Large System Analysis of Downlink MISO-NOMA System via Regularized Zero-Forcing Precoding with Imperfect CSIT

Author 1, Author 2, Author 3, Author 4, and Author 5

***Abstract*—This letter studies the multiple-input single-output (MISO) non-orthogonal multiple-access (NOMA) downlink using regularized zero-forcing (RZF) precoding with imperfect channel state information (CSI). We first propose a new user scheduling scheme based on imperfect CSI and a model to characterize the channel correlation between the weak and strong users. Then we derive an approximate expression of the ergodic sum-rate using large-system random matrix theory. This approximation permits us to derive the optimal power allocation scheme that satisfies the rate requirement of the weak users. Simulation results are presented to confirm the accuracy of the approximation and reveal the relationships between the ergodic sum-rate, the channel correlation, and other system parameters.**

***Index Terms*—NOMA, imperfect CSI, regularized zero-forcing precoding, user clustering.**

I. INTRODUCTION

Non-orthogonal multiple-access (NOMA) is regarded as the radio access technique to address the massive connectivity problem for Internet of Things (IoT) [1, 2]. With successive interference cancellation (SIC) at the receiver and superposition coding at the transmitter [3], NOMA has the capability to deliver greater performance, such as higher user fairness and improved spectrum efficiency, when compared to orthogonal multiple access (OMA) methods.

User clustering, scheduling, power allocation, and multiuser beamforming are among the methods that can further improve the sum-rate and energy efficiency of the multiple-input single-output (MISO)-NOMA downlink [4, 5]. To reduce users' inter-cluster interferences, [4] applied zero forcing (ZF) precoding and clustering to MISO-NOMA. In [5], user clustering, beamforming, and power allocation methods were proposed to minimize the transmit power with rate constraints.

In the above works, it is assumed that the base station (BS) has perfect channel state information (CSI). Nevertheless, due to estimation errors, it is impossible to have perfect CSI in practice. In [6], the authors studied the robust beamforming design problem for NOMA in MISO channels by maximizing the worst-case achievable sum-rate. Not only is the knowledge of CSI an important factor affecting the performance of NOMA, the correlation between the user channels in each cluster is another crucial parameter [4, 7]. In [4], the authors provided a user clustering method based on the correlation between the user channels of the MISO-NOMA system. It was shown that the higher the correlation between the channels of two users, the less the inter-cluster interference of the weak user. This work was extended to the imperfect CSI scenario with regularized ZF (RZF) precoding in [7].

The main issue for the two power allocation schemes in [4, 7] is that Monte-Carlo averaging over all channel realizations, which has high computational complexity, is needed. In this letter, we consider a downlink MISO-NOMA system with imperfect CSI using RZF precoding and aim to obtain a closed-form expression of the optimal power allocation by maximizing an approximate expression of the achievable ergodic sum-rate. However, it is difficult to analyze the achievable ergodic sum-rate due to the correlation between the user channels. We tackle this by proposing a model for the channel correlation. Specifically, our contributions are summarized as follows:

- By proposing a model to characterize the channel correlation, we derive a large-system approximate expression for the ergodic sum-rate. This approximation permits us to obtain a closed-form expression for the optimal power allocation, which depends only on the statistical CSI. Therefore, the computational complexity of the proposed method is much lower than that in [4, 7].
- We propose a new user scheduling scheme by exploiting the correlation between the channels and the knowledge of imperfect CSI. We also show the significance of this correlation and reveal that for high correlation between the channels, NOMA can achieve a higher gain than OMA; otherwise, OMA is better.
- The optimal regularization scalar in the RZF precoding is also obtained via a low-complexity linear search method. The simulation results reveal that the optimal regularization scalar monotonically increases with the channel correlation, does not depend on the rate requirement of the weak user, and monotonically increases with the amount of channel uncertainty.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

A downlink multiuser beamforming system comprising one BS and $2K$ users is considered. The BS is equipped with N antennas and serves $2K(\geq N)$ single-antenna users. Users are grouped into K clusters where each cluster is assumed to have two users, i.e., a weak user and a strong user. Note that our results can be easily extended to the case with more than two users per cluster. Let $x_k = \sqrt{p_{k,1}}s_{k,1} + \sqrt{p_{k,2}}s_{k,2}$ denote the superposition coded signal transmitted from the BS to the k-th cluster, where $s_{k,1}$ and $s_{k,2}$ are the signals of strong and weak users, respectively. The signals are assumed independent and identically distributed (i.i.d.) complex data symbols with zero mean and unit variance. Also, $p_{k,1}$ and $p_{k,2}$ denote the power allocation parameters for the strong and weak users, and they satisfy $p_{k,1}+p_{k,2}=1$, for $\forall k$.

Using SIC decoding, the strong user in the k-th cluster first decodes the signal vector of the weak user $s_{k,2}$ by treating $s_{k,1}$ as unknown interference since the signal $\sqrt{p_{k,2}}s_{k,2}$ is stronger than $\sqrt{p_{k,1}}s_{k,1}$ [1, 2]. Then the strong user removes it from the received signal and decodes $s_{k,1}$ from the remaining part of the received signal. Meanwhile, the weak user decodes its own signal $s_{k,2}$ directly by treating $s_{k,1}$ as unknown interference. To do so, the received signals $y_{k,1}$ and $y_{k,2}$ of the strong and weak users in the k-th cluster can be, respectively, given as

$$y_{k,1}=\mathbf{h}_{k,1}^{H}\mathbf{g}_{k}\sqrt{p_{k,1}}s_{k,1}+\mathbf{h}_{k,1}^{H}\sum_{j\neq k}^{K}\mathbf{g}_{j}x_{j}+n_{k,1}, \quad (1)$$

$$y_{k,2}=\mathbf{h}_{k,2}^{H}\mathbf{g}_{k}(\sqrt{p_{k,2}}s_{k,2}+\sqrt{p_{k,1}}s_{k,1}) + \mathbf{h}_{k,2}^{H}\sum_{j\neq k}^{K}\mathbf{g}_{j}x_{j}+n_{k,2}, \quad (2)$$

where $\mathbf{g}_k\in\mathbb{C}^{N\times 1}$ is the precoding vector between the BS and the k-th cluster, $\mathbf{h}_{k,1}^{H}$ and $\mathbf{h}_{k,2}^{H}\in\mathbb{C}^{1\times N}$ denote the channel vectors between BS and the strong/weak users in the k-th cluster, respectively, $n_{k,1}$ and $n_{k,2}$ are i.i.d. complex Gaussian noises with zero mean and variance of σ^2.

We assume that the channels $\mathbf{h}_{k,i}^{H}$ for $i=1,2$ and $k=1,2,\ldots,K$ are modeled as $\mathbf{h}_{k,i}^{H}=\sqrt{\beta_{k,i}}\mathbf{z}_{k,i}^{H}$, where $\beta_{k,1}$ and $\beta_{k,2}$ denote the large-scale fading coefficients between the BS and the strong and weak users in the k-th cluster, respectively, and satisfy $\beta_{k,1}>\beta_{k,2}$, $\mathbf{z}_{k,i}^{H}$ denotes the fast fading channel vector and has i.i.d. entries with zero-mean and variance of $\frac{1}{N}$. We also define $\boldsymbol{\beta}_1=[\beta_{1,1},\ldots,\beta_{K,1}]^T$.

The overall precoding matrix of the BS is denoted by $\mathbf{G}=[\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_K]\in\mathbb{C}^{N\times K}$. We assume that the BS meets the transmit power constraint

$$\mathsf{tr}\{\mathbf{G}\mathbf{G}^H\}\leq NP, \quad (3)$$

where $P>0$ is the power budget of the BS.

It is assumed that the imperfect channels $\hat{\mathbf{h}}_{k,1}$ and $\hat{\mathbf{h}}_{k,2}$ are available at the BS, modeled as

$$\hat{\mathbf{h}}_{k,i}^{H}=\sqrt{\beta_{k,i}}\left(\psi_{k,i}\mathbf{z}_{k,i}^{H}+\tau_{k,i}\mathbf{q}_{k,i}^{H}\right), \quad (4)$$

where $\mathbf{q}_{k,i}^{H}$ has i.i.d. entries with zero-mean and variance of $\frac{1}{N}$ and is independent from $\mathbf{z}_{k,i}^{H}$, and $\tau_{k,i}\in[0,1]$ denotes the amount of uncertainty in $\hat{\mathbf{h}}_{k,i}^{H}$ and $\psi_{k,i}=\sqrt{1-\tau_{k,i}^2}$ for $i=1,2$ and $k=1,2,\ldots,K$. For frequency-division-duplex (FDD) systems, the model (4) reflects the imperfect channel knowledge due to the finite-bandwidth feedback links, whereas for time-division-duplex (TDD) systems, the model (4) reflects the imperfection due to finite training sequence length [8]. We employ the RZF precoding with imperfect CSI of the strong users to reduce the multi-user interference by [8, 9]

$$\mathbf{G}=\xi(\hat{\mathbf{H}}_1^H\hat{\mathbf{H}}_1+\alpha\mathbf{I}_N)^{-1}\hat{\mathbf{H}}_1^H, \quad (5)$$

where $\alpha>0$ represents the regularization scalar, $\hat{\mathbf{H}}_1=[\hat{\mathbf{h}}_{1,1},\hat{\mathbf{h}}_{2,1},\ldots,\hat{\mathbf{h}}_{K,1}]^H\in\mathbb{C}^{K\times N}$, and ξ denotes a normalization parameter to fulfil the BS transmit power constraint (3), hence $\xi^2=\frac{NP}{\mathsf{tr}(\mathbf{W}\hat{\mathbf{H}}_1^H\hat{\mathbf{H}}_1\mathbf{W})}$, where $\mathbf{W}\triangleq(\hat{\mathbf{H}}_1^H\hat{\mathbf{H}}_1+\alpha\mathbf{I}_N)^{-1}$.

Substituting (5) into (1) and (2), the signal-to-interference plus noise ratios (SINRs) of the strong and weak users in the k-th cluster are given, respectively, by

$$\gamma_{k,1}=\frac{p_{k,1}|\mathbf{h}_{k,1}^H\mathbf{W}\hat{\mathbf{h}}_{k,1}|^2}{\mathbf{h}_{k,1}^H\mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H\hat{\mathbf{H}}_{1,[k]}\mathbf{W}\mathbf{h}_{k,1}+\phi}, \quad (6)$$

$$\gamma_{k,2}=\frac{p_{k,2}|\mathbf{h}_{k,2}^H\mathbf{W}\hat{\mathbf{h}}_{k,1}|^2}{p_{k,1}|\mathbf{h}_{k,2}^H\mathbf{W}\hat{\mathbf{h}}_{k,1}|^2+\mathbf{h}_{k,2}^H\mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H\hat{\mathbf{H}}_{1,[k]}\mathbf{W}\mathbf{h}_{k,2}+\phi}, \quad (7)$$

where $\hat{\mathbf{H}}_{1,[k]}=[\hat{\mathbf{h}}_{1,1},\ldots,\hat{\mathbf{h}}_{k-1,1},\hat{\mathbf{h}}_{k+1,1},\ldots,\hat{\mathbf{h}}_{K,1}]^H\in\mathbb{C}^{(K-1)\times N}$, $\phi=\frac{1}{\rho N}\mathsf{tr}(\mathbf{W}\hat{\mathbf{H}}_1^H\hat{\mathbf{H}}_1\mathbf{W})$, and $\rho=\frac{P}{\sigma^2}$ denotes the signal-to-noise ratio (SNR). We can obtain the rates of strong and weak users in the k-th cluster as

$$R_{k,1}=\mathsf{E}_{\mathbf{H}_1,\mathbf{H}_2}\{\log(1+\gamma_{k,1})\}, \quad (8)$$

$$R_{k,2}=\mathsf{E}_{\mathbf{H}_1,\mathbf{H}_2}\{\log(1+\gamma_{k,2})\}. \quad (9)$$

Therefore, the achievable ergodic sum-rate of all users can be expressed as

$$R_{\mathsf{sum}}=R_{\mathsf{sum}_1}+R_{\mathsf{sum}_2}, \quad (10)$$

where $R_{\mathsf{sum}_1}=\sum_{k=1}^K R_{k,1}$ and $R_{\mathsf{sum}_2}=\sum_{k=1}^K R_{k,2}$.

B. Problem Formulation

In this letter, we aim to maximize the achievable ergodic sum-rate under the ergodic minimum rate constraint for the weak users and the transmit power constraint by finding the optimal power allocation $\{p_{k,1}^{\mathsf{opt}},p_{k,2}^{\mathsf{opt}}\}_{\forall k}$ and the regularization parameter α^{opt}. This problem can be formulated as

$$\max_{\{p_{k,1}\}_{\forall k},\alpha} R_{\mathsf{sum}} \quad (11)$$
$$\text{s.t.}\quad R_{k,2}\geq R_{k,0}, p_{k,1}+p_{k,2}=1, 0\leq p_{k,1},p_{k,2}\leq 1, \forall k,$$

where $R_{k,0}$ denotes the minimum rate requirement of the weak user in the k-th cluster to ensure the quality of service. Notice that the power constraint (3) is absorbed into R_{sum} by ξ. However, from (8) and (9), it is required to evaluate the achievable ergodic sum-rate R_{sum} using Monte-Carlo methods averaging over all channel realizations which have very high computational complexity. To tackle this challenge, we present an approach to solve the problem (11) in the next section.

III. Performance Analysis

A. User Selection and Channel Modeling of Weak Users

In power-domain NOMA, to enhance system performance and reduce interference, a user-selected scheme is necessary before signal transmission, which includes the gain-difference and correlation between the user channels in each cluster.

The high correlation between the channels of users in each cluster can improve the sum-rate of NOMA systems [4]. In the k-th cluster, the correlation between the channels of the weak and strong users Corr_k is expressed by

$$\mathsf{Corr}_k = \frac{|\mathbf{h}_{k,1}^H \mathbf{h}_{k,2}|}{|\mathbf{h}_{k,1}^H||\mathbf{h}_{k,2}^H|} = \frac{|\mathbf{z}_{k,1}^H \mathbf{z}_{k,2}|}{|\mathbf{z}_{k,1}^H||\mathbf{z}_{k,2}^H|}, \tag{12}$$

for $k = 1, 2, \ldots, K$. However, only imperfect channels $\hat{\mathbf{h}}_{k,1}^H$ and $\hat{\mathbf{h}}_{k,2}^H$ are available at the BS. Therefore, we use the following approximation

$$\widehat{\mathsf{Corr}}_k = \frac{|\hat{\mathbf{h}}_{k,1}^H \hat{\mathbf{h}}_{k,2}|}{\psi_{k,1}\psi_{k,2}|\hat{\mathbf{h}}_{k,1}^H||\hat{\mathbf{h}}_{k,2}^H|}. \tag{13}$$

Thus, a new user scheduling scheme is proposed, i.e., the selected users in a cluster satisfy the following conditions

$$\widehat{\mathsf{Corr}}_k \geq \mathsf{Corr}_0 \quad \text{and} \quad |\beta_{k,1}\psi_{k,1}^2 - \beta_{k,2}\psi_{k,2}^2| \geq \beta_0, \quad \text{for} \quad \forall k, \tag{14}$$

where β_0 and $\mathsf{Corr}_0 \in [0,1]$ are the prescribed large-scale and correlation thresholds, respectively.

It is well known that to deal with the term $\mathbf{x}^H\mathbf{A}\mathbf{x}$ using large dimension random matrix theory, we always require that $\mathbf{x}$ and $\mathbf{A}$ are independent [10, Lemma 2.3]. Otherwise, it would be required to separate the component depending on $\mathbf{x}$ from $\mathbf{A}$ [10, Lemma 2.1]. However, it is difficult to separate the component depending on $\mathbf{h}_{k,2}$ from $\mathbf{W}$ in the numerator and denominator of (7) due to the correlation between $\mathbf{h}_{k,1}$ and $\mathbf{h}_{k,2}$. To tackle this challenge, we propose a model to characterize the relationship between the fast fading channels of the weak and strong users in each cluster as

$$\mathbf{z}_{k,2}^H = \theta_k \mathbf{z}_{k,1}^H + \sqrt{1-\theta_k^2}\mathbf{v}_k^H, \tag{15}$$

where $\mathbf{v}_k^H$ has i.i.d. entries with zero-mean and variance of $\frac{1}{N}$ and is independent from $\mathbf{z}_{k,1}^H$ and $\mathbf{q}_k^H$, and θ_k is a constant.

Substituting (15) into (12), we therefore get the correlation coefficient

$$\mathsf{Corr}_k = \frac{|\theta_k \mathbf{z}_{k,1}^H \mathbf{z}_{k,1} + \sqrt{1-\theta_k^2}\mathbf{z}_{k,1}^H \mathbf{v}_k|}{|\mathbf{z}_{k,1}^H||\theta_k \mathbf{z}_{k,1}^H + \sqrt{1-\theta_k^2}\mathbf{v}_k^H|}. \tag{16}$$

By applying [10, Lemma 2.3] and the fact that $\mathbf{z}_{k,1}^H$ and $\mathbf{v}_k^H$ are independent, we have

$$\mathsf{Corr}_k - \theta_k \xrightarrow{a.s.} 0, \ \text{as} \ N \to \infty, \tag{17}$$

where a.s. denotes 'almost sure' convergence. Similarly,

$$\widehat{\mathsf{Corr}}_k - \theta_k \xrightarrow{a.s.} 0, \ \text{as} \ N \to \infty. \tag{18}$$

It is confirmed that the proposed model in (15) can asymptotically characterize the correlation between the channels of the weak and strong users through the parameter θ_k and we can use $\widehat{\mathsf{Corr}}_k$ to approximate Corr_k. Fig. 1 shows the relationship between the correlation coefficient Corr_k (or $\widehat{\mathsf{Corr}}_k$) and the parameter θ_k for different number of antennas, which is consistent with our analytical results in (17) and (18).

B. Large System Analysis

Using (15), we aim to obtain an approximate expression of (10) using large dimensional random matrix theory. First, it is assumed that N and $K \to \infty$ with the ratio $c = \frac{K}{N}$.

According to [8, Theorem 1] and taking some mathematical derivations, we can obtain the following theorem.

Theorem 1: As $N \to \infty$, we have $\gamma_{k,1} - \overline{\gamma}_{k,1} \xrightarrow{a.s.} 0$ and $\gamma_{k,2} - \overline{\gamma}_{k,2} \xrightarrow{a.s.} 0$, for $k = 1, \ldots, K$, where

$$\overline{\gamma}_{k,1} = \frac{p_{k,1}\psi_{k,1}^2 e_k}{\left(1 - \frac{2\psi_{k,1}^2 e_k}{1+e_k} + \frac{\psi_{k,1}^2 e_k^2}{(1+e_k)^2} + \frac{1}{\rho\beta_{k,1}}\right) u_k}, \tag{19}$$

$$\overline{\gamma}_{k,2} = \frac{p_{k,2}\theta_k^2\psi_{k,1}^2 e_k}{p_{k,1}\theta_k^2\psi_{k,1}^2 e_k + \left(1 - \frac{2\theta_k^2\psi_{k,1}^2 e_k}{1+e_k} + \frac{\theta_k^2\psi_{k,1}^2 e_k^2}{(1+e_k)^2} + \frac{1}{\rho\beta_{k,2}}\right) u_k}, \tag{20}$$

with

$$u_k = (1+e_k)^2\left(1 - \frac{\alpha e_k}{\beta_{k,1}}(1 - \boldsymbol{\beta}_1^T \boldsymbol{\Theta}^{-1}\boldsymbol{\eta})\right), \tag{21a}$$

$$[\boldsymbol{\Theta}]_{kl} = \begin{cases} \dfrac{-1}{N}\dfrac{e_k e_l}{(e_k+1)^2}, & \text{for } k \neq l; \\ 1 - \dfrac{1}{N}\dfrac{e_k^2}{(e_k+1)^2}, & \text{for } k = l, \end{cases} \tag{21b}$$

$$[\boldsymbol{\eta}]_k = -\frac{1}{N}\frac{e_k^2}{(e_k+1)^2} \tag{21c}$$

and e_k's are the unique solution of the following K equations

$$e_k = \frac{\beta_{k,1}}{\frac{1}{N}\sum_{k=1}^K \frac{\beta_{k,1}}{e_k+1} + \alpha} \ \text{for} \ k = 1, \ldots, K. \tag{22}$$

Proof: For $\gamma_{k,1}$ in (6), using [8, Theorem 1] (setting the number of BSs $M = 1$ and correlation matrix $\mathbf{T}_k = \beta_{k,1}\mathbf{I}$ in [8]), we can obtain

$$\mathbf{h}_{k,1}^H \mathbf{W}\hat{\mathbf{h}}_{k,1} - \frac{\psi_{k,1}e_k}{1+e_k} \xrightarrow{a.s.} 0, \tag{23a}$$

$$\mathbf{h}_{k,1}^H \mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W}\mathbf{h}_{k,1} - \left(1 - \frac{2\psi_{k,1}^2 e_k}{1+e_k} + \frac{\psi_{k,1}^2 e_k^2}{(1+e_k)^2}\right) \times \left(e_k - \frac{\alpha e_k^2}{\beta_{k,1}}(1 - \boldsymbol{\beta}_1^T\boldsymbol{\Theta}^{-1}\boldsymbol{\eta})\right) \xrightarrow{a.s.} 0, \tag{23b}$$

$$\phi - \frac{1}{\rho\beta_{k,1}}\left(e_k - \frac{\alpha e_k^2}{\beta_{k,1}}(1 - \boldsymbol{\beta}_1^T\boldsymbol{\Theta}^{-1}\boldsymbol{\eta})\right) \xrightarrow{a.s.} 0, \tag{23c}$$

where $\boldsymbol{\Theta}$, $\boldsymbol{\eta}$, and e_k are given by (21b), (21c), and (22), respectively. Substituting (23) into (6), we can obtain (19).

For $\gamma_{k,2}$ in (7), substituting (15) into (7), we get

$$\mathbf{h}_{k,2}^H \mathbf{W}\hat{\mathbf{h}}_{k,1} = \sqrt{\delta_k}\theta_k \mathbf{h}_{k,1}^H \mathbf{W}\hat{\mathbf{h}}_{k,1} + \sqrt{\beta_{k,2}}\sqrt{1-\theta_k^2}\mathbf{v}_k^H \mathbf{W}\hat{\mathbf{h}}_{k,1} \tag{24}$$

$$\mathbf{h}_{k,2}^H \mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W}\mathbf{h}_{k,2} = \delta_k\theta_k^2 \mathbf{h}_{k,1}^H \mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W}\mathbf{h}_{k,1} + \beta_{k,2}(1-\theta_k^2)\mathbf{v}_k^H \mathbf{W}\hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W}\mathbf{v}_k, \tag{25}$$

where $\delta_k = \frac{\beta_{k,2}}{\beta_{k,1}}$. Using [10, Lemmas 2.1–2.3], we have

$$\mathbf{v}_k^H \mathbf{W}\hat{\mathbf{h}}_{k,1} \xrightarrow{a.s.} 0 \tag{26}$$

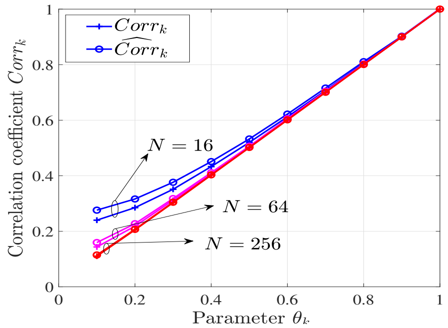

Fig. 1. The relationship between the correlation coefficient Corr_k and the parameter θ_k.

$$\mathbf{v}_k^H \mathbf{W} \hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W} \mathbf{v}_k \xrightarrow{a.s.} \frac{1}{N} \mathsf{tr}\left(\mathbf{W} \hat{\mathbf{H}}_1^H \hat{\mathbf{H}}_1 \mathbf{W} \right) = \rho\phi. \tag{27}$$

Substituting (23a) and (26) into (24), we get

$$\mathbf{h}_{k,2}^H \mathbf{W} \hat{\mathbf{h}}_{k,1} - \sqrt{\delta_k} \frac{\theta_k \psi_{k,1} e_k}{1+e_k} \xrightarrow{a.s.} 0. \tag{28}$$

Substituting (23b), (23c), and (27) into (25), we also have

$$\mathbf{h}_{k,2}^H \mathbf{W} \hat{\mathbf{H}}_{1,[k]}^H \hat{\mathbf{H}}_{1,[k]} \mathbf{W} \mathbf{h}_{k,2} - \left(1 - \frac{2\theta_k^2 \psi_{k,1}^2 e_k}{1+e_k} + \frac{\theta_k^2 \psi_{k,1}^2 e_k^2}{(1+e_k)^2} \right) \times \delta_k \left(e_k - \frac{\alpha e_k^2}{\beta_{k,1}} (1 - \boldsymbol{\beta}_1^T \boldsymbol{\Theta}^{-1} \boldsymbol{\eta}) \right) \xrightarrow{a.s.} 0. \tag{29}$$

Combining (23c), (28), and (29), we establish (20). ■

The large-system approximation in Theorem 1 provides accurate estimates for the ergodic sum-rate even for small numbers of antennas. According to the continuous mapping theorem [11, Theorem 25.7–Corollary 2], we get a large-system approximation $\bar{R}_{\mathsf{sum}}$ for R_{sum} in (10), such that $R_{\mathsf{sum}} - \bar{R}_{\mathsf{sum}} \xrightarrow{a.s.} 0$, as $N \to \infty$, where

$$\bar{R}_{\mathsf{sum}} = \sum_{k=1}^{K} (\bar{R}_{k,1} + \bar{R}_{k,2}), \tag{30}$$

$\bar{R}_{k,1} = \log(1+\overline{\gamma}_{k,1})$, and $\bar{R}_{k,2} = \log(1+\overline{\gamma}_{k,2})$. Thus, we recast the problem (11) into

$$\max_{\{p_{k,1}\}_{\forall k}, \alpha} \bar{R}_{\mathsf{sum}} \tag{31}$$
$$\text{s.t.} \quad R_{k,2} \geq R_{k,0}, p_{k,1} + p_{k,2} = 1, 0 \leq p_{k,1}, p_{k,2} \leq 1, \forall k,$$

C. Solving the Optimization Problem

From (30), we see that $\bar{R}_{\mathsf{sum}}$ only depends on the statistical CSI, such as the large-scale fading coefficients $\{\beta_{k,1}, \beta_{k,2}\}_{\forall k}$, the amount of uncertainty of the channel $\{\tau_k\}_{\forall k}$, and the correlation between the channels of the weak and strong users $\{\theta_k\}_{\forall k}$. However, the optimal regularization scalar α^{opt} still does not permit closed-form solutions. Thus, we decompose the joint optimization problem (31) into two subproblems:

1) Given that $\{p_{k,1}\}_{\forall k}$ is fixed, the optimal regularization scalar $\alpha^{\mathsf{opt}} := \arg\max_{\bar{R}_{k,2} \geq R_{k,0}} \bar{R}_{\mathsf{sum}}$ can be obtained efficiently using one-dimensional linear search.

2) For a fixed α^{opt}, $\{p_{k,1}\}_{\forall k}^{\mathsf{opt}}$ satisfies the following optimization problem

$$\max_{\{p_{k,1}\}_{\forall k}} \bar{R}_{\mathsf{sum}} \tag{32}$$
$$\text{s.t.} \quad R_{k,2} \geq R_{k,0}, p_{k,1} + p_{k,2} = 1, 0 \leq p_{k,1}, p_{k,2} \leq 1, \forall k.$$

According to (32), (19), and (20), we find that $p_{k,1}$ only depends on the sum-rate in the k-th cluster. Therefore, the optimization problem (31) can be equivalent to the maximization of the sum-rate in each cluster, separately. That is, for $\forall k$, $p_{k,1}^{\mathsf{opt}}$ can be derived by the following problem

$$\max_{p_{k,1}} \bar{R}_{k,1} + \bar{R}_{k,2} \tag{33}$$
$$\text{s.t.} \quad R_{k,2} \geq R_{k,0}, p_{k,1} + p_{k,2} = 1, 0 \leq p_{k,1}, p_{k,2} \leq 1.$$

By taking the second derivative of $\bar{R}_{k,1} + \bar{R}_{k,2}$ with respect to $p_{k,1}$, we have $\frac{\partial^2 (\bar{R}_{k,1}+\bar{R}_{k,2})}{\partial p_{k,1}^2} = \Big(p_{k,1} + \big((1+\frac{1}{\rho\beta_{k,2}})\frac{1}{\theta_k^2 \psi_{k,1}^2 e_k} - \frac{2}{1+e_k} + \frac{e_k}{(1+e_k)^2}\big)u_k\Big)^{-2} - \Big(p_{k,1} + \big((1+\frac{1}{\rho\beta_{k,1}})\frac{1}{\psi_{k,1}^2 e_k} - \frac{2}{1+e_k} + \frac{e_k}{(1+e_k)^2}\big)u_k\Big)^{-2}$. Since $\beta_{k,1} > \beta_{k,2}$ and $\theta_k \leq 1$, we get $\frac{\partial^2 (\bar{R}_{k,1}+\bar{R}_{k,2})}{\partial p_{k,1}^2} \leq 0$. Thus, the above problem is convex. From the Karush-Kuhn-Tucker (KKT) conditions, we can derive the optimal power allocation factor $p_{k,1}^{\mathsf{opt}}$ as

$$p_{k,1}^{\mathsf{opt}} = \frac{1}{2^{R_{k,0}}} - \frac{(2^{R_{k,0}}-1)u_k}{2^{R_{k,0}} \theta_k^2 \psi_{k,1}^2 e_k} \times \left(1 - \frac{2\theta_k^2 \psi_{k,1}^2 e_k}{1+e_k} + \frac{\theta_k^2 \psi_{k,1}^2 e_k^2}{(1+e_k)^2} + \frac{1}{\rho\beta_{k,2}} \right). \tag{34}$$

Substituting $p_{k,1}^{\mathsf{opt}}$ in (34) and $p_{k,2}^{\mathsf{opt}} = 1 - p_{k,1}^{\mathsf{opt}} (\forall k)$ into (19) and (20), we get $\bar{R}_{\mathsf{sum}}(\alpha)$ from (30), where $\bar{R}_{\mathsf{sum}}(\alpha)$ only depends on α but does not depends on $p_{k,1}$ and $p_{k,2}$ since $p_{k,1}^{\mathsf{opt}}$ in (34) is a closed-form solution. Thus, we can directly obtain $\alpha^{\mathsf{opt}} := \arg\max \bar{R}_{\mathsf{sum}}(\alpha)$ using one-dimensional linear search, where the minimum rate constraint for the weak users vanishes since it is absorbed into $p_{k,1}^{\mathsf{opt}}$. Then, the unique solution $\{e_k\}_{\forall k}$ is calculated by substituting α^{opt} into (22). Finally, we can obtain $p_{k,1}^{\mathsf{opt}}$ using (34). Hence, we can obtain α^{opt} and $p_{k,1}^{\mathsf{opt}}$ by using an alternating optimization algorithm. Since the calculation of α or $p_{k,1}$ is along the monotonically increasing direction of $\bar{R}_{\mathsf{sum}}(\alpha, p_{k,1})$ at each step, the alternating algorithm is guaranteed to converge.

IV. Numerical Results

In this section, we provide some simulation results to evaluate the proposed power allocation scheme. In the simulations, we consider that the number of antennas at the BS is $N = 64$, the number of users is 128 ($K = 64$), and we have the large-scale fading factors $\beta_{k,i} = 128.1 + 37.6 \log 10(r_{k,i})$ dB, where $r_{k,i}$ (km) is the distance between user i in the k-th cluster and the BS. Also, the transmit power at the BS is 30dBm, the noise density at the users is -169dBm/Hz, and the bandwidth is 10MHz. All users are randomly located with uniform distribution in the cell with 1km radius.

In Fig. 2, we illustrate the ergodic sum-rate and the optimal regularization scalar α^{opt} versus θ_k with $\{\tau_{k,1}^2 = \tau_{k,2}^2 =$

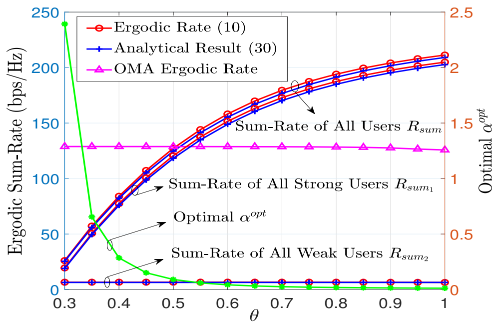

Fig. 2. Achievable ergodic sum-rate and optimal α^{opt} *vs* θ_k.

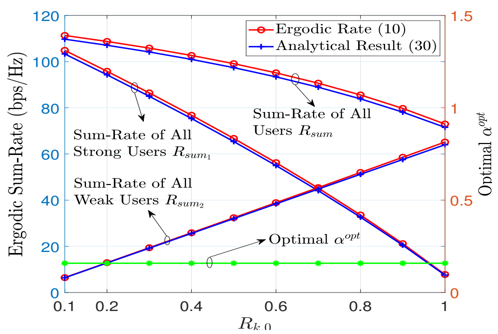

Fig. 3. Achievable ergodic sum-rate and optimal α^{opt} *vs* $R_{k,0}$.

$0, R_{k,0} = 0.1\}$, respectively. It can be seen that the analytical results (blue curves) almost agree with the simulation results (red curves) achieved by Monte-Carlo averaging over 10^4 independent channel realizations of $\{\mathbf{H}_1, \mathbf{H}_2\}$ even with small numbers of antennas. Fig. 2 also compares the performances of NOMA and OMA. We see that when the correlation between the channels of the weak and strong users θ_k increases, the ergodic sum-rate of NOMA increases while α^{opt} decreases monotonically. It means that the ZF procoding is near-optimal for high correlation. However, the ergodic sum-rate of OMA is almost unaffected by correlation. It reveals that for high correlation between the channels of the weak and strong users, NOMA can achieve a better gain than OMA; otherwise, OMA is better. This insight is different from the many existing works, which infer that NOMA is always better than OMA.

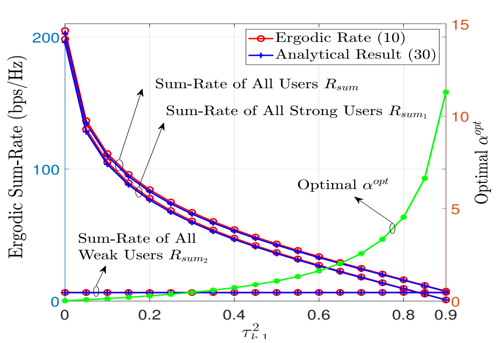

Fig. 4. Achievable ergodic sum-rate and optimal α^{opt} *vs* $\tau_{k,1}^2$.

Fig. 3 depicts the ergodic sum-rate versus $R_{k,0}$ with $\{\tau_{k,1}^2 = \tau_{k,2}^2 = 0.1, \theta_k = 0.85\}$. We can see that with increasing $R_{k,0}$, the sum-rate decreases. This is due to the fact that when $R_{k,0}$ increases, more power is allocated to the weak users in order to satisfy the constraint $\{\bar{R}_{k,2} \geq R_{k,0}\}_{\forall k}$. It causes the sum-rate of all strong users to drop even more. Fig. 3 shows that α^{opt} does not depend on $R_{k,0}$ since α only controls the inter-cluster user interference while does not balance the rates between the strong user and weak user in each cluster.

Finally, the ergodic sum-rate results versus $\tau_{k,1}^2$ with $\{\tau_{k,1}^2 = \tau_{k,2}^2, \theta_k = 0.85, R_{k,0} = 0.1\}$ are illustrated in Fig. 4. We observe that higher performance is achieved if more CSI is available at the BS and α^{opt} monotonically increases with increasing $\tau_{k,1}^2$.

V. Conclusion

We addressed the achievable ergodic sum-rate maximization problem subject to the rate constraint of the weak user in the MISO-NOMA system with imperfect CSI at the BS employing RZF precoding. We proposed a model to characterize the correlation between the fast fading channels of the weak and strong users and derived a large-system approximate expression of the achievable ergodic sum-rate. By maximizing this approximate expression, a closed-form solution of the optimal power allocation has been derived.

References

[1] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[2] K. Yang, N. Yang, N. Ye, M. Jia, Z. Gao, and R. Fan, "Non-orthogonal multiple access: Achieving sustainable future radio access," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 116–121, Feb. 2019.

[3] S. Liu, C. Zhang, and G. Lyu, "User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system," in *IEEE Int. Conf. Commun. Workshop*, London, UK, 2015, pp. 2561–2565.

[4] B. Kim, S. Lin, and C. Lin, "Non-orthogonal multiple access in a downlink multi-user beamforming system," in *IEEE Military Commun. Conf. (MILCOM)*, San Diego, USA, 2013, pp. 1278–1283.

[5] Y. Jeong, C. Lee, and Y. H. Kim, "Power minimizing beamforming and power allocation for MISO-NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6187–6191, Jun. 2019.

[6] Q. Zhang, Q. Li, and J. Qi, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10 231–10 236, Dec. 2016.

[7] X. Liu, J. Zhang, and S. Cai, "An optimal power allocation scheme in downlink multi-user NOMA beamforming system with imperfect CSI," in *IEEE Conf. on Communi. Systems (ICCS)*, Chengdu, China, 2018, pp. 1–5.

[8] J. Zhang, C.-K. Wen, S. Jin, X. Q. Gao, and K.-K. Wong, "Large system analysis of cooperative multi-cell downlink transmission via regularized channel inversion with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4801–4813, Oct. 2013.

[9] J. Zhang, C. Yuen, C.-K. Wen, S. Jin, and X. Gao, "Ergodic secrecy sum-rate for multiuser downlink transmission via regularized channel inversion: Large system analysis," *IEEE Commun. Letters*, vol. 18, no. 9, pp. 1627–1630, Sep. 2014.

[10] Z. Bai, Y. Chen, and Y.-C. Liang, *Random Matrix Theory and its Applications*. World Scientific Publishing Company, 2009.

[11] P. Billingsley, *Probability and Measure*. Hoboken, NJ, USA: Wiley, 2011.