_{i,j}$	The element located in the i-th row and j-th column of $\tilde{\mathbf{W}} = \mathbf{M}_k\mathbf{M}_p$
κ	$\kappa = \lambda(\mathbf{Q})$ where $\mathbf{Q}$ is the complex Wishart matirx
$\mathrm{B}(x, y)$	The Beta function

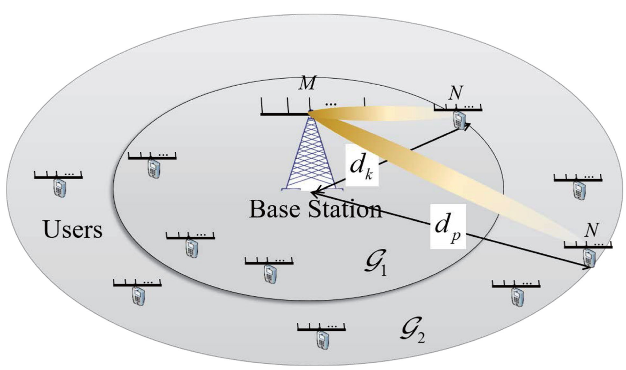


Fig. 1. A MIMO-NOMA system with multiple users.

II. SYSTEM MODEL

In this paper, an uplink multiuser MIMO-NOMA system model is considered which is illustrated in Fig. 1. The BS is equipped with M antennas. In the MIMO-NOMA system, for simplicity, we consider that two multi-antenna users can be selected to access to the BS simultaneously. It is worthy to note that the presented model can be connected to various MIMO systems such as the internet of things (IoT) and internet of vehicles. Here, it is assumed that the number of antennas at each user is N, and the users can be divided into two groups, i.e., $\mathcal{G}_1$ and $\mathcal{G}_2$. Compared with the users in $\mathcal{G}_2$, users in $\mathcal{G}_1$ are much closer to the BS. Note that in this paper, two user groups are considered and it is easy to be extended to multi-group communication systems. For the sake of convenience, we assume that users u_k and u_p are selected from $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, in this paper.

Thus, the received signals at the BS can be given as [31]

$$\mathbf{y} = \mathbf{H}_k\mathbf{s}_k + \mathbf{H}_p\mathbf{s}_p + \mathbf{n}, \tag{1}$$

where $\mathbf{H}_k \in \mathbb{C}^{M\times N}$ is the channel matrix of u_k in $\mathcal{G}_1$. $\mathbf{H}_p$ is the channel matrix of u_p in $\mathcal{G}_2$ where $p > |\mathcal{G}_1|$, noise $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ is a circularly symmetric complex gaussian (CSCG) variable. In Eqn. (1), $\mathbf{H}_k = \frac{\mathbf{M}_k}{\sqrt{L(d_k)}}$ where $\mathbf{M}_k$ is a Rayleigh fading channel matrix [32]. The distance between the BS and u_k (or u_p) is denoted by d_k (or d_p) while $L(d_k) = d_k^{\alpha}$ is the path loss. Here, as mentioned in [14], we assume attenuation factor $\alpha = 3$ throughout this paper. Given a signal constellation table $\mathcal{S}$, vectors $\mathbf{s}_k, \mathbf{s}_p \in \mathcal{S}$ denote the signals which are transmitted from u_k and u_p, respectively. Here, the channel state information is known at the BS. Then, we have

$$\mathbf{y} = \frac{\mathbf{M}_k}{\sqrt{L(d_k)}}\mathbf{s}_k + \frac{\mathbf{M}_p}{\sqrt{L(d_p)}}\mathbf{s}_p + \mathbf{n}, \tag{2}$$

where d_p is larger than d_k. In Eq. (2), we assume that the channel is static during a regular symbol interval (e.g., no more than T symbols). For different users, their channel matrices are independent of each other statistically. In the following sections, $\sigma_u^2 = 1$ is used as the equivalent variance of the CSCG variables in $\mathbf{M}_k$ and $\mathbf{M}_p$ while the power of the transmitted signal is equal to σ_s^2.

III. MULTIUSER SELECTION CRITERIA

In this section, we mainly focus on the design of the user selection criterion and the performance analysis of the signal detection at the BS with different signal detectors, we also discuss the impact of the user pairing on the PEP of the detectors. Note that there are many methods which have been proposed to detect transmitted signals. In this section, we consider two typical detectors (e.g., the optimal ML detector and the linear ZF detector). After brief analysis, different reasonable selection criteria are proposed (i.e., MEI and MED methods, illustrated in the following parts) for the MIMO-NOMA system.

From Eqn. (2), based on the ML method given in [10], $\mathbf{s}_k$ is detected by[1]

$$\tilde{\mathbf{s}}_k = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{y} - \mathbf{H}_k\mathbf{s}\|^2. \tag{3}$$

To obtain the estimation of $\mathbf{s}_p$, we firstly assume that the signals transmitted from u_k have been estimated. Then, according to Eqn. (3), the output of the ML detector becomes

$$\tilde{\mathbf{s}}_p = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{y} - \mathbf{H}_k\tilde{\mathbf{s}}_k - \mathbf{H}_p\mathbf{s}\|^2. \tag{4}$$

[1]Here, we assume that signals transmitted from u_k are detected at first.

Similarly, for the linear detector, we have

$$\tilde{\mathbf{s}}_k = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{W}_k\mathbf{y} - \mathbf{s}\|^2 \tag{5}$$

and

$$\tilde{\mathbf{s}}_p = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{W}_p(\mathbf{y} - \mathbf{H}_k\tilde{\mathbf{s}}_k) - \mathbf{s}\|^2, \tag{6}$$

where $\mathbf{W}_k$ and $\mathbf{W}_p$ are the linear detectors of different signals (from u_k and u_p). According to the above information, Eqns. (3) and (4) (as well as Eqs. (5) and (6)) can be viewed as a two-step detection process. Although different types of detectors can be employed during the process, we only consider the most typical modes in this paper, e.g., the ML-ML detector (utilizing ML detectors to estimate signals $\mathbf{s}_k$ and $\mathbf{s}_p$) and the ZF-ZF/MMSE-MMSE detector. From Eqns. (4) and (6), we can learn that the BER performance for $\mathbf{s}_p$ depends on not only the channel matrix $\mathbf{H}_p$, but also the BER performance for $\mathbf{s}_k$, which is dependent on the channel matrix $\mathbf{H}_k$. Thus, to derive a reasonable multiuser selection criterion for the system, the relationship between the channel matrices and the BER performance should be clarified, which will be analyzed in the following parts.

A. Detection With the ML-ML Detector

Inspired by [10], with the ML-ML detector, the PEP of $\mathbf{s}_k$ can be given as

$$\begin{aligned} P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) &= P(\|\mathbf{y} - \mathbf{H}_k\hat{\mathbf{s}}_k\|^2 \le \|\mathbf{y} - \mathbf{H}_k\mathbf{s}_k\|^2) \\ &= P(\|\mathbf{H}_k\mathbf{d}_k + \mathbf{H}_p\mathbf{s}_p + \mathbf{n}\|^2 \le \|\mathbf{H}_p\mathbf{s}_p + \mathbf{n}\|^2) \\ &= P(\|\mathbf{H}_k\mathbf{d}_k\|^2 + 2\langle \mathbf{t} + \mathbf{n}, \mathbf{H}_k\mathbf{d}_k\rangle \le 0), \end{aligned} \tag{7}$$

where $\hat{\mathbf{s}}_k$ is an erroneously detected signal vector, $\mathbf{d}_k = \mathbf{s}_k - \hat{\mathbf{s}}_k$, $\langle \mathbf{t} + \mathbf{n}, \mathbf{H}_k\mathbf{d}_k\rangle = \mathrm{Re}\{(\mathbf{t}+\mathbf{n})^{\mathrm{H}}\mathbf{H}_k\mathbf{d}_k\}$, and $\mathbf{t} = \mathbf{H}_p\mathbf{s}_p$. $\mathrm{Re}\{\mathbf{n}\}$ returns the real part of $\mathbf{n}$. Letting $\mathbf{x} = \langle \mathbf{t} + \mathbf{n}, \mathbf{H}_k\mathbf{d}_k\rangle$, we have

$$\mathbf{x} \sim \mathcal{N}\left(\mathrm{Re}\{(\mathbf{H}_k\mathbf{d}_k)^{\mathrm{H}}\mathbf{t}\}, \frac{N_0}{2}\|\mathbf{H}_k\mathbf{d}_k\|^2\mathbf{I}\right),$$

where $\mathcal{N}(a, b)$ stands for a guassian distribution with mean a and variance b. Then, there exists

$$P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) = \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2 + 2\,\mathrm{Re}\{\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{t}\}}{\sqrt{2N_0\|\mathbf{H}_k\mathbf{d}_k\|^2}}\right), \tag{8}$$

where $\mathcal{Q}(x) = \int_x^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\mathrm{d}t$. For $\mathbf{s}_p$, with the same method, we can derive

$$\begin{aligned} &P(\mathbf{s}_p \to \hat{\mathbf{s}}_p \mid \mathbf{s}_k \to \hat{\mathbf{s}}_k) \\ &= P(\|\mathbf{y} - \mathbf{H}_k\hat{\mathbf{s}}_k - \mathbf{H}_p\hat{\mathbf{s}}_p\|^2 \le \|\mathbf{y} - \mathbf{H}_k\hat{\mathbf{s}}_k - \mathbf{H}_p\mathbf{s}_p\|^2) \\ &= \mathcal{Q}\left(\frac{\|\mathbf{H}_p\mathbf{d}_p\|^2 + 2\,\mathrm{Re}\{\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{d}_k\}}{\sqrt{2N_0\|\mathbf{H}_p\mathbf{d}_p\|^2}}\right), \end{aligned} \tag{9}$$

where $\mathbf{d}_p = \mathbf{s}_p - \hat{\mathbf{s}}_p$.

To find the optimal users, one feasible strategy is to minimize the upper bound on the PEP [25], [29]. From Eqn. (9), we can see that the multiuser criterion is hard to find directly. Then, we consider utilizing the approximation of the PEP to derive a reasonable criterion.

In Eqn. (8), according to the discrete characteristic of the $\mathbf{s}_p$ (for example, if the M-ary PSK is employed to the system, $\mathbf{s}_p$ can be regarded as a special discrete Gaussian random variable (RV) and can be assumed that $\mathbf{s}_p \sim \mathcal{CN}(\mathbf{0}, \sigma_s^2\mathbf{I})$), we have $X = 2\,\mathrm{Re}\{\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{t}\} \sim \mathcal{N}(0, 2\,\|\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2\sigma_s^2)$. Then, the PEP can be approximated as

$$\mathrm{E}_{\mathbf{s}_p}[P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)] = \int_X \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2 + x}{\sqrt{2N_0\|\mathbf{H}_k\mathbf{d}_k\|^2}}\right) f_X(x)\mathrm{d}x, \tag{10}$$

where $f_X(x)$ is the PDF of the RV X. According to (Eqn. (3.66) in [33]), we have

$$\mathrm{E}_Z[\mathcal{Q}(u + \lambda z)] = \mathcal{Q}\left(\frac{u}{\sqrt{1+\lambda^2}}\right),\ Z \sim \mathcal{N}(0, 1). \tag{11}$$

Combining Eqns. (10) and (11), it can be derived that

$$\begin{aligned} &P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) \\ &\approx \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2/\sqrt{2N_0\|\mathbf{H}_k\mathbf{d}_k\|^2}}{\sqrt{1 + (\sigma_s^2\|\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2)/(N_0\|\mathbf{H}_k\mathbf{d}_k\|^2)}}\right) \\ &= \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{\sqrt{2(N_0\|\mathbf{H}_k\mathbf{d}_k\|^2 + \sigma_s^2\|\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2)}}\right). \end{aligned} \tag{12}$$

Based on the total probability theorem, $P(\mathbf{s}_p \to \hat{\mathbf{s}}_p)$ becomes

$$\sum_{\mathbf{d}_k\in\mathcal{D}} P(\mathbf{s}_p \to \hat{\mathbf{s}}_p \mid \mathbf{d}_k)P(\mathbf{d}_k), \tag{13}$$

where $\mathcal{D}$ denotes the error space of $\mathbf{s}_k$ (or $\mathbf{s}_p$). In order to obtain the optimal multiuser selection criterion for the uplink communication system, certain users need to be chosen as follows [8]:

$$\{k^\star, p^\star\} = \arg\min_{k,p}\max_{\mathbf{d}_k,\mathbf{d}_p\in\mathcal{D}}(BER_k + BER_p), \tag{14}$$

where $BER_k = \sum_{\mathbf{s}_k\in\mathbf{s}} P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)$ and $BER_p = \sum_{\mathbf{s}_p\in\mathbf{s}} P(\mathbf{s}_p \to \hat{\mathbf{s}}_p)$, respectively. However, the problem in Eqn. (14) is difficult to be solved directly. This is because that the value of $P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)$ and the value of $P(\mathbf{s}_p \to \hat{\mathbf{s}}_p)$ are coupled to each other. Here, we consider the loose coupling and obtain the following conclusion.

Proposition 1: Suppose that $|\mathcal{G}_1|, |\mathcal{G}_2| \ll \infty$ and signals transmitted from u_k are detected first. Then, under high SNR, for the ML-ML detector, to optimize the PEP performance of the communication system, the problem above can be simplified as

$$\{k^\star, p^\star\} = \arg\min_{k,p}\max_{\mathbf{d}_k,\mathbf{d}_p\in\mathcal{D}}(PEP_k),$$

where $PEP_k = P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)$.

Proof: Please see Appendix A. ■

According to Proposition 1, the group pairing problem could be modified as

$$\{k^\star, p^\star\} = \arg\max_{k,p}\min_{\mathbf{d}_k\in\mathcal{D}}\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{\sqrt{2\mathbf{d}_k^{\mathrm{H}}(N_0\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_s^2\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k)\mathbf{d}_k}}$$

and the strategy presented above is the so-called MDist criterion [8]. When the SNR is high, the criterion becomes

$$\{k^\star, p^\star\} = \arg\max_{k,p} \min_{\mathbf{d}_k \in \mathcal{D}} \frac{\|\mathbf{H}_k \mathbf{d}_k\|^2}{\sqrt{2\mathbf{d}_k^{\mathrm{H}}(\sigma_s^2 \mathbf{H}_k^{\mathrm{H}} \mathbf{H}_p \mathbf{H}_p^{\mathrm{H}} \mathbf{H}_k)\mathbf{d}_k}}. \tag{15}$$

Based on the max-min eigenvalue (ME) criterion [8], we have the upper bound on the PEP as follows:

$$P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) \leq \mathcal{Q}\left(\frac{\|\mathbf{d}_k\| \lambda_{\min}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)}{\sqrt{2\lambda_{\max}(N_0 \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_s^2 \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k)}}\right). \tag{16}$$

Here, $\lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(\mathbf{H})$ return the maximum eigenvalue and minimum eigenvalue of $\mathbf{H}$. Then, we can find the optimal $\mathbf{H}_k$ and $\mathbf{H}_p$ from all candidates to minimize $\frac{\lambda_{\min}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)}{\sqrt{2\lambda_{\max}(N_0 \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_s^2 \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k)}}$. This problem could be solved by exhaust algorithms, however, the computational complexity could be high. Besides, given two independent matrices $\mathbf{H}_k$ and $\mathbf{H}_p$, no significant analytic relationship could be found between $\lambda(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)$ and $\lambda(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k)$ [34], where $\lambda(\mathbf{A})$ denotes the eigenvalue of the $\mathbf{A}$. Alternatively, since $\|\mathbf{H}_k\mathbf{d}_k\|^2 \leq \|\mathbf{d}_k\|^2 \lambda_{\max}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)$, the PEP in Eqn. (12) is upper bounded as

$$\begin{aligned} P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) &= \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{\sqrt{2(\mathbf{H}_k\mathbf{d}_k)^{\mathrm{H}}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})(\mathbf{H}_k\mathbf{d}_k)}}\right) \\ &\leq \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|}{\sqrt{2\lambda_{\max}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p)}}\right) \\ &\leq \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|\sqrt{\lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})}}{\sqrt{2\lambda_{\max}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p)}}\right). \end{aligned} \tag{17}$$

Therefore, a suboptimal multiuser selection criterion for the ML-ML detector can be given by

$$k^\star = \arg\max_{k} \lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}), \tag{18a}$$

$$p^\star = \arg\min_{p} \lambda_{\max}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}). \tag{18b}$$

Thus, at first, the BS can choose the closer one whose $\lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})$ is maximal. Meanwhile, another user who has the minimum $\lambda_{\max}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})$ will also be selected to access to the BS. Here, the criterion above is referred as the max-min eigenvalue after decoupling (MED) criterion.

B. *Detection With Linear Detectors*

The ML method is regarded as the optimal detector. However, its computational complexity is extremely high. In this case, low-complexity detectors (e.g., MMSE and ZF) are designed for the signal detection. In this subsection, we analyze the PEP of linear detectors and derive the multiuser selection criterion accordingly.

Suppose that the MMSE-MMSE[2] detector is employed to estimate the user messages, the obtained signals are $\tilde{\mathbf{s}}_k = \mathrm{Dec}_{\mathcal{S}}(\mathbf{W}_k\mathbf{y})$ and $\tilde{\mathbf{s}}_p = \mathrm{Dec}_{\mathcal{S}}(\mathbf{W}_p(\mathbf{y} - \mathbf{H}_k\tilde{\mathbf{s}}_k))$ where $\mathbf{W}_k = (\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + N_0\mathbf{I})^{-1}\mathbf{H}_k^{\mathrm{H}}$ and $\mathbf{W}_p = (\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p + N_0\mathbf{I})^{-1}\mathbf{H}_p^{\mathrm{H}}$. Here, $\mathrm{Dec}_{\mathcal{S}}(\mathbf{x}) = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{x} - \mathbf{s}\|^2$. Then, for the ZF-ZF detector, the PEP is given by

$$\begin{aligned} P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) &= P(\|\mathbf{W}_k\mathbf{y} - \hat{\mathbf{s}}_k\|^2 \leq \|\mathbf{W}_k\mathbf{y} - \mathbf{s}_k\|^2) \\ &= \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|^2 + 2\,\mathrm{Re}\{\mathbf{d}_k^{\mathrm{H}}\mathbf{W}_k\mathbf{H}_p\mathbf{s}_p\}}{\sqrt{2N_0\|\mathbf{W}_k^{\mathrm{H}}\mathbf{d}_k\|^2}}\right), \end{aligned} \tag{19}$$

$$\begin{aligned} &P(\mathbf{s}_p \to \hat{\mathbf{s}}_p \mid \mathbf{s}_k \to \hat{\mathbf{s}}_k) \\ &= P(\|\mathbf{W}_p(\mathbf{y} - \mathbf{H}_k\hat{\mathbf{s}}_k) - \hat{\mathbf{s}}_p\|^2 \leq \|\mathbf{W}_p(\mathbf{y} - \mathbf{H}_k\hat{\mathbf{s}}_k) - \mathbf{s}_p\|^2) \\ &= P(\|\mathbf{W}_p\mathbf{H}_k\mathbf{d}_k + \mathbf{d}_p + \mathbf{W}_p\mathbf{n}\|^2 \leq \|\mathbf{W}_p\mathbf{H}_k\mathbf{d}_k + \mathbf{W}_p\mathbf{n}\|^2) \\ &= \mathcal{Q}\left(\frac{\|\mathbf{d}_p\|^2 + 2\,\mathrm{Re}\{\mathbf{d}_p^{\mathrm{H}}\mathbf{W}_p\mathbf{H}_k\mathbf{d}_k\}}{\sqrt{2N_0\|\mathbf{W}_p^{\mathrm{H}}\mathbf{d}_p\|^2}}\right), \end{aligned} \tag{20}$$

where $\mathbf{d}_p = \mathbf{s}_p - \hat{\mathbf{s}}_p$. According to Eqns. (10)–(12), Eqn. (19) can be approximated as

$$P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) \approx \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|^2}{\sqrt{2(N_0\|\mathbf{W}_k^{\mathrm{H}}\mathbf{d}_k\|^2 + \sigma_s^2\|\mathbf{d}_k^{\mathrm{H}}\mathbf{W}_k\mathbf{H}_p\|^2)}}\right) \tag{21a}$$

$$= \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|^2}{\sqrt{2(\mathbf{d}_k^\star)^{\mathrm{H}}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})(\mathbf{d}_k^\star)}}\right), \tag{21b}$$

where $\mathbf{d}_k^\star = \mathbf{W}_k^{\mathrm{H}}\mathbf{d}_k$. From Eqn. (21a), we can easily get the suboptimal multiuser selection criterion for the ZF-ZF detector as follows:

$$\{k^\star, p^\star\} = \arg\min_{k,p} \lambda_{\max}(N_0\mathbf{W}_k\mathbf{W}_k^{\mathrm{H}} + \sigma_s^2\mathbf{W}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{W}_k^{\mathrm{H}}). \tag{22}$$

When $N_0 \to 0$, the criterion becomes

$$\{k^\star, p^\star\} = \arg\min_{k,p} \lambda_{\max}(\mathbf{W}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{W}_k^{\mathrm{H}}). \tag{23}$$

Here, the criterion in Eqn. (23) is referred to as the min-max eigenvalue of inter-user interference (MEI) criterion. Similarly, for the ML-ML detector (based on Eqn. (15)), the interference part becomes $\|\mathbf{d}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2$. Then, the MEI criterion for ML-ML detector is given by

$$\{k^\star, p^\star\} = \arg\min_{k,p} \lambda_{\max}(\mathbf{H}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}). \tag{24}$$

Meanwhile, according to Eqn. (21b), we have

$$(21\text{b}) \leq \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|^2\|\mathbf{d}_k^\star\|^{-1}}{\sqrt{2\lambda_{\max}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}}\right).$$

Since $\|\mathbf{W}_k^{\mathrm{H}}\mathbf{d}_k\|^{-1} \geq \|\mathbf{W}_k^{\mathrm{H}}\|^{-1}\|\mathbf{d}_k\|^{-1}$, it can also be shown that

$$(21\text{b}) \leq \mathcal{Q}\left(\frac{\|\mathbf{d}_k\|\|\mathbf{H}_k^{\mathrm{H}}\|}{\sqrt{2\lambda_{\max}(N_0\mathbf{I} + \sigma_s^2\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}}\right).$$

[2]Since the MMSE detector will become the ZF detector as SNR $\to \infty$, we mainly analyze the ZF detector in the following parts.

For the ZF-ZF detector, the MED criterion is same to the criterion in Eqn. (18).

Based on Eqns. (20), (22), and (23), we find that minimizing $\lambda_{\max}(\mathbf{W}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{W}_k^{\mathrm{H}})$ can help improve both the PEP_k and PEP_p. In addition, as stated in Proposition 1, under the circumstance where $K=|\mathcal{G}_1|$ and $P=|\mathcal{G}_2|$ are limited, Eqn. (23) will be a considerable multiuser selection criterion for the ZF-ZF detector.[3]

IV. Impact of the Number of Different User Groups

In this section, to see the impact of the number of different user groups (i.e., $\mathcal{G}_1, \mathcal{G}_2$) on the detection performance for the proposed criteria, we study the performance analysis of the detection system with the proposed multiuser selection methods.

A. Multiuser Diversity of Different Detectors With the Proposed Criteria

Since the inter-user interference and the noise are the critical factors to the detection performance, to derive the multiuser diversity gain, we firstly analyze the system performance in this part.

Given random gaussian matrices (RGMs) $\mathbf{M}_k$ and $\mathbf{M}_p$, and further assuming $M=N$ for the sake of simplicity, the eigenvalues of the RGM [34], [35] are randomly distributed over the area of

$$\mathcal{A}=\{(x,iy);\ |x|,|y|\leq\sqrt{N}\},\tag{25}$$

while the eigenvalues of the Wishart matrix (e.g., $\mathbf{M}_k\mathbf{M}_k^{\mathrm{H}}$) [36] are distributed over the following support:

$$\mathcal{L}=\{0\leq x\leq 4\,N\}.\tag{26}$$

In addition, for the $X=L(d_k)\lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})$ and $Z=L(d_p)\lambda_{\max}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})$, we can know that the corresponding PDF is $f_X(x)=Ne^{-Nx}$, $x\leq\infty$ [36]. The CDF of RV Z is $F_Z(z)=e^{-\frac{(z-4\,N)^3}{192\,N}}$ [37]. Then, let $V=\max\{X_1,\ldots,X_K\}$ and $U=\min\{Z_{p=K+1},\ldots,Z_{p=K+P}\}$. The CDF of V and U can be given by $f_V(v)=KN(1-e^{-Nv})^{K-1}e^{-Nv}$ and $f_U(u)=[1-(1-F_Z(u))^P]'$, respectively.

Definition 1: Given an uplink multiuser detection system where more than two users can access the BS simultaneously. Then, for u_k, the equivalent multiuser diversity can be defined as[4]

$$\mathcal{D}_k=-\frac{\mathrm{d}PEP_k}{\mathrm{d}G},$$

where G denotes the SNR.

Based on the proposed multiuser selection criterion, we can have the following results.

Remark 1: Given u_k and $|\mathcal{G}_2|=P$, based on the MED criterion in Eqn. (18), it can be proved that the PEP are upper bounded as $P(\mathbf{s}_k\rightarrow\hat{\mathbf{s}}_k)\leq[1-F_Z(0)]^P\mathcal{Q}(\eta_1)$, where $\eta_1\in[0,\sqrt{\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{2N_0}}]$.

Proof: Please see Appendix B. ∎

Assume that $\mathbf{s}_k$ is firstly detected, we can see that the PEP of the detector can decrease as $|\mathcal{G}_2|$ grows. Since $F_Z(0)=e^{-\frac{N^2}{3}}$, when the antenna number N is large and u_k is fixed, the system performance will not be improved obviously as $|\mathcal{G}_2|$ grows.

Remark 2: Given $|\mathcal{G}_1|=K$ and u_p, based on the MED criterion in Eqn. (18), it can be proved that the PEP are upper bounded as $P(\mathbf{s}_k\rightarrow\hat{\mathbf{s}}_k)\leq[1-m]^K\mathcal{Q}(\eta_2)$, where $m=e^{-\frac{2Na_k^2L(d_k)(N_0+\sigma_s^2\lambda_p)}{\|\mathbf{d}_k\|^2}}, 0\leq\eta_2\leq\frac{\|\mathbf{d}_k\|\sqrt{\lambda_{\min}(\mathbf{M}_k^{\mathrm{H}}\mathbf{M}_k)}}{\sqrt{2L(d_k)(N_0+\sigma_s^2\lambda_p)}}, a_k=\max_k\{\lambda_{\min}(\mathbf{M}_k^{\mathrm{H}}\mathbf{M}_k)\}$ and $\lambda_p=\lambda_{\max}(\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p)$ is fixed.

Proof: We omit the detailed proof here since it is similar to the proof of Remark 1. ∎

Similarly, when u_p is given in advance, we can see that the PEP of the detector can decrease as $|\mathcal{G}_1|$ grows. However, since $m\sim e^{-N^2}$, the system performance will not be improved obviously when the antenna number N increases.

Theorem 1: Considering the circumstance where limited number of users exists, based on the proposed MED criterion, then $\mathcal{D}_k$ is no less than $|\mathcal{G}_1|$ while $\mathcal{D}_p\geq 1$.

Proof: Please see Appendix C. ∎

Although the system performance is influenced by the interference term and the noise term simultaneously under low SNR, with limited $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$, we can still obtain considerable detection results based on the proposed criteria. Thus, the inter-user interference can be reduced to a lower level as $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$ increase.

In Section III, we have discussed different multiuser criteria for ML-ML and MMSE-MMSE detectors (i.e., MDist criterion, MED criterion and MEI criterion). Here, we mainly focus on analyzing the detection performance of the given detectors. Firstly, for the ML-ML detector, let $w_{i,j}$ be the element located in the i-th row and j-th column of $\tilde{\mathbf{W}}=\mathbf{M}_k\mathbf{M}_p$, and we utilize Eqn. (24) as the multiuser selection criterion. Then, there exists $w_{i,j}\sim\mathcal{CN}(0,N)$. When $K,P\ll\infty$, based on the MEI criterion, we have (details are given in Appendix D)

$$\mathrm{E}_{\mathbf{H}_p}[P_{\mathrm{ML-ML}}(\mathbf{s}_k\rightarrow\hat{\mathbf{s}}_k)]\leq[1-F_Z(0)]^{D_1}[\mathcal{Q}(\eta_4)],\tag{27}$$

where $\max\{K,P\}<D_1\leq PK, \eta_4\in[0,\sqrt{\frac{\lambda_k\|\mathbf{d}_k\|^2}{2N_0}}]$, and $\lambda_{\min}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)\leq\lambda_k\leq\lambda_{\max}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)$. From Eqn. (27), we can easily find that the upper bound on the average PEP is directly proportional to the number of the users. In addition, u_k with larger $\lambda_{\min}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k)$ should be selected in the premise of the low inter-user interference. Since $\mathbf{M}$ is a Gaussian matrix, based on the MED criterion, there exists (details are given in Appendix E)

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}[P_{\mathrm{ML-ML}}(\mathbf{s}_k\rightarrow\hat{\mathbf{s}}_k)]\leq\frac{2^{K+\frac{1}{2}}N^K\Gamma(K+\frac{1}{2})L(d_p)}{\sqrt{2\pi}\sigma_s^2}\left[\frac{\|\mathbf{d}_k\|^2}{2L(d_k)(N_0L(d_p)+4\sigma_s^2N)}\right]^{-K}[1-F_Z(\eta_3)]^P,\tag{28}$$

[3]That is to say, the minimization of $\lambda_{\max}(\mathbf{W}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{W}_k)$ decides the error floor of the PEP_p while the value of $\lambda_{\max}(\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p)$ affects the performance gap between the PEP_p and PEP_k.

[4]Although the conventional multiuser diversity is defined at SNR $\longrightarrow\infty$, due to the inter-user interference, it may not be practical for MIMO-NOMA systems. Thus, in this subsection, we first analyze the multiuser diversity at low SNR, while the lower bound of the PEP at high SNR will be analyzed in the next subsection.

where $\eta_3 \in [0,\, 4N]$. From Eqn. (28), we can easily see that the upper bound on the average PEP is directly proportional to $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$. When the noise power (i.e., N_0) decreases, the system performance can be improved accordingly. In addition, we can also learn that the upper bound might be enlarged as N grows.

For the ZF-ZF detector, according to Proposition 1, Eqns. (22) and (23), when the number of users is limited, the optimal performance depends only on the $\lambda_{\max}(\mathbf{W}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{W}_k^{\mathrm{H}})$ under high SNR. Then, after employing the MEI criterion, it can be derived that $\mathrm{E}_{\mathbf{H}_p}[P_{\mathrm{ZF-ZF}}(\mathbf{s}_k \to \hat{\mathbf{s}}_k)]$ is similar to Eqn. (28).

B. PEP Approximation With Random u_k and u_p

In this subsection, the PEP performance of different detectors is estimated while the users are selected randomly from $\mathcal{G}_1$ and $\mathcal{G}_2$ (or $K=1, P=1$). For the optimal ML-ML detector, based on Eqn. (21b), we have

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_k \to \hat{\mathbf{s}}_k)\right] \overset{(a)}{\geq} \mathcal{Q}\left(\mathrm{E}_{\mathbf{H}_k}\left[\frac{\|\mathbf{H}_k\mathbf{d}_k\|}{\sqrt{2N_0}}\right]\sqrt{\mathrm{E}_{\mathbf{H}_p}\left[\eta_p\right]}\right) \tag{29}$$

and

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_p \to \hat{\mathbf{s}}_p|\mathbf{s}_k \to \hat{\mathbf{s}}_k)\right] \geq \mathcal{Q}\left(\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[\frac{\|\mathbf{H}_p\mathbf{d}_p\|^2 + 2\,\mathrm{Re}\{\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{d}_k\}}{\sqrt{2N_0\|\mathbf{d}_p\|^2}\sqrt{\lambda(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}}\right]\right)$$
$$= \mathcal{Q}\left(\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[\frac{\|\mathbf{H}_p\mathbf{d}_p\|}{\sqrt{2N_0}} + \frac{2\,\|\mathbf{d}_k\|\sqrt{N}\lambda(\mathbf{H}_{\mathrm{G}})}{\sqrt{2N_0L_p}\sqrt{\lambda(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}}\right]\right), \tag{30}$$

where $\eta_p = \frac{1}{\lambda(\mathbf{I}+\frac{\sigma_s^2}{N_0}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}$ and $\mathbf{H}_{\mathrm{G}} = \frac{1}{2}(\mathbf{H}_k + \mathbf{H}_k^{\mathrm{H}})$. In Eqn. (29), inequality (a) is due to Jensen's inequality. Here, we consider the complex Wishart matirx $\mathbf{Q}$ of the form $\mathbf{Q} = \mathbf{M}^{\mathrm{H}}\mathbf{M}$, where $\mathbf{M}$ is a RGM. According to Eqns. (25), (26), and (Eqn. (1.21) in [34]), the PDF of the $\kappa = \lambda(\mathbf{Q})$ becomes

$$f_\kappa(x) = \frac{1}{2N\pi}\sqrt{4 - \left(\frac{x}{2\,N}\right)^2}, \quad 0 \leq x \leq 4N. \tag{31}$$

Then, we can derive that

$$\mathrm{E}[\sqrt{\lambda(\mathbf{Q})}] = \int_0^{4\,N}\sqrt{x}f_\kappa(x)\mathrm{d}x = \frac{8\sqrt{N}}{\pi}\int_0^1 x^{\frac{1}{2}}\sqrt{1-x^2}\mathrm{d}x$$
$$= \frac{4\sqrt{N}}{\pi}\mathrm{B}\left(\frac{3}{4},\frac{3}{2}\right), \tag{32}$$

where $\mathrm{B}(x,y) = 2\int_0^1 t^{2x-1}(1-t^2)^{y-1}\mathrm{d}t$ is the Beta function [38], [39]. For $\mathrm{E}[\eta_p]$ in Eqn. (29), based on (Eqn. (2.116) in [34]), we have

$$\beta = \frac{1-\mathrm{E}[\eta_p]}{\xi(\tilde{\gamma}\mathrm{E}[\eta_p])},\ \xi(\tilde{\gamma}) = \frac{1}{1+\tilde{\gamma}},\ \tilde{\gamma} = \frac{N\sigma_s^2}{L(d_p)N_0}. \tag{33}$$

Thus, Eqns. (29) and (30) become

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_k \to \hat{\mathbf{s}}_k)\right] \geq \mathcal{Q}\left(\frac{2\sqrt{2\,N}\sqrt{\sqrt{4\tilde{\gamma}+1}-1}\|\mathbf{d}_k\|}{\pi\sqrt{N_0\tilde{\gamma}L(d_k)}}\mathrm{B}\left(\frac{3}{4},\frac{3}{2}\right)\right) \tag{34}$$

and

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_p \to \hat{\mathbf{s}}_p|\mathbf{s}_k \to \hat{\mathbf{s}}_k)\right] \geq \mathcal{Q}\left(\frac{4\,N\|\mathbf{d}_p\|}{\pi\sqrt{N_0L(d_p)}}\mathrm{B}\left(\frac{3}{4},\frac{3}{2}\right)\right), \tag{35}$$

respectively. From Eqns. (34) and (35), we can learn that the lower bound on the PEP performance decreases as N grows. Similarly, for the ZF-ZF detector with the MEI criterion, we can get the same performance bound for the detection system. Note that when the system noise is weak, the PEP for u_p is close to but no less than $P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)$.

The lower bound in Eqn. (35) is loose especially under high SNR. Thus, under high SNR, the original bound may not be suitable for the system, since the denominator in Eqn. (19) is 0 ($N_0 = 0$). Therefore, according to Eqns. (10) and (12), we have

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_k \to \hat{\mathbf{s}}_k)\right] = \mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[\mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{\sqrt{2\|\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2\sigma_s^2}}\right)\right]$$
$$\geq \mathcal{Q}\left(\sqrt{\frac{L(d_p)\|\mathbf{d}_k\|^2}{2L(d_k)\sigma_s^2}}\mathrm{E}\left[\sqrt{\frac{\lambda(\mathbf{Q}_1)}{\lambda(\mathbf{Q}_2)}}\right]\right)$$
$$= \mathcal{Q}\left(\frac{2}{\pi^2}\sqrt{\frac{2L(d_p)\|\mathbf{d}_k\|^2}{L(d_k)\sigma_s^2}}b_1b_2\right) \tag{36}$$

$$\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_2 \to \hat{\mathbf{s}}_2|\mathbf{s}_1 \to \hat{\mathbf{s}}_1)\right] \approx \mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\left[P_{\mathrm{ML-ML}}(\mathbf{s}_1 \to \hat{\mathbf{s}}_1)\right], \tag{37}$$

where the Wishart matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are independent of each other. $\mathrm{E}[\frac{1}{\sqrt{\lambda(\mathbf{Q}_2)}}] = \frac{1}{\sqrt{N}\pi}\mathrm{B}(\frac{1}{4},\frac{3}{2})$, $b_1 = \mathrm{B}(\frac{1}{4},\frac{3}{2})$, and $b_2 = \mathrm{B}(\frac{3}{4},\frac{3}{2})$. Under high SNR, we can see that the PEP performance is independent of N. Instead, the PEP performance mainly depends on the value of distance gap between the selected two users.

V. Simulation Results

In this section, we present the simulation results to show the performance of our analysis with the quadrature phase shift keying (QPSK) used for signaling. We assume $KU = K + P$ and the transmitted signal power is set to be $\sigma_s^2 = 1$. Note that $\mathrm{SNR}' = \sigma_s^2 L(d_k)/N_0$ and $d_p > d_k$ (i.e., u_k is the near user). For convenience, the following simulations are considered in this section.

- "Analytical results, ML(or ZF), near user": simulations using the analytical PEP based on Eqn. (8) or (19) (Eqn. (8)

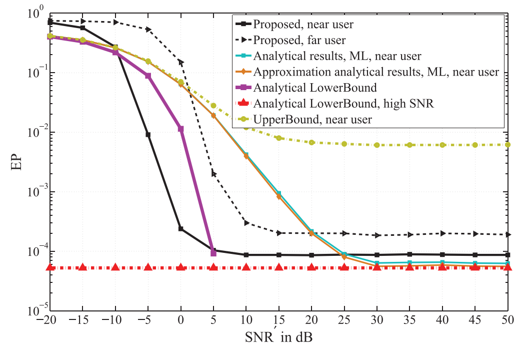


Fig. 2. PEP performance approximation of the ML-ML detector.

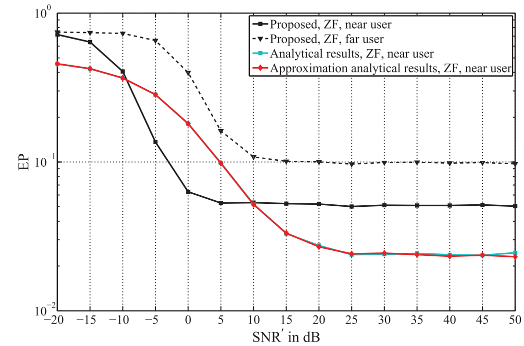


Fig. 3. PEP performance approximation of the ZF-ZF detector.

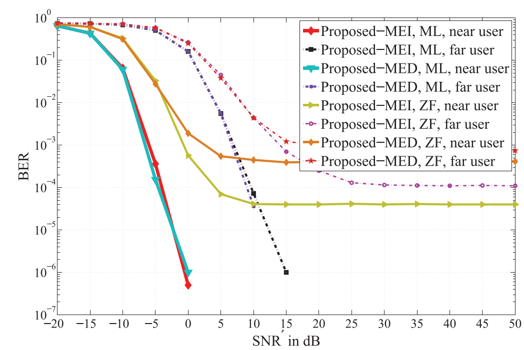


Fig. 4. BER versus SNR′ with $K = 5, P = 5$, and $M = N = 2$.

is for the ML-ML detector while Eqn. (19) is for the ZF-ZF detector).

- "Approximation analytical results, ML (or ZF), near user": simulations using the average PEP based on Eqn. (12) or (21) (Eqn. (12) is for the ML-ML detector while Eqn. (21) is for the ZF-ZF detector).
- "UpperBound, near user": simulations using the PEP upper bound of Eqn. (12), where $\|\mathbf{H}_k\mathbf{H}_p\|^2$ is replaced by $\lambda_{\max}(\mathbf{H}_k\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}})$.
- "Analytical LowerBound": simulations using the lower bound on the average PEP (which is based on Eqn. (34)).
- "Analytical LowerBound, high SNR": simulations using the lower bound on the average PEP under high SNR (which is based on Eqn. (36)).
- "Proposed-MEI (or MED), ML (or ZF), near user (or far user)": BER simulations for near user u_k (or far user u_p) based on the MEI (or MED) criterion (with ML-ML (or ZF-ZF) detector).
- "REF-$\cdots$": simulations with the user selection strategy proposed in [14], where the user group is selected by maximizing $|\mathbf{H_k}\mathbf{H_k^H}|/|\mathbf{H_p}\mathbf{H_p^H}|$.
- "Baseline, near user(or far user)": BER performance of the ZF-ZF detector for u_k (or u_p) with the random multiuser selection strategy in Section IV-B.
- "Sumrate-OMA": sumrate results of the MIMO-NOMA system with the orthogonal multiple access (OMA) strategy [40].

A. Performance Analysis of the Proposed System With Random u_k and u_p

In this subsection, we study the PEP performance of the uplink MIMO-NOMA system where two random users are selected to transmit signals to the BS. Because of the high computational complexity of the ML detector, we consider the 4×4 MIMO-NOMA system here. Besides, the theoretical upper bound and average PEP performance are also illustrated in the following results.

Fig. 2 shows the PEP performance of the ML-ML detector without user selection where the theoretical PEP lower bound for u_k is shown to be close to the PEP approximation in Eqns. (8) and (12). We can also find that the error floor exists due to the inter-user interference, which are predicted by Eqn. (17) and the approximated PEP of the ML detector with the MEI criterion. In addition, based on Eqn. (12), the maximization of the PEP for u_k begin to stabilize under high SNR. From Fig. 2, we can see that there exists an error floor even with the employment of the optimal detector.

In Fig. 3, the PEP performance of the ZF-ZF detector is shown with random selected u_k and u_p. From the simulation results, we can learn that the performance of the PEP is poor. The existence of the error floor is consistent with the analysis in Section III-A and the corresponding inter-user interference will be the main limiting factor in the signal detection under high SNR. Moreover, since the detected users are selected randomly, we can see that the poor PEP results are obtained when the linear ZF-ZF detector is employed. Therefore, reasonable multiuser selection criteria are necessary to suppress the interference and realize the signal estimation effectively.

B. BER Performance of the MIMO-NOMA System

In this subsection, based on different multiuser selection criteria, we mainly show the BER performance improvement of the NOMA-MIMO system with different signal detectors.

Fig. 4 shows the BER performance of the proposed detection system for different SNR′. Note that the results are obtained by averaging 10000 runs. The results shown in Fig. 4 illustrate that the BER performance with the proposed user selection criteria will increase when the SNR grows, and the inter-user interference has been suppressed effectively through the multiuser diversity. For the ZF-ZF detector, error floor could exist because of the strong multiuser interference. However, a good detection

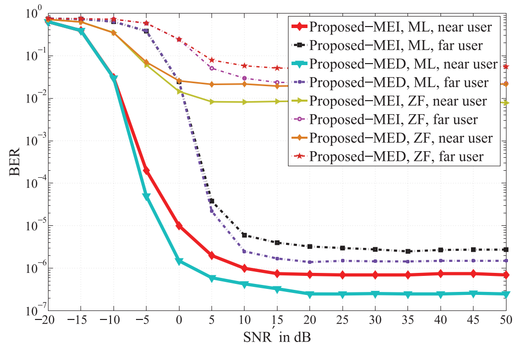


Fig. 5. BER performance of the 4×4 MIMO-NOMA system where $K = 5, P = 5$.

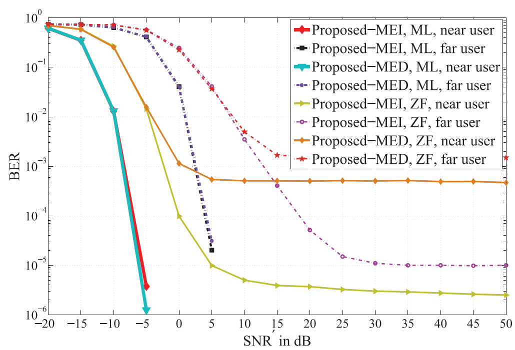


Fig. 6. BER performance of the 4×4 MIMO-NOMA system where $K = 35, P = 35$.

performance can be realized by employing the proposed multiple user selection criteria even under the circumstance where a small number of users exist, which means the strategy proposed in Proposition 1 can effectively improve the detection performance.

The BER performance of the 4×4 system (i.e., $N = M = 4$) is presented in Fig. 5 and Fig. 6. With fixed $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$, it is shown that the performance of the optimal detector (or linear detector) may not be improved as N increases. This phenomenon can be explained by Eqn. (28). When the number of the equipped antennas grows, the upper bound on the PEP performance will be declined since the ranges of the eigenvalues of the channel matrix ($\mathbf{W}_k$ for the ZF/MMSE detector) are enlarged as N increases. Moreover, given a number of users, the error floor can be effectively reduced by the multiuser diversity. From Fig. 6, based on the proposed MEI criterion, we can see that considerably low BER can be obtained with the linear detector.

In Fig. 7 and Fig. 8, as expected, we can see that the BER performance will be improved as the number of the total users increases. Besides, although the BER performance for u_p is worse than that for u_k, under high SNR′, the signal detection performance for both u_k and u_p is considerable. For u_p, when the SNR′ is low (e.g., less than 0 dB), the BER performance can hardly be improved even with large KU. This is because that signals $\mathbf{s}_p$ could be effectively estimated only when $\mathbf{s}_k$ are well detected. When the ML-ML detector is employed at the BS, we can also learn that the BER performance of the proposed methods is similar to that of the reference method in [14] under

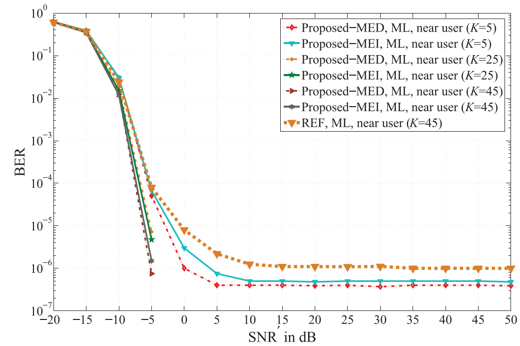


Fig. 7. BER versus SNR′ for u_k where $N = 4$ and the ML-ML detector is employed.

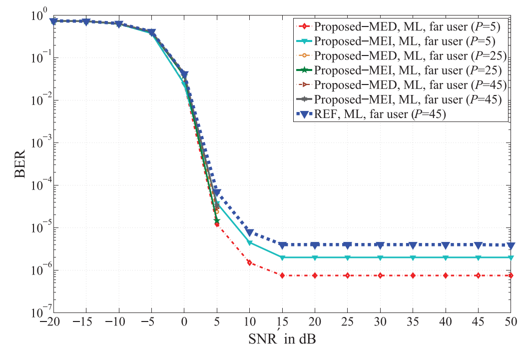


Fig. 8. BER versus SNR′ for u_p where $N = 4$ and the ML-ML detector is employed.

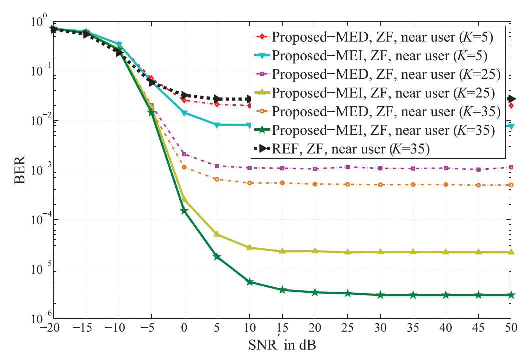


Fig. 9. BER versus SNR′ for u_k where $N = 4$ and the ZF-ZF detector is employed.

low SNR′, but superior to that of the reference method under high SNR′. From the simulation results, we can also find that the error floor of the system is eliminated when there are a large number of user candidates.

Fig. 9 and Fig. 10 show the BER performance of the ZF-ZF detector under the 4×4 MIMO-NOMA system where different number of users exists. In comparison with the 2×2 system (which is shown in Fig. 4), to achieve a considerable detection performance, much more user candidates are required to eliminate the multiuser interference for the 4×4 system. This is because the eigenvalue range of the larger matrix will increase as matrix dimensions grow, which means that the interference

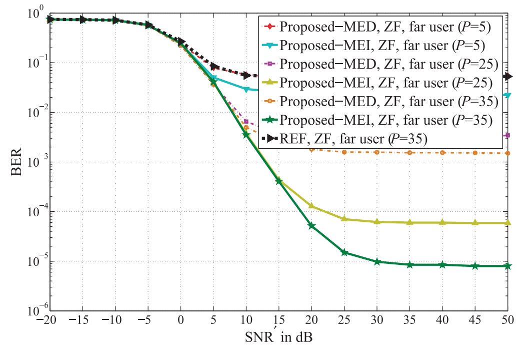


Fig. 10. BER versus SNR$'$ for u_p where $N = 4$ and the ZF-ZF detector is employed.

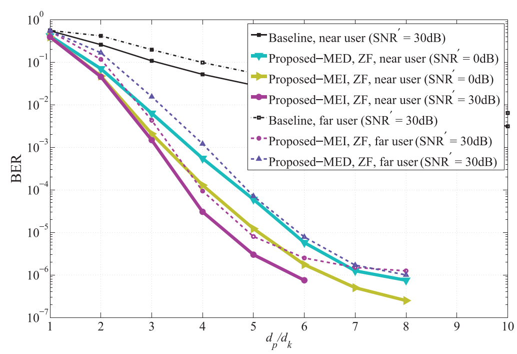


Fig. 11. BER versus $\frac{d_p}{d_k}$ where $N = 4$ and different multiuser criteria are employed.

will be enlarged. With the MEI criterion, the PEP performance for the u_p with $N = 4$ is slightly better than that with $N = 2$ when $P(\mathbf{s}_k \rightarrow \hat{\mathbf{s}}_k | N = 2) \approx P(\mathbf{s}_k \rightarrow \hat{\mathbf{s}}_k | N = 4)$. This phenomenon can be explained by the performance analysis of signal detection in the point-to-point MIMO system, where the upper bound on the PEP is proportional to the received antennas [8]. Moreover, based on the user pairing method in [14], we can see that the transmitted signals cannot be successfully estimated since the receiver constraints are not taken into consideration in the reference method.

To observe the impact of the channel gap between $\mathbf{H}_k$ and $\mathbf{H}_p$ on the system performance, simulations are performed to obtain the BER results for various $\frac{d_p}{d_k}$. Here, two channel matrices are generated randomly. For other selection criteria, we set $K = P = 5$. In Fig. 11, under a high SNR$'$ (e.g., 30 dB), we can see the BER performance of the ZF-ZF detector becomes satisfactory when $\frac{d_p}{d_k} = 5$. When the distance gap increases, the inter-user interference becomes lower. Although the signal attenuation is serious for the weak user, the detection quality could be guaranteed under high SNR.

Fig. 12 shows the system sum-rate results [41] based on different user selection criteria. In comparison with the MIMO-OMA system, higher spectrum efficiency can be realized by the multiuser MIMO-NOMA system with the proposed methods. When the ML-ML detector is adopted, we can see that the sum-rate performance of our proposed MEI (or MED) selection criterion is close to that of the reference method in [14]. However, as

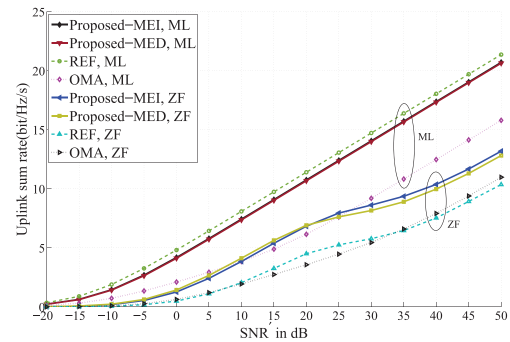


Fig. 12. Sum rate versus SNR$'$ where $N = 4$ and $\frac{d_p}{d_k} = 8$.

shown in Fig. 9 and Fig. 10, the system detection performance of the comparing user-pairing method cannot be guaranteed, which means that the proposed methods are applicable to the given MIMO-NOMA system. Moreover, for the ZF-ZF suboptimal detector, it shows that the performance of the proposed MEI/MED criterion is much better than that of method in [14]. Therefore, our proposed methods are suitable to the MIMO-NOMA system even with suboptimal detectors.

VI. Conclusion

In this paper, we have investigated the signal detection of the MIMO-NOMA system and proposed an optimal multiuser selection criterion for the proposed system. Since two users can access the BS simultaneously, the inter-user interference could be the main factor limiting the total system performance under high SNR. Thus, sub-optimal and considerable multiuser criteria are designed to provide a solution for the MIMO-NOMA system. Given a limited number of users, we find that the strategy to improve the detection performance of the stronger user is viable which determines the upper-limit of the PEP performance of the weaker user. By exploiting the multiuser diversity, in the simulation results, we have shown that the proposed approaches are feasible and perform well under the multiuser circumstance. Moreover, considering the case with a small number of users, we have derived the approximated PEP bound of different signal detectors and derived a closed PEP estimation when the users are selected randomly.

In the future work, we will consider studying the user selection methods for massive MIMO systems [42], [43] where precoding could be adopted at the user side. The work of this paper will be extended to design the precoders for the selected users by optimizing the detection performance at the BS. Besides, for the future networks, the signal detection against interference background could be one of the key technologies [44], [45]. We consider combining our scheme with other 5G technologies, such as device-to-device communication [42] and satellite communications [46]–[47]. These might be our future research directions.

Appendix A
Proof for Proposition 1

From Eqns. (9) and (13), we can directly derive that the PEP_p is no larger than the PEP_k. When the detection performance for

$\mathbf{s}_k$ is not good enough, the signals transmitted from u_p cannot be correctly estimated. Thus, to guarantee the detection performance for $\mathbf{s}_p$, the PEP_k needs to be small enough. According to Eqns. (9) and (12), we can find that the inter-user interference is the main factor influencing the detection performance. Since

$$P(\mathbf{s}_p \to \hat{\mathbf{s}}_p \mid \mathbf{s}_k \to \hat{\mathbf{s}}_k) \leq \mathcal{Q}\left(\frac{\|\mathbf{H}_p\mathbf{d}_p\|^2 - 2\,|\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{d}_k|}{\sqrt{2N_0\|\mathbf{H}_p\mathbf{d}_p\|^2}}\right),$$

and the interference term $2\,|\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{d}_k|$ might be larger than $\|\mathbf{H}_p\mathbf{d}_p\|^2$, the primary task is to suppress the inter-user interference.

When the SNR is high (i.e., $N_0 \to 0$), and assuming that certain $\mathbf{H}_k$ and $\mathbf{H}_p$ are selected based on Eqn. (14), for a given $\mathbf{d}_k$, we have

$$P(\mathbf{s}_k \to \hat{\mathbf{s}}_k) = \mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{\sqrt{\sigma_s^2\|\mathbf{d}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\|^2)}}\right), \tag{38}$$

which can be minimized with the exhaustive search method. The denominator of Eqn. (38) can be viewed as the inter-interference term. Since $\mathcal{Q}(x)$ is the descending function of x, from Eqn. (8), we can know that the minimization of the interference term can help improve the detection performance for $\mathbf{s}_k$. On the other hand, to maximize the numerator $\|\mathbf{H}_k\mathbf{d}_k\|^2$ is exactly related to the improvement of the PEP performance of $\mathbf{s}_k$ in single-user communication systems [10]. Suppose that $|\mathcal{G}_1|, |\mathcal{G}_2| \ll \infty$, based on the above analysis, the problem in Eqn. (13) can be simplified as

$$\{k^\star, p^\star\} = \arg\min_{k,p}\max_{\mathbf{d}_k \in \mathcal{D}}(PEP_k),$$

where $PEP_k = P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)$. Therefore, we finish the proof.

Appendix B
Proof for Remark 1

Given user u_k, based on the MED criterion, for both the ML-ML detector and the ZF-ZF detector, we have

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_p}[P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)] \\
&\leq \mathrm{E}\left[\mathcal{Q}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|}{\sqrt{2N_0 + 2\min\lambda_{\max}(\sigma_s^2\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})}}\right)\right] \\
&= \int_0^\infty \int_{\frac{\|\mathbf{H}_k\mathbf{d}_k\|}{\sqrt{2N_0 + 2\frac{\sigma_s^2 u}{L(d_p)}}}}^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}} f_U(u)\mathrm{d}x\mathrm{d}u \\
&= \int_0^{\sqrt{\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{2N_0}}} \int_{\frac{L(d_p)}{\sigma_s^2}\left(\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{2x^2} - N_0\right)}^\infty f_U(u)\mathrm{d}u \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}\mathrm{d}x \\
&\quad + \int_{\sqrt{\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{2N_0}}}^\infty \int_0^\infty f_U(u)\mathrm{d}u \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}\mathrm{d}x \\
&= [1 - F_Z(0)]^P \int_{\eta_1}^\alpha \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\mathrm{d}x + [1 - F_Z(0)]^P \mathcal{Q}(\alpha),
\end{aligned} \tag{39}$$

where $\alpha = \sqrt{\frac{\|\mathbf{H}_k\mathbf{d}_k\|^2}{2N_0}}$ and $\eta_1 \in [0, \alpha]$.

Appendix C
Proof for Theorem 1

For $\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{s}_k$, according to [39], we have

$$-\|\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{s}_k\| \leq \mathrm{Re}\{\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{s}_k\} \leq \|\mathbf{d}_p^{\mathrm{H}}\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k\mathbf{s}_k\|. \tag{40}$$

Then, for the ZF-ZF detector, we can also derive that

$$(19) \leq \mathcal{Q}\left(\frac{\sqrt{\lambda_{\min,\mathbf{H}_k}}\|\mathbf{d}_k\| - 2\sqrt{\lambda_{\max,\mathbf{H}_p}}\,\|\mathbf{s}_p\|}{\sqrt{2N_0}}\right),$$

where $\lambda_{\min,\mathbf{H}_k} = \lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})$ and $\lambda_{\min,\mathbf{H}_p} = \lambda_{\min}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})$. Then, based on the MED criterion and integral theorem [38], we can get

$$\begin{aligned}
PEP_{\mathbf{s}_k} &\leq \int_0^\infty \int_0^\infty \mathcal{Q}\left(\sqrt{\frac{\gamma}{2}}(m_k\|\mathbf{d}_k\|\sqrt{v} - 2m_p\|\mathbf{s}_p\|\sqrt{u})\right) \\
&\qquad \times f(v)f(u)\mathrm{d}v\mathrm{d}u &\text{(41a)}\\
&= \int_0^\infty \int_0^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}} g(v,K)f_U(u)\mathrm{d}x\mathrm{d}v\mathrm{d}u &\text{(41b)}\\
&\quad - \int_{\mathrm{B}}^\infty \int_0^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}} g(v,K)f_U(u)\mathrm{d}x\mathrm{d}v\mathrm{d}u &\text{(41c)}\\
&\quad + \int_{\mathrm{B}}^\infty \int_0^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}} f_V(v)f_U(u)\mathrm{d}x\mathrm{d}v\mathrm{d}u, &\text{(41d)}
\end{aligned}$$

where $m_k = \sqrt{L(d_k)}, m_p = \sqrt{L(d_p)}, g(v,K) = a_0v^{K-1} + o(v^{K-1+\tau}), r = \sqrt{N_0/2}[\|\mathbf{d}_k\|\sqrt{v/L(d_k)} - 2\,\|\mathbf{s}_p\|\sqrt{u/L(d_p)}]$, and $\gamma = 1/N_0$. Here, $V = \max\{X_1, \ldots, X_K\}$ and $U = \min\{Z_{p=K+1}, \ldots, Z_{p=K+P}\}$. Note that B is small enough that the PDF of V can be approximated as $f_V(v) = a_0v^t + o(v^{t+\epsilon})$, where a_0 is a constant, $\epsilon > 0$, and t depends on V. For $Z = L(d_p)\lambda_{\max}(\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}})$, when N is large, the ranges of the value are narrow. Thus, we have

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}_k}[P(\mathbf{s}_k \to \hat{\mathbf{s}}_k)] \\
&\leq \int_0^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}[a_0v^{K-1} + o(v^{K-1+\tau})]\mathrm{d}x\mathrm{d}v \\
&\quad - \int_{\mathrm{B}}^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}[a_0v^{K-1} + o(v^{K-1+\tau})]\mathrm{d}x\mathrm{d}v \\
&\quad + \int_{\mathrm{B}}^\infty \int_r^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}} f_V(v)\mathrm{d}x\mathrm{d}v.
\end{aligned} \tag{42}$$

After a brief derivation [20], we can show that Eqn. (41b) is upper bounded by

$$\begin{aligned}
&\frac{a_0}{2K}\sqrt{\frac{\gamma}{\pi}}\frac{1}{(\|\mathbf{d}_k\|/\sqrt{L(d_k)})^{2K}}\left(\frac{\gamma}{2}\right)^{\frac{-(2K+1)}{2}}\Gamma(2K+1) \\
&\quad \times e^{-\frac{\|\mathbf{s}_p\|^2u\gamma}{2L(d_p)}} D_{-(2K+1)}(-l_0) \\
&\quad + o\left(\left(\frac{\gamma}{2}\right)^{\frac{-(2K+1)}{2}}\Gamma(2K+1)e^{-\frac{\|\mathbf{s}_p\|^2u\gamma}{2L(d_p)}} D_{-(2K+1)}(l_0)\right).
\end{aligned} \tag{43}$$

$l_0 = -\|\mathbf{s}_p\|\sqrt{2u\gamma/L(d_p)}$ and $D_p(x) = \frac{e^{\frac{-x^2}{4}}}{\Gamma(-p)}\int_0^\infty e^{-xz-\frac{z^2}{2}} z^{-p-1}\mathrm{d}z$. We can learn that the upper bound of Eqn. (41c) is $o\,(\text{Eqn.}(43))$ while Eqn. (41d) is upper bounded by $\mathcal{Q}(\sqrt{\gamma/2}\,[\|\mathbf{d}_k\|\sqrt{\mathrm{B}/L(d_k)}-2\,\|\mathbf{s}_p\|\sqrt{u/L(d_p)}])$.

Thus, we can learn that Eqn. (41d) is $o(\gamma^{-(K+1)})$ when $u=0$. Then, based on Eqns. (42) and (43), we can get

$$\begin{aligned}&\mathrm{E}_{\mathbf{H}_k}[P(\mathbf{s}_k\to\hat{\mathbf{s}}_k)]\\&\leq\frac{a_0}{2\,K}\sqrt{\frac{\gamma}{\pi}}\frac{1}{[\|\mathbf{d}_k\|^2/L(d_k)]^K}\left(\frac{\gamma}{2}\right)^{\frac{-(2K+1)}{2}}\Gamma(2K+1)\\&\quad\times e^{-\frac{\|\mathbf{s}_p\|^2u\gamma}{2L(d_p)}}D_{-(2K+1)}\left(-\|\mathbf{s}_p\|\sqrt{2u\gamma/L(d_p)}\right)\\&\quad+\mathcal{Q}\left(\sqrt{\gamma/2}\left[\|\mathbf{d}_k\|\sqrt{\mathrm{B}/L(d_k)}-2\,\|\mathbf{s}_p\|\sqrt{u/L(d_p)}\right]\right).\end{aligned}\tag{44}$$

For the second term in Eqn. (44), we have

$$\begin{aligned}&\mathrm{E}_U\left[\mathcal{Q}\left(\sqrt{\gamma/2}\left[\|\mathbf{d}_k\|\sqrt{\mathrm{B}/L(d_k)}-2\,\|\mathbf{s}_p\|\sqrt{u/L(d_p)}\right]\right)\right]\\&\leq\int_0^\infty\mathcal{Q}\left(\sqrt{\frac{\gamma}{2}}\,\|\mathbf{d}_k\|\sqrt{\frac{B}{L(d_k)}}-2\,\|\mathbf{s}_p\|\sqrt{\frac{u}{L(d_p)}}\right)f_U(u)\mathrm{d}u\\&=\int_0^\infty\int_{\sqrt{\frac{\gamma}{2}}\,\|\mathbf{d}_k\|\sqrt{\frac{B}{L(d_k)}}-2\,\|\mathbf{s}_p\|\sqrt{\frac{u}{L(d_p)}}}^\infty\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}f_U(u)\mathrm{d}x\mathrm{d}u\\&=\int_{c_1-2\,c_2\sqrt{N}}^{c_1}\int_{(\frac{x-c_1}{c_2})^2}^{4\,N}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}[1-(1-\mathrm{F}_Z(u))^K]'\mathrm{d}u\mathrm{d}x\\&\quad+\int_{c_1}^\infty\int_0^{4\,N}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}[1-(1-\mathrm{F}_Z(u))^K]'\mathrm{d}u\mathrm{d}x\\&=\int_{-2\,c_2\sqrt{N}}^0\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-(x+c_1)^2}{2}}[1-\mathrm{F}_Z(x^2/c_2^2)]^K\mathrm{d}x\\&\quad+\int_{c_1}^\infty\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-(x+c_1)^2}{2}}[1-\mathrm{F}_Z(0)]^K\mathrm{d}x.\end{aligned}\tag{45}$$

In Eqn. (45), $\mathrm{F}_Z(z)=e^{\frac{-|4N-z|^3}{192\,N}}, c_1=\sqrt{\frac{B\gamma}{2L(d_k)}}\,\|\mathbf{d}_k\|$, and $c_2=\sqrt{2\gamma/L(d_p)}\,\|\mathbf{s}_p\|$. Based on the mean value theorem for integrals, we can see that

$$\begin{aligned}&\mathrm{E}_U\left[\mathcal{Q}\left(\sqrt{\gamma/2}\left[\|\mathbf{d}_k\|\sqrt{\mathrm{B}/L(d_k)}-2\,\|\mathbf{s}_p\|\sqrt{u/L(d_p)}\right]\right)\right]\\&\leq(1-\mathrm{F}_Z(0))^K\left(\int_{c_1+\varepsilon_1}^{c_1}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-(x)^2}{2}}\mathrm{d}x+\mathcal{Q}(2c_1)\right)\\&=\left(1-e^{-\frac{N^2}{3}}\right)^K[\mathcal{Q}(c_1)-\mathcal{Q}(c_1+\varepsilon_1)+\mathcal{Q}(2c_1)],\end{aligned}\tag{46}$$

where $\varepsilon_1\in[-2\,c_2\sqrt{N},\,0]$. From Eqn. (46), given the value of c_1, we can learn that the PEP performance can increase with the K. Since $u\approx bN, b\geq0$, after substituting Eqn. (46) in Eqn. (44), it can be derived that

$$\begin{aligned}&\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}[P(\mathbf{s}_k\to\hat{\mathbf{s}}_k)]\leq\frac{a_0}{2\,K}\sqrt{\frac{\gamma}{\pi}}\frac{[L(d_k)]^K}{\|\mathbf{d}_k\|^2}\left(\frac{\gamma}{2}\right)^{\frac{-(2K+1)}{2}}\\&\quad\times\Gamma(2K+1)e^{\frac{-\|\mathbf{s}_p\|^2bN\gamma}{2L(d_p)}}D_{-(2K+1)}\left(-\|\mathbf{s}_p\|\sqrt{2bN\gamma/L(d_p)}\right)\\&\quad+\left(1-e^{-\frac{N^2}{3}}\right)^K[\mathcal{Q}(c_1)-\mathcal{Q}(c_1+\varepsilon_1)+\mathcal{Q}(2c_1)].\end{aligned}\tag{47}$$

Then, based on the similar method, we have

$$\mathrm{E}[P(\mathbf{s}_p\to\hat{\mathbf{s}}_p\mid\mathbf{s}_k\to\hat{\mathbf{s}}_k)]\leq c_3(\gamma)^{-1}+o(\gamma^{-1}),\tag{48}$$

where $c_3=\frac{L(d_p)\Gamma(\frac{3}{2})}{\sqrt{\pi}\|\mathbf{d}_p\|^2}$ and $\mathbf{s}_k$ are estimated correctly (i.e., $\hat{\mathbf{s}}_k=\mathbf{s}_k$) for simplicity. Since $D_p(z)\sim e^{-\frac{z^2}{4}}z^p(1-\frac{p(p-1)}{2Z^4}+\frac{p(p-1)(p-2)(p-3)}{2\times4z^4}-\cdots)$ when $|z|\gg1$, under high SNR (i.e., $|z|\gg K$), Eqn. (47) becomes

$$\begin{aligned}&\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}[P(\mathbf{s}_k\to\hat{\mathbf{s}}_k)]\leq\frac{a_0}{2\,K}\sqrt{\frac{\gamma}{\pi}}\frac{[L(d_k)]^K}{\|\mathbf{d}_k\|^2}(\gamma)^{-(2K+1)}\\&\quad\times\Gamma(2K+1)e^{-\frac{\|\mathbf{s}_p\|^2bN\gamma}{L(d_p)}}\left|-\|\mathbf{s}_p\|\sqrt{bN/L(d_p)}\right|^{-(2K+1)}\\&\quad+\left(1-e^{-\frac{N^2}{3}}\right)^K[\mathcal{Q}(c_1)-\mathcal{Q}(c_1+\varepsilon_1)+\mathcal{Q}(2c_1)].\end{aligned}\tag{49}$$

In Eqn. (49), the second term can be regarded as the interference. From Eqn. (49), we know that the BER performance increases as the SNR increases and the equivalent diversity order is larger than the number of users in $\mathcal{G}_1$ (i.e., $|\mathcal{G}_1|$). Furthermore, from Eqn. (48), we know that $\mathcal{D}_p$ is limited. Fortunately, since PEP_p is dependent on PEP_k, a considerable improvement on the detection performance can be obtained when the messages transmitted form $\mathbf{u}_k$ are detected successfully.

APPENDIX D
PROOF FOR (27)

$$\begin{aligned}&\mathrm{E}_{\mathbf{H}_p}[P_{\mathrm{ML-ML}}(\mathbf{s}_k\to\hat{\mathbf{s}}_k)]\\&\leq\mathrm{E}_{\mathbf{H}_p}\left[\mathcal{Q}\left(\frac{\|\mathbf{d}_k\|^2\lambda_k}{\sqrt{2N_0\lambda_k+2\min\lambda_{\max}(\sigma_s^2\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_p\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_k)}}\right)\right]\\&=\mathrm{E}_{\mathbf{M}_p}\left[\mathcal{Q}\left(\frac{\|\mathbf{d}_k\|\lambda_k}{\sqrt{2N_0\lambda_k+2\frac{N\sigma_s^2}{L(d_k)L(d_p)}\min\lambda_{\max}(\mathbf{M}_p^{\mathrm{H}}\mathbf{M}_p)}}\right)\right]\\&\overset{(a)}{=}\int_0^\infty\int_{\frac{\|\mathbf{d}_k\|\lambda_k}{\sqrt{2N_0\lambda_k+2uN\sigma_s^2/[L(d_k)L(d_p)]}}}^\infty\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}f_U(u)\mathrm{d}x\mathrm{d}u\\&=\int_0^{\sqrt{\frac{\lambda_k\|\mathbf{d}_k\|^2}{2N_0}}}\int_{\frac{L(d_k)L(d_p)}{N\sigma_s^2}(\frac{\|\mathbf{d}_k\|^2(\lambda_k)^2}{2x^2}-N_0\lambda_k)}^\infty f_U(u)\mathrm{d}u\frac{1}{\sqrt{2\pi}}\\&\quad\times\mathrm{e}^{\frac{-x^2}{2}}\mathrm{d}x+\int_{\sqrt{\frac{\lambda_k\|\mathbf{d}_k\|^2}{2N_0}}}^\infty\int_0^\infty f_U(u)\mathrm{d}u\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}\mathrm{d}x\\&\overset{(b)}{=}[1-F_Z(0)]^D[\mathcal{Q}(\eta_4)],\end{aligned}\tag{50}$$

where $\lambda_{\min}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k) \le \lambda_k \le \lambda_{\max}(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k), \max\{K,P\} < D \le PK$, and $\eta_4 \in [0, \sqrt{\frac{\lambda_k\|\mathbf{d}_k\|^2}{2N_0}}]$. In Equality $(a), U = \min \lambda_{\max}(\mathbf{M}_p^{\mathrm{H}}\mathbf{M}_p)$ and its PDF is $f_U(u)$. Equality (b) is attributed to the mean value theorem of integrals [38].

Appendix E
Proof for (28)

$$
\begin{aligned}
&\mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}[P_{\mathrm{ML-ML}}(\mathbf{s}_k \to \hat{\mathbf{s}}_k)] \\
&\le \mathrm{E}_{\mathbf{H}_k,\mathbf{H}_p}\mathcal{Q}\left(\frac{\|\mathbf{d}_k\|\sqrt{\lambda_{\min}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}})}}{\sqrt{2\lambda_{\max}(N_0\mathbf{I}+\sigma_s^2\mathbf{H}_p^{\mathrm{H}}\mathbf{H}_p)}}\right) \\
&= \mathrm{E}_{U,V}\left[\mathcal{Q}\left(\frac{\|\mathbf{d}_k\|\sqrt{\frac{v}{L(d_k)}}}{\sqrt{2N_0+2(\frac{\sigma_s^2}{L(d_p)}u)}}\right)\right] \\
&\overset{(c)}{=} \int_0^\infty\int_0^\infty\int_{g(u,v)}^\infty \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}}\mathrm{d}x f_U(u) f_V(v)\mathrm{d}v\mathrm{d}u \\
&\overset{(d)}{=} \int_0^{4N}\left[\frac{2^{K+\frac{1}{2}}KN^K\Gamma(K+\frac{1}{2})}{\sqrt{2\pi}K}[\bar{\gamma}(u)]^{-K} + o([\bar{\gamma}(u)]^{-K})\right] \\
&\quad\times f_U(u)\mathrm{d}u \\
&\approx \frac{2^{K+\frac{1}{2}}N^K\Gamma(K+\frac{1}{2})L(d_p)}{\sqrt{2\pi}\sigma_s^2}\left[\frac{\|\mathbf{d}_k\|^2}{2L(d_k)}\right]^{-K}[1-F_Z(\eta_3)]^P \\
&\quad\times(N_0+4\frac{\sigma_s^2}{L(d_p)}N)^K, \qquad (51)
\end{aligned}
$$

where $g(u,v) = \frac{\|\mathbf{d}_k\|\sqrt{L(d_p)v}}{\sqrt{2N_0L(d_k)L(d_p)+2(\sigma_s^2L(d_k)u)}}, \bar{\gamma}(u) = \frac{\|\mathbf{d}_k\|^2}{2L(d_k)(N_0+\frac{\sigma_s^2u}{L(d_p)})}$, and $\eta_3 \in [0, 4N]$. In Equality $(c), V = \max\lambda_{\min}(\mathbf{M}_k^{\mathrm{H}}\mathbf{M}_k)$ and its PDF is $f_V(v)$. Equality (d) is attributed to the result in [11].

References

[1] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 112–125, Jan. 2014.

[2] M. Matthaiou, C. Zhong, and T. Ratnarajah, "Novel generic bounds on the sum rate of MIMO ZF receivers," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4341–4353, Sep. 2011.

[3] D. A. Basnayaka, P. J. Smith, and P. A. Martin, "Performance analysis of macrodiversity MIMO systems with MMSE and ZF receivers in flat rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2240–2251, May 2013.

[4] J. Choi and J. Ha, "Rate optimization to minimize distortion for source-channel coded H-BLAST with SIC decoding," *IEEE Commun. Lett.*, vol. 13, no. 2, pp. 115–117, Feb. 2009.

[5] P. Liu and I. Kim, "Exact and closed-form error performance analysis for hard MMSE-SIC detection in MIMO Systems," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2463–2477, Sep. 2011.

[6] X. Ma and W. Zhang, "Performance analysis for MIMO systems with lattice-reduction aided linear equalization," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 309–318, Feb. 2008.

[7] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. 35th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2001, pp. 1194–1199.

[8] J. Choi and F. Adachi, "User selection criteria for multiuser systems with optimal and suboptimal LR based detectors," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5463–5468, Oct. 2010.

[9] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antenna," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[10] L. Bai and J. Choi, *Low Complexity MIMO Detection*. Berlin, Germany: Springer, 2012.

[11] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[12] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink nonorthogonal multiple access with statistical channel state information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4718–4731, Jul. 2016.

[13] L. Zhang, J. Liu, M. Xiao, G. Wu, Y.-C. Liang, and S. Li, "Performance analysis and optimization in downlink NOMA systems with cooperative full-duplex relaying," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2398–2412, Oct. 2017.

[14] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[15] L. Bai, L. Zhu, T. Li, J. Choi, and W. Zhuang, "An efficient hybrid transmission method: Using non-orthogonal multiple access and multiuser diversity," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2276–2288, Mar. 2018.

[16] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.

[17] X. Chen, Z. Zhang, C. Zhong, and D. W. K. Ng, "Exploiting multiple-antenna techniques for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2207–2220, Oct. 2017.

[18] 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.

[19] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.

[20] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[21] H. Lee, D. Kim, and D. Hong, "Mode selection in multi-user full-duplex systems considering inter-user interference," in *Proc. 24th Eur. Signal Process. Conf.*, Aug. 2016, pp. 769–772.

[22] W. Sun, Y. Ge, Z. Zhang, and W. Wong, "An analysis framework for interuser interference in IEEE 802.15.6 body sensor networks: A stochastic geometry approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8567–8577, Oct. 2016.

[23] S. Huang, H. Yin, J. Wu, and V. C. M. Leung, "User selection for multiuser MIMO downlink with zero-forcing beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3084–3097, Sep. 2013.

[24] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[25] L. Bai, C. Chen, J. Choi, and L. Cong, "Greedy user selection using a lattice reduction updating method for multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 136–147, Jan. 2011.

[26] J. Liu, H. Nishiyama, N. Kato, and J. Guo, "On the outage probability of device-to-device-communication-enabled multichannel cellular networks: An RSS-threshold-based perspective," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 163–175, Jan. 2016.

[27] Y. S. Cho, D. Kim, T. K. Kim, and G. H. Im, "Improved channel-dependent user selection for an SC-FDMA system," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1608–1611, Oct. 2012.

[28] W. Peng, W. Gao, and J. Liu, "AI enabled massive devices multiple access for smart city," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7623–7634, Oct. 2019, doi: 10.1109/JIOT.2019.2902448.

[29] R. Zeng and C. Tepedelenliolu, "Underlay cognitive multiuser diversity with random number of secondary users," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5571–5581, Oct. 2014.

[30] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[32] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1794–1803, Jul. 2002.

[33] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[34] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Boston, MA, USA: Now, 2004.

[35] Z. Bai, Y. Chen, and Y. C. Liang, *Random Matrix Theory and Its Application*. Singapore: World Scientific, 2009.

[36] A. Edelman, "Eigenvalues and condition numbers of random matrices," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, Oct. 1988.

[37] P. Vivo, S. N. Majumdar, and O. Bohigas, "Large deviations of the maximum eigenvalue in Wishart random matrices," *J. Phys. A-Math. Theor.*, vol. 40, no. 16, pp. 4317–4337, Mar. 2007.

[38] S. Ghahramani, *Fundamentals of Probability*. Englewood Cliffs, NJ, USA.: Prentice-Hall, 2000.

[39] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. New York, NY, USA: Academic, 1994.

[40] K. Janghel and S. Prakriya, "Performance of adaptive OMA/cooperative-NOMA scheme with user selection," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2092–2095, Oct. 2018.

[41] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[42] W. Peng, W. Gao, and J. Liu, "A novel perspective on multiple access in 5G network: Framework and solutions," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 154–160, Jun. 2019, doi: 10.1109/MWC.2019.1800315.

[43] J. Chen, H. Chen, and F. Zhao, "Energy-efficient joint user association and power allocation in relay-aided massive MIMO systems," *J. Commun. Inf. Netw.*, vol. 3, no. 2, pp. 58–65, Jun. 2018.

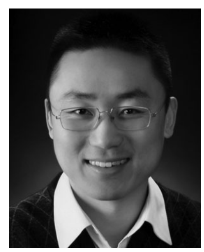
[44] H. Guo, J. Liu, and J. Zhang, "Computation offloading for multi-access mobile edge computing in ultra-dense networks," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 14–19, Aug. 2018.

[45] X. Liu, R. Li, K. Luo, and T. Jiang, "Downlink and uplink decoupling in heterogeneous networks for 5G and beyond," *J. Commun. Inf. Netw.*, vol. 3, no. 2, pp. 1–13, Jun. 2018.

[46] J. Liu, Y. Shi, L. Zhao, Y. Cao, W. Sun, and N. Kato, "Joint placement of controllers and gateways in SDN-enabled 5G-satellite integrated network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 221–232, Feb. 2018, doi: 10.1109/JSAC.2018.2804019.

[47] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surv. Tut.*, vol. 20, no. 4, pp. 2714–2741, Oct.–Dec. 2018, doi: 10.1109/COMST.2018.2841996.

Chao Han (S'18) received the B.Sc. degree in biological medical engineering from Changzhi Medical College, Shanxi, China, in 2013 and the M.Sc. degree in electronic science and technology from Beijing Jiaotong University, Beijing, China, in 2016. He is currently working toward the Ph.D. degree in communication and information systems with the School of Electronic and Information Engineering, Beihang University, Beijing, China. His research interests include signal processing, transceiver design of wireless communications, and hybrid beamforming.



Lin Bai (M'13–SM'17) received the B.Sc. degree in electronic and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, the M.Sc. (Hons.) degree in communication systems from the University of Wales, Swansea, U.K., in 2007, and the Ph.D. degree in advanced telecommunications from the School of Engineering, Swansea University, U.K., in 2010. Since 2011, he has been with Beihang University (Beijing University of Aeronautics and Astronautics, BUAA), Beijing, China, as an Associate Professor. He is currently with the School of Cyber Science and Technology. His research interests include multiple-input multiple-output (MIMO), Internet of Things (IoT), and unmanned aerial vehicle (UAV) communications. Dr. Bai received the IEEE Communications Letters Exemplary Reviewers Certificate for 2012 and the IEEE Wireless Communications Letters Top Editor Award for 2018. He is currently an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, and a Managing Editor of *Journal of Communications and Information Networks* (JCIN). He has also served as a Lead Guest Editor for the IEEE WIRELESS COMMUNICATIONS and a Guest Editor for the IEEE INTERNET OF THINGS JOURNAL.

Quan Yu (SM'15) received the B.S. degree in information physics from Nanjing University, Nanjing, China, in 1986, the M.S. degree in radio wave propagation from Xidian University, Xian, China, in 1988, and the Ph.D. degree in fiber optics from the University of Limoges, Limoges, France, in 1992. In 1992, he joined the Institute of China Electronic System Engineering Corporation, Beijing, China, where he has been a Senior Engineer, and is currently a Research Fellow. He is an Academician with the Chinese Academy with the Engineering (CAE). His research interests include architecture of wireless networks, optimization of protocols, and cognitive radios.

Jingchao Wang received the Ph.D. degree in electronics engineering from Tsinghua University, Beijing, China. His current research interests include space information networks and satellite communications.

Jinho Choi (SM'02) was born in Seoul, South Korea. He received B.E. (*magna cum laude*) degree in electronics engineering in 1989 from Sogang University, Seoul, South Korea, and the M.S.E. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1991 and 1994, respectively. He is currently with the School of Information Technology, Deakin University, Geelong VIC, Australia, as a Professor. Prior to joining Deakin in 2018, he was with Swansea University, Swansea, U.K., as a Professor/Chair in wireless, and Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, as a Professor. His research interests include the Internet of Things (IoT), wireless communications, and statistical signal processing. He has authored two books published by Cambridge University Press in 2006 and 2010. Dr. Choi was the recipient of the 1999 Best Paper Award for Signal Processing from EURASIP and 2009 Best Paper Award from WPMC (Conference). He is currently an Editor for the IEEE TRANSACTIONS COMMUNICATIONS and IEEE WIRELESS COMMUNICATIONS LETTERS and had served as an Associate Editor or Editor of other journals including IEEE COMMUNICATIONS LETTERS, *Journal of Communications and Networks* (JCN), IEEE TRANSACTIONS VEHICULAR TECHNOLOGY, and *ETRI* journal.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include resource management, wireless network security, social networks, smart grids, and vehicular ad hoc and sensor networks. Dr. Shen was the recipient of the R. A. Fessenden Award from the IEEE, Canada, in 2019, the James Evans Avant Garde Award from the IEEE Vehicular Technology Society in 2018, and the Joseph LoCicero Award and the Education Award from the IEEE Communications Society in 2015 and 2017, respectively. He was also the recipient of the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award five times from the University of Waterloo and the Premiers Research Excellence Award from the Province of Ontario, Canada, in 2003. He served as the Technical Program Committee Chair/Co-Chair for the 2016 IEEE Global Communications Conference, the 2014 IEEE International Conference on Computer Communications, the 2010 IEEE Vehicular Technology Conference Fall, and the 2007 IEEE Global Communications Conference, the Symposia Chair for the 2010 IEEE International Conference on Communications, the Tutorial Chair for the 2011 IEEE Vehicular Technology Conference Spring, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He is the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL and the Vice President on Publications of the IEEE Communications Society. He is a registered Professional Engineer of Ontario, Canada, a fellow of the Engineering Institute of Canada, the Canadian Academy of Engineering, and the Royal Society of Canada, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and the IEEE Communications Society.