SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2018, Vol. 61 042303:1–042303:14
doi: 10.1007/s11432-016-9126-1

Robust MMSE precoding for massive MIMO transmission with hardware mismatch

Yan CHEN[1], Xiqi GAO[1*], Xiang-Gen XIA[2] & Li YOU[1]

[1]*National Mobile Communication Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[2]*Department of Electrical and Computer Engineering, University of Delaware, Newark* DE 19716*, USA*

Received 20 October 2016/Revised 21 February 2017/Accepted 5 May 2017/Published online 20 November 2017

Abstract Due to hardware mismatch, the channel reciprocity of time-division duplex massive multiple-input multiple-output system is impaired. Under this condition, there exist several different approaches for base station (BS) to obtain downlink (DL) channel information based on the minimum mean-square-error (MMSE) estimation method. In this paper, we show that with the hardware mismatch parameters BS can obtain the same DL channel information via these different approaches. As the obtained DL channel information is usually imperfect, we propose a precoding technique based on the criterion that minimizes the mean-square-error (MSE) of signal detection at the user terminals (UTs). The proposed precoding is robust to the channel estimation error and significantly improves the system performance compared to the conventional regularized zero-forcing precoding. Furthermore, we derive an asymptotic approximation of the ergodic sum rate for the proposed precoding using the large dimensional random matrix theory, which is tight as the number of antennas both at the BS and UT approach infinity with a fixed non-zero and finite ratio. This approximation can provide a reliable sum rate prediction at a much lower computation cost than Monte Carlo simulations. Simulation results show that the approximation is accurate even for a realistic system dimension.

Keywords robust precoding, massive MIMO, hardware mismatch, channel estimation, large dimensional RMT

Citation Chen Y, Gao X Q, Xia X-G, et al. Robust MMSE precoding for massive MIMO transmission with hardware mismatch. Sci China Inf Sci, 2018, 61(4): 042303, doi: 10.1007/s11432-016-9126-1

1 Introduction

Massive multiple-input multiple-output (MIMO) systems, equipped with a large number of antennas at a base station (BS) to serve several user terminals (UTs) simultaneously, show the potential of significant improvement in both spectral and energy efficiency [1]. Therefore, massive MIMO has become a promising technology for next generation wireless communication systems, and has attracted tremendous research interest recently [2, 3].

In massive MIMO systems, the downlink (DL) channel state information (CSI) is required at the BS for multi-user precoding [4]. For time-division duplex (TDD) operation, the BS can obtain the DL channel information directly from the uplink (UL) channel estimate which is obtained from UL training based on the channel reciprocity [1, 5]. The channel reciprocity, however, is usually impaired, because the hardware employed in the transmitting and receiving modules of the BS and UT antennas are usually imperfect [6, 7]. This results in serious degradation the massive MIMO system performance [8–10]. In

* Corresponding author (email: xqgao@seu.edu.cn)

recent studies, several estimation and calibration methods of the hardware mismatch parameters have been investigated [11–16]. When one compares partial calibration [11, 15, 17] and full calibration [12, 14, 16, 18], there is a tradeoff between complexity and performance. The Full-Pre-precoding Calibration scheme in [8] is shown to be able to suit any precoding scheme.

With the knowledge of full DL CSI, the interference can be efficiently mitigated by appropriate precoding techniques at the BS. In the case of UTs equipped with single antenna, linear precoding techniques such as zero-forcing [19], regularized zero-forcing (RZF) [20] or transmit Wiener filter [21], which are shown to be near-optimal with low complexity, are more practical in massive MIMO systems compared to non-linear precoding [22, 23]. However, the DL channel information at the BS is always imperfect. Thus, it is essential to derive a linear precoding robust to the channel estimation error, which is the focus of our work.

After reciprocity calibration, the impact of the residual hardware mismatch impairments on DL transmission performance is studied in [9, 24]. However, these researches focus on the additive distortions introduced by the residual hardware mismatch impairments. The effect of hardware mismatch on the DL sum rate is investigated in [8, 25], but the results are based on the conventional linear precoding.

In this paper, we investigate several approaches for BS to obtain the DL channel coefficients from UL training based on the minimum mean-square-error (MMSE) estimation theory. We show that BS obtains the same DL channel estimate via these different approaches with hardware mismatch parameters, which means that the DL channel information obtained with relative calibration matrices [8, 17] is optimal for MMSE criterion. Based on the imperfect channel estimate with estimation error, we further propose a robust precoding technique which can minimize the mean-square-error (MSE) of signal detection at the UTs, achieving significant system performance improvements compared with the conventional RZF precoding. Furthermore, we derive an asymptotic approximation of the ergodic sum rate for the proposed precoding using the large dimensional random matrix theory, which is tight as the number of antennas both at the BS and user terminals approach infinity with a fixed non-zero and finite ratio.

Taking the hardware mismatch into account this analytical expression is for the proposed robust precoding, which is different from the results of [8, 9]. This asymptotic approximation can show the effect of hardware mismatch on the DL sum rate and provide a reliable sum rate prediction at a much lower computation cost than Monte Carlo simulations. Simulation results show that the approximation is accurate even for a realistic system dimension.

The rest of the paper is organized as follows. In Section 2, the system model accounting for RF circuits is described first, then the DL channel estimation approaches are studied and compared. In Section 3, the proposed robust MMSE precoding technique is developed. An asymptotic approximation of ergodic sum rate is derived in Section 4. Simulation results are shown in Section 5, and the conclusion is drawn in Section 6.

Throughout this paper, upper (lower) case boldface letters are used to denote matrices (column vectors). $\boldsymbol{I}_N$ stands for the N by N identity matrix. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ denote conjugate, transpose and Hermitian transpose, respectively. $\mathrm{E}\left\{\cdot\right\}$ represents the expectation operator. $\mathcal{CN}\left(\boldsymbol{m}, \boldsymbol{C}\right)$ is used to denote the circular symmetric complex Gaussian distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$. $\mathcal{U}\left(a, b\right)$ denotes the uniform distribution on the interval $[a, b]$. We use $\|\boldsymbol{A}\|$ to denote the spectral norm of a matrix $\boldsymbol{A}$. $\mathrm{tr}\left(\boldsymbol{A}\right)$ is the trace of $\boldsymbol{A}$. $\mathrm{vec}\left(\boldsymbol{A}\right)$ is the column vector obtained by stacking the columns of $\boldsymbol{A}$. $\mathrm{diag}\left\{a_1, \ldots, a_N\right\}$ represents an N by N diagonal matrix with diagonal entries given by $a_1, \ldots, a_N$.

2 System model

In this section, we first introduce the channel model which includes the consideration of the RF circuits, and we then study different estimation approaches based on the MMSE estimation method which can be used for the BS to obtain the DL channel estimate.

2.1 Channel model

Consider a single cell TDD massive MIMO transmission, where the BS equipped with N antennas serves $K(\leqslant N)$ single-antenna noncooperative UTs. We assume that channels vary in time according to the block fading model. We consider narrow-band transmission.

In reality, if the RF circuits in the transmitting and receiving modules of the BS and UTs antennas are taken into account, the UL channel matrix $\boldsymbol{H}_{\mathrm{u}} \in \mathbb{C}^{N\times K}$ and the DL channel matrix $\boldsymbol{H}_{\mathrm{d}} \in \mathbb{C}^{K\times N}$ can be modeled as [7, 17]

$$\boldsymbol{H}_{\mathrm{u}} = \boldsymbol{R}_{\mathrm{b}}\boldsymbol{V}_{\mathrm{u}}\boldsymbol{T}_{\mathrm{u}}, \quad \boldsymbol{H}_{\mathrm{d}} = \boldsymbol{R}_{\mathrm{u}}\boldsymbol{V}_{\mathrm{d}}\boldsymbol{T}_{\mathrm{b}}, \tag{1}$$

where $\boldsymbol{R}_{\mathrm{b}} = \mathrm{diag}\{r_{\mathrm{b}1}, \ldots, r_{\mathrm{b}N}\}$ and $\boldsymbol{T}_{\mathrm{b}} = \mathrm{diag}\{t_{\mathrm{b}1}, \ldots, t_{\mathrm{b}N}\}$ are diagonal matrices of complex gains, introduced by the BS side receive and transmit RF circuits, respectively; $\boldsymbol{R}_{\mathrm{u}} = \mathrm{diag}\{r_{\mathrm{u}1}, \ldots, r_{\mathrm{u}K}\}$ and $\boldsymbol{T}_{\mathrm{u}} = \mathrm{diag}\{t_{\mathrm{u}1}, \ldots, t_{\mathrm{u}K}\}$ are receive and transmit RF circuit gains introduced by the UT side; $\boldsymbol{V}_{\mathrm{u}}$ and $\boldsymbol{V}_{\mathrm{d}}$ are the corresponding UL and DL physical wireless channel coefficients, respectively.

In a TDD system, it is true that $\boldsymbol{V}_{\mathrm{d}} = \boldsymbol{V}_{\mathrm{u}}^{\mathrm{T}}$ due to the wireless channel reciprocity. However, for the channel model (1), if the RF circuits are mismatched, i.e., $\boldsymbol{R}_{\mathrm{b}} \neq \boldsymbol{T}_{\mathrm{b}}$ and $\boldsymbol{R}_{\mathrm{u}} \neq \boldsymbol{T}_{\mathrm{u}}$, then $\boldsymbol{H}_{\mathrm{d}} \neq \boldsymbol{H}_{\mathrm{u}}^{\mathrm{T}}$. In other words, the channel reciprocity does not hold for channel model (1) owning to the RF circuit mismatch. The relation between the DL channel and the UL channel can be written as $\boldsymbol{H}_{\mathrm{d}} = \boldsymbol{R}_{\mathrm{u}}\boldsymbol{T}_{\mathrm{u}}^{-1}\boldsymbol{H}_{\mathrm{u}}^{\mathrm{T}}\boldsymbol{R}_{\mathrm{b}}^{-1}\boldsymbol{T}_{\mathrm{b}}$.

2.2 Channel estimation

For multi-user MIMO systems, DL channel coefficients are required at the BS used for precoding. In TDD systems, the BS can obtain the DL channel estimate from UL training. During the UL training phase, the combined received signal matrix $\boldsymbol{Y}_{\mathrm{r}} \in \mathbb{C}^{N\times K}$ can be expressed as

$$\boldsymbol{Y}_{\mathrm{r}}^{\mathrm{T}} = \boldsymbol{S}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{u}}^{\mathrm{T}} + \boldsymbol{N}^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{S} \in \mathbb{C}^{K\times L}$ represents the known training matrix, $\boldsymbol{N} \in \mathbb{C}^{N\times L}$ is the independent additive Gaussian noise matrix with elements distributed as independently and identically $\mathcal{CN}\left(0, \sigma_{\mathrm{t}}^2\right)$, L is the training interval length, and σ_{t}^2 is the noise power during the training phase. The transpose operation is used for convenience. We assume that orthogonal training sequences are used, i.e., $\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} = \sigma_{\mathrm{s}}^2\boldsymbol{I}_k$. After decorrelation and power normalization, the received signal can be rewritten as

$$\boldsymbol{Y}_{\mathrm{t}}^{\mathrm{T}} = \frac{1}{\sigma_{\mathrm{s}}^2}\boldsymbol{S}^*\boldsymbol{Y}_{\mathrm{r}}^{\mathrm{T}} = \boldsymbol{H}_{\mathrm{u}}^{\mathrm{T}} + \frac{1}{\sigma_{\mathrm{s}}^2}\boldsymbol{S}^*\boldsymbol{N}^{\mathrm{T}}. \tag{3}$$

Based on the property of unitary transformation, the noise term $\frac{1}{\sigma_{\mathrm{s}}^2}\boldsymbol{S}^*\boldsymbol{N}^{\mathrm{T}}$ in (3) is still independent Gaussian with elements distributed as independently and identically $\mathcal{CN}(0, \frac{1}{\rho_{\mathrm{t}}})$ in which we define the UL channel training signal-to-noise ration (SNR) as $\rho_{\mathrm{t}} = \sigma_{\mathrm{s}}^2/\sigma_{\mathrm{t}}^2$.

By vectorizing the signal of (3) and denoting $\boldsymbol{y}_{\mathrm{t}} = \mathrm{vec}(\boldsymbol{Y}_{\mathrm{t}}^{\mathrm{T}})$, $\boldsymbol{h}_{\mathrm{u}} = \mathrm{vec}(\boldsymbol{H}_{\mathrm{u}}^{\mathrm{T}})$, $\boldsymbol{h}_{\mathrm{d}} = \mathrm{vec}(\boldsymbol{H}_{\mathrm{d}})$, $\boldsymbol{v}_{\mathrm{w}} = \mathrm{vec}(\boldsymbol{V}_{\mathrm{u}}^{\mathrm{T}})$ and $\boldsymbol{n}_{\mathrm{t}} = \mathrm{vec}(\frac{1}{\sigma_{\mathrm{s}}^2}\boldsymbol{S}^*\boldsymbol{N}^{\mathrm{T}})$, the channel observation in (3) can be rewritten as

$$\begin{aligned} \boldsymbol{y}_{\mathrm{t}} &= \boldsymbol{h}_{\mathrm{u}} + \boldsymbol{n}_{\mathrm{t}} && (4)\\ &= (\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})\,\boldsymbol{v}_{\mathrm{w}} + \boldsymbol{n}_{\mathrm{t}} && (5)\\ &= \left[\boldsymbol{R}_{\mathrm{b}}\boldsymbol{T}_{\mathrm{b}}^{-1} \otimes \boldsymbol{T}_{\mathrm{u}}\boldsymbol{R}_{\mathrm{u}}^{-1}\right]\boldsymbol{h}_{\mathrm{d}} + \boldsymbol{n}_{\mathrm{t}}, && (6) \end{aligned}$$

where $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$ is used, and Eq. (5) follows from (1) and Eq. (6) is obtained by substituting $\boldsymbol{H}_{\mathrm{u}} = \boldsymbol{R}_{\mathrm{b}}\boldsymbol{T}_{\mathrm{b}}^{-1}\boldsymbol{H}_{\mathrm{d}}^{\mathrm{T}}\boldsymbol{R}_{\mathrm{u}}^{-1}\boldsymbol{T}_{\mathrm{u}}$ into (3). Since we do not consider the condition that some of the RF circuits are out of work, i.e., some entries of the RF circuit gain matrix are zeroes, we assume that the RF circuit gains matrices are invertible. We also assume $\boldsymbol{v}_{\mathrm{w}}$ is modeled as a Gaussian vector following $\boldsymbol{v}_{\mathrm{w}} \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{vv}}\right)$, and $\boldsymbol{C}_{\boldsymbol{vv}}$ is perfectly known at the BS [26].

From the received training signal, there exist several different approaches to derive the DL channel estimate $\widehat{\boldsymbol{H}}_{\mathrm{d}}$: we can derive the DL channel estimate directly from (6), or obtain intermediate variable estimates first, for examples, UL channel estimate from (4), wireless channel estimate from (5), then derive the DL channel estimation based on their relations with the DL channel.

By applying the MMSE estimation [27], different approaches to deriving the DL channel information from the linear Gaussian model (4)–(6) are summarized as follows.

Approach 1. The UL channel estimate $\widehat{\boldsymbol{h}}_{\mathrm{u}}$ is obtained from (4) by applying MMSE estimation, then $\widehat{\boldsymbol{H}}_{\mathrm{d}}$ is derived based on the relationship $\widehat{\boldsymbol{h}}_{\mathrm{d1}} = \left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{R}_{\mathrm{b}}^{-1} \otimes \boldsymbol{R}_{\mathrm{u}}\boldsymbol{T}_{\mathrm{u}}^{-1}\right)\widehat{\boldsymbol{h}}_{\mathrm{u}}$.

Approach 2. The UL wireless channel estimate $\widehat{\boldsymbol{v}}_{\mathrm{w}}$ is obtained from (5) by applying MMSE estimation, then $\widehat{\boldsymbol{H}}_{\mathrm{d}}$ is derived based on the relationship $\widehat{\boldsymbol{h}}_{\mathrm{d2}} = (\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})\,\widehat{\boldsymbol{v}}_{\mathrm{w}}$.

Approach 3. The DL channel estimate $\widehat{\boldsymbol{h}}_{\mathrm{d3}}$ is derived from (6) directly by applying MMSE estimation.

Proposition 1. The DL channel estimates $\widehat{\boldsymbol{h}}_{\mathrm{d1}}$, $\widehat{\boldsymbol{h}}_{\mathrm{d2}}$ and $\widehat{\boldsymbol{h}}_{\mathrm{d3}}$ obtained via the three approaches mentioned above are the same and given by

$$\widehat{\boldsymbol{h}}_{\mathrm{d}} = (\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})^{\mathrm{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\mathrm{t}}\boldsymbol{y}_{\mathrm{t}}}^{-1}\boldsymbol{y}_{\mathrm{t}}, \tag{7}$$

where $\boldsymbol{C}_{\boldsymbol{y}_{\mathrm{t}}\boldsymbol{y}_{\mathrm{t}}} = (\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})^{\mathrm{H}} + \frac{1}{\rho_{\mathrm{t}}}\boldsymbol{I}$. The covariance matrix of the estimation error $\boldsymbol{e}_{\mathrm{d}} = \boldsymbol{h}_{\mathrm{d}} - \widehat{\boldsymbol{h}}_{\mathrm{d}}$ is given by

$$\boldsymbol{C}_{\boldsymbol{ee}} = \boldsymbol{C}_{\boldsymbol{h}_{\mathrm{d}}\boldsymbol{h}_{\mathrm{d}}} - (\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})^{\mathrm{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\mathrm{t}}\boldsymbol{y}_{\mathrm{t}}}^{-1}\,(\boldsymbol{R}_{\mathrm{b}} \otimes \boldsymbol{T}_{\mathrm{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})^{\mathrm{H}}, \tag{8}$$

where $\boldsymbol{C}_{\boldsymbol{h}_{\mathrm{d}}\boldsymbol{h}_{\mathrm{d}}} = (\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\mathrm{b}} \otimes \boldsymbol{R}_{\mathrm{u}})^{\mathrm{H}}$.

Proof. See Appendix A.

Remark 1. Since Approach 1 can be considered as applying the Full-Pre-precoding Calibration scheme [8] to the UL channel estimation only with relative calibration matrices, and we can conclude that the DL channel information obtained in this way is optimal for MMSE criterion. Several methods to obtain the relative calibration matrices are provided in [11–13, 15–18].

Remark 2. $\boldsymbol{R}_{\mathrm{b}}$, $\boldsymbol{T}_{\mathrm{u}}$, $\boldsymbol{R}_{\mathrm{u}}$ and $\boldsymbol{T}_{\mathrm{b}}$ are assumed to be diagonal, which is due to the system structure considered in this paper. However, it should be noted that Proposition 1 can be extended to a general case that $\boldsymbol{R}_{\mathrm{b}}$, $\boldsymbol{T}_{\mathrm{u}}$, $\boldsymbol{R}_{\mathrm{u}}$ and $\boldsymbol{T}_{\mathrm{b}}$ are arbitrary invertible matrices, because the diagonal property is not necessary in the proof.

The channel in [1] is assumed to be as independent identically distributed, but in practice this is usually not true [28–31]. The DL channel of UT k, written as $\boldsymbol{h}_k = \boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k}$, should be modeled as correlated, in which the corresponding wireless channel $\boldsymbol{v}_k$ is modeled as

$$\boldsymbol{v}_k = \sqrt{N}\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k, \tag{9}$$

where $\boldsymbol{\Phi}_k$ is the UT k channel correlation matrix, $\boldsymbol{z}_k$ with i.i.d. entries satisfies $\boldsymbol{z}_k \sim \mathcal{CN}\left(\boldsymbol{0}, \frac{1}{N}\boldsymbol{I}_N\right)$. We assume that the channels of different UTs are uncorrelated, and denote the wireless channel correlation matrix can be rewritten as $\boldsymbol{C}_{\boldsymbol{vv}} = \mathrm{diag}\left\{\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_K\right\}$.

Based on (8), the DL channel estimation error of UT k can be modeled as

$$\boldsymbol{e}_k = \sqrt{N}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Phi}_{\boldsymbol{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k}r_{\mathrm{u}k}, \tag{10}$$

where $\boldsymbol{e}_{\mathrm{v}k} \sim \mathcal{CN}\left(\boldsymbol{0}, \frac{1}{N}\boldsymbol{I}_N\right)$ is independent of $\boldsymbol{z}_k$, $\boldsymbol{\Phi}_{\boldsymbol{e}_k}$ can be derived from (8).

By defining the DL channel estimation error matrix as $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K]^{\mathrm{T}}$, the DL channel matrix can be expressed as

$$\boldsymbol{H}_{\mathrm{d}} = \widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{E} = \boldsymbol{R}_{\mathrm{u}}\left(\widehat{\boldsymbol{V}}_{\mathrm{d}} + \boldsymbol{E}_{\mathrm{v}}\right)\boldsymbol{T}_{\mathrm{b}}, \tag{11}$$

where $\widehat{\boldsymbol{H}}_{\mathrm{d}} = \boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{V}}_{\mathrm{d}}\boldsymbol{R}_{\mathrm{u}}$ is defined as the DL channel matrix estimate based on Proposition 1. $\widehat{\boldsymbol{V}}_{\mathrm{d}}$ is the wireless channel matrix estimate, and $\boldsymbol{E}_{\mathrm{v}} = [\sqrt{N}\boldsymbol{\Phi}_{\boldsymbol{e}_1}^{1/2}\boldsymbol{e}_{\mathrm{v}1}, \ldots, \sqrt{N}\boldsymbol{\Phi}_{\boldsymbol{e}_K}^{1/2}\boldsymbol{e}_{\mathrm{v}K}]^{\mathrm{T}}$ is the corresponding estimation error matrix.

3 Robust MMSE precoding

During the DL data transmission phase, the signal received at the UTs at a specific time instant can be written as

$$\boldsymbol{y}_{\mathrm{d}} = \boldsymbol{H}_{\mathrm{d}}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}_{\mathrm{d}} = \boldsymbol{R}_{\mathrm{u}}\left(\widehat{\boldsymbol{V}}_{\mathrm{d}} + \boldsymbol{E}_{\mathrm{v}}\right)\boldsymbol{T}_{\mathrm{b}}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}_{\mathrm{d}}, \tag{12}$$

where $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ denotes the Gaussian signaling, i.e., $\boldsymbol{s} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K)$, and the k-th entry s_k is the signal for UT k, $\boldsymbol{n}_{\mathrm{d}} \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma_{\mathrm{d}}^2\boldsymbol{I}_K\right)$ is the independent Gaussian noise, and $\boldsymbol{P} \in \mathbb{C}^{N\times K}$ is the precoding matrix which satisfies the power constraint

$$\mathrm{tr}\left\{\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right\} \leqslant P. \tag{13}$$

Given the DL channel estimate $\widehat{\boldsymbol{H}}_{\mathrm{d}} = \boldsymbol{R}_{\mathrm{u}}\widehat{\boldsymbol{V}}_{\mathrm{d}}\boldsymbol{T}_{\mathrm{b}}$, we can design a precoding matrix to minimize the MSE of signal detection at the UTs. Therefore, the optimization problem with power constraint can be formulated as

$$\min_{\boldsymbol{P},\alpha} \ \mathrm{E}\left\{\|\alpha\boldsymbol{y}_{\mathrm{d}} - \boldsymbol{s}\|_2^2\right\} \quad \text{s.t.} \ \mathrm{tr}\left\{\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right\} \leqslant P, \tag{14}$$

where α is a real optimization variable corresponding to the adaptive gain control, and the expectation is with respect to $\boldsymbol{s}$, $\boldsymbol{n}_{\mathrm{d}}$ and $\boldsymbol{E}_{\mathrm{v}}$. The solution is given in the following proposition.

Proposition 2. The optimal solution to the problem (14) is given by

$$\begin{aligned} \boldsymbol{P}_{\mathrm{opt}} &= \frac{1}{\alpha_{\mathrm{opt}}}\left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}\widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}}\boldsymbol{I}_N\right)^{-1}\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}, \\ \alpha_{\mathrm{opt}} &= \sqrt{\frac{\mathrm{tr}\left\{\widehat{\boldsymbol{H}}_{\mathrm{d}}\left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}\widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}}\boldsymbol{I}_N\right)^{-2}\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}\right\}}{P}}, \end{aligned} \tag{15}$$

where $\boldsymbol{C}_{\mathrm{E}} = \mathrm{E}\left\{(\boldsymbol{R}_{\mathrm{u}}\boldsymbol{E}_{\mathrm{v}}\boldsymbol{T}_{\mathrm{b}})^{\mathrm{H}}(\boldsymbol{R}_{\mathrm{u}}\boldsymbol{E}_{\mathrm{v}}\boldsymbol{T}_{\mathrm{b}})\right\}$, and the downlink SNR is defined as $\rho_{\mathrm{d}} = \frac{P}{\sigma_{\mathrm{d}}^2}$.

Proof. See Appendix B.

In our robust MMSE precoding (15), the channel estimation error is taken into account. While the conventional MMSE based precoding schemes, such as RZF [20] and transmit Wiener filter [21], usually assume full CSI at the BS. It can be seen from the closed-form expression of the proposed precoding that the computational complexity is $\mathcal{O}\left(N^3\right)$, which is acceptable in practice [24]. Note that, when $\boldsymbol{C}_{\mathrm{E}} \to \boldsymbol{0}$, the robust MMSE precoding reduces to the conventional RZF precoding.

4 Sum rate of robust MMSE precoding

For the DL data transmission with precoding, the signal received by UT k can be written as

$$y_k = \boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{p}_k s_k + \sum_{j=1,j\neq k}^{K}\boldsymbol{h}_k^{\mathrm{T}}\boldsymbol{p}_j s_j + n_k, \tag{16}$$

where $\boldsymbol{p}_k$ is the precoding vector for UT k.

The signal-to-interference plus noise ratio (SINR) of user k under the robust MMSE precoding given by Proposition 2 takes the form [32]

$$\gamma_k = \frac{\left|(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}}\widehat{\boldsymbol{W}}(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^*\right|^2}{\sum_{j=1,j\neq k}^{K}\left|(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}}\widehat{\boldsymbol{W}}(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j})^*\right|^2 + \frac{\xi}{\rho_{\mathrm{d}}}}, \tag{17}$$

where $\widehat{\boldsymbol{W}} = (\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}\widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}}\boldsymbol{I}_N)^{-1}$ and $\xi = \mathrm{tr}\{\widehat{\boldsymbol{H}}_{\mathrm{d}}\boldsymbol{W}^2\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}\}$. The UT k channel estimate $\widehat{\boldsymbol{h}}_k$ can be written as $\widehat{\boldsymbol{h}}_k = \boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k}$, in which $\widehat{\boldsymbol{v}}_k = \sqrt{N}(\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k - \boldsymbol{\Phi}_{\boldsymbol{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k})$ according to $\boldsymbol{h}_k = \widehat{\boldsymbol{h}}_k + \boldsymbol{e}_k$.

The rate R_k is given by

$$R_k = \log\left(1 + \gamma_k\right), \tag{18}$$

and the ergodic sum rate $R_{\rm sum}$ is given by

$$R_{\rm sum} = \sum_{k=1}^{K} \mathrm{E}\left\{R_k\right\}. \tag{19}$$

Under the assumption that all correlation matrices $\boldsymbol{\Phi}_k$ have uniformly bounded spectral norm on N, i.e., $\limsup_{N,K\to\infty} \sup_{1\leqslant k\leqslant K} \|\boldsymbol{\Phi}_k\| < \infty$, a deterministic equivalent γ_k° of γ_k is provided in the following proposition.

Proposition 3. When $N \to \infty$ and $1 \leqslant N/K < c < \infty$ for a constant c, we have that $\gamma_k - \gamma_k^\circ \to 0$ almost surely, where γ_k° is given by

$$\gamma_k^\circ = \frac{\nu_k^\circ}{\sum_{j=1,j\neq k}^{K} \omega_j^\circ + \frac{\xi^\circ}{\rho}} \tag{20}$$

with

$$\nu_k^\circ = \frac{\left|r_{\mathrm{u}k}\right|^2 m_k^\circ}{1 + \left|r_{\mathrm{u}k}\right|^2 m_{\mathrm{eq}_k}^\circ}, \tag{21}$$

$$\xi^\circ = \frac{1}{N}\sum_{k=1}^{K} \frac{\left|r_{\mathrm{u}k}\right|^2 \mu_{\mathrm{I}k}^\circ}{\left(1 + \left|r_{\mathrm{u}k}\right|^2 m_{\mathrm{eq}_k}^\circ\right)^2}, \tag{22}$$

$$\omega_j^\circ = \frac{\left|r_{\mathrm{u}j}\right|^2 \left|r_{\mathrm{u}k}\right|^2}{N\left(1 + \left|r_{\mathrm{u}j}\right|^2 m_{\mathrm{eq}_j}^\circ\right)^2} \left[\mu_{\mathrm{T}j}^\circ + \frac{\left|r_{\mathrm{u}k}\right|^4 \left|m_k^\circ\right|^2 \mu_{\mathrm{Te}j}^\circ}{\left(1 + \left|r_{\mathrm{u}k}\right|^2 m_{\mathrm{eq}_k}^\circ\right)^2} - \frac{2\left|r_{\mathrm{u}k}\right|^2 \Re\left\{m_k^\circ \mu_{\mathrm{T}j}^\circ\right\}}{\left(1 + \left|r_{\mathrm{u}k}\right|^2 m_{\mathrm{eq}_k}^\circ\right)}\right], \tag{23}$$

in which

$$m_k^\circ = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_k^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \boldsymbol{\Psi}\right\}, \tag{24}$$

$$m_{\mathrm{eq}_k}^\circ = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \boldsymbol{\Psi}\right\}, \tag{25}$$

where $\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}} = \boldsymbol{\Phi}_k^{\mathrm{T}} + \boldsymbol{\Phi}_{\mathrm{e}_k}^{\mathrm{T}}$, and $\boldsymbol{\Psi}$ is defined as

$$\boldsymbol{\Psi} = \left(\frac{1}{N}\sum_{j=1}^{K} \frac{\left|r_{\mathrm{u}k}\right|^2 \boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}}}{1 + e_j} + \frac{1}{N}\boldsymbol{C}_{\mathrm{E}} + \frac{K}{N\rho_{\mathrm{d}}}\boldsymbol{I}_N\right)^{-1}, \tag{26}$$

and the elements of $\boldsymbol{e} = [e_1, \ldots, e_K]$ are defined as $e_k = \lim_{t\to\infty} e_k^{(t)}$, where for $t = 1, 2, \ldots$.

$$e_k^{(t)} = \frac{1}{N}\mathrm{tr}\left\{\left|r_{\mathrm{u}k}\right|^2 \boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \left(\frac{1}{N}\sum_{j=1}^{K} \frac{\left|r_{\mathrm{u}k}\right|^2 \boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}}}{1 + e_j^{(t-1)}} + \frac{1}{N}\boldsymbol{C}_{\mathrm{E}} + \frac{K}{N\rho_{\mathrm{d}}}\boldsymbol{I}_N\right)^{-1}\right\} \tag{27}$$

with initial values $e_k^{(0)} = 1/\rho_d$ for all k. Moreover, $\mu_{\mathrm{T}j}^\circ$, $\mu_{\mathrm{Te}j}^\circ$ and $\mu_{\mathrm{I}j}^\circ$ are given by

$$\mu_{\mathrm{T}j}^\circ = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_j^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \boldsymbol{\Psi}_{\mathrm{T}}'\right\}, \tag{28}$$

$$\mu_{\mathrm{Te}j}^\circ = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \boldsymbol{\Psi}_{\mathrm{T}}'\right\}, \tag{29}$$

$$\mu_{\mathrm{I}k}^\circ = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \boldsymbol{\Psi}_{\mathrm{I}}'\right\}, \tag{30}$$

where $\boldsymbol{\Psi}'_{\mathrm{T}} = \boldsymbol{\Psi}'(\boldsymbol{D})$ for $\boldsymbol{D} = \boldsymbol{T}_{\mathrm{b}}^{*}\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}$, and $\boldsymbol{\Psi}'_{\mathrm{I}} = \boldsymbol{\Psi}'(\boldsymbol{D})$ for $\boldsymbol{D} = \boldsymbol{I}$, in which the function $\boldsymbol{\Psi}'(\boldsymbol{D})$ is defined as

$$\boldsymbol{\Psi}'(\boldsymbol{D}) = \boldsymbol{\Psi}\left[\frac{1}{N}\sum_{j=1}^{K}\frac{\boldsymbol{T}_{\mathrm{b}}^{*}\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}e'_j(\boldsymbol{D})}{(1+e_j)^2} + \boldsymbol{D}\right]\boldsymbol{\Psi}, \tag{31}$$

and $\boldsymbol{e}'(\boldsymbol{D}) = [e'_1(\boldsymbol{D})\cdots e'_K(\boldsymbol{D})]^{\mathrm{T}}$ is calculated as

$$\boldsymbol{e}'(\boldsymbol{D}) = [\boldsymbol{I}_n - \boldsymbol{J}]^{-1}\boldsymbol{q}(\boldsymbol{D}), \tag{32}$$

where $\boldsymbol{J}$ and $\boldsymbol{q}(\boldsymbol{D})$ are defined as

$$[\boldsymbol{J}]_{i,j} = \frac{\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^{*}\boldsymbol{\Phi}_{\mathrm{eq}_i}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\boldsymbol{T}_{\mathrm{b}}^{*}\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}}{N(1+e_j)^2}, \qquad 1 \leqslant i,j \leqslant K, \tag{33}$$

$$[\boldsymbol{q}(\boldsymbol{D})]_i = \frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^{*}\boldsymbol{\Phi}_{\mathrm{eq}_i}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\boldsymbol{D}\boldsymbol{\Psi}\right\}, \qquad 1 \leqslant i \leqslant K. \tag{34}$$

Proof. See Appendix C.

Remark 3. The fixed-point algorithm in Proposition 3 is proved to converge in Proposition 1 of [32,33].

We are interested in the individual rate R_k as well as the ergodic system sum rate R_{sum}. Since the logarithm is a continuous function, using the continuous mapping theorem [34, 35], we can obtain an asymptotic approximation of the individual rate, given by

$$R_k - R_k^{\circ} \xrightarrow[N\to\infty]{\text{a.s.}} 0, \tag{35}$$

where $R_k^{\circ} = \log(1+\gamma_k^{\circ})$. An asymptotic approximation R_{sum}° of the ergodic sum rate R_{sum} is obtained by replacing the instantaneous individual rate R_k with its asymptotic approximation R_k°, i.e.,

$$R_{\mathrm{sum}}^{\circ} = \sum_{k=1}^{K}\log(1+\gamma_k^{\circ}). \tag{36}$$

5 Simulation results

In this section, we present some simulation results to evaluate the performance of the proposed robust MMSE based precoding and validate the accuracy of the ergodic sum rate approximation given by (36).

Similar to [8], the amplitudes of the RF gains are assumed to be identically and uniformly distributed, i.e., $\mathcal{U}(1-0.5\delta, 1+0.5\delta)$, where δ reflects the distribution range. We also assume that the phase of the RF gains is well compensated in the simulation. In the channel model of [32], the parameter τ is used to reflect the accuracy of DL channel information ($\tau = 0$ corresponds to the perfect CSI case, whereas $\tau = 1$ indicates that the DL channel information is completely uncorrelated to the true channel), while we use the parameter ρ_{t} to represent the channel estimation accuracy. Comparing (8) and (6) of [32], we can see that $\rho_{\mathrm{t}} = 5$ dB is approximately equal to $\tau = 0.1$ which means that the DL channel information is highly accurate, and $\rho_{\mathrm{t}} = -5$ dB is approximately equal to $\tau = 0.5$ which means that the estimation error is large. The correlation $\boldsymbol{\Phi}_k$ of the kth UT is modeled as [32, 36]

$$[\boldsymbol{\Phi}_k]_{ij} = \frac{1}{\phi_{k,\max} - \phi_{k,\min}}\int_{\phi_{k,\min}}^{\phi_{k,\max}} \mathrm{e}^{\mathrm{i}\frac{2\pi}{\lambda}d_{ij}\cos(\phi)}\mathrm{d}\phi, \tag{37}$$

where λ denotes the signal wavelength, and d_{ij} denotes the distance between transmit antennas i and j. We assume that the UTs distribute uniformly around the BS at an angle $\phi_k = 2\pi k/K$, and the azimuth angle at which the waves impinge the UT k uniformly ranging from $\phi_{k,\min}$ to $\phi_{k,\max}$. In our example, we choose $\phi_{k,\min} = -\pi$ and $\phi_{k,\max} = \phi_k - \pi$. Note that for some UT, whose azimuth angle range is small (i.e., $\phi_{k,\max} - \phi_{k,\min}$ is small), the correlation matrix is rank-deficient, since the signal arrives from a very narrow angle. To ensure that $\|\boldsymbol{\Phi}_k\|$ is bounded as N grows large, we assume that the distance between adjacent antennas is half wavelength, for the BS equipped with a uniform linear array. In all the simulations, we take equal power allocation.

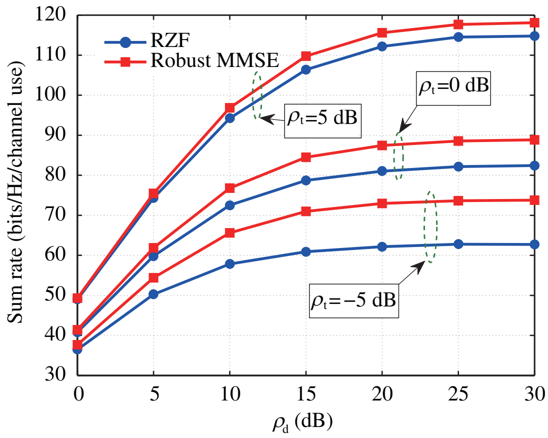

Figure 1 (Color online) Comparison of the ergodic sum rates between the robust MMSE precoding and the RZF precoding, results are shown versus the SNR with $N = 128$, $K = 32$ and $\delta = 0.4$.

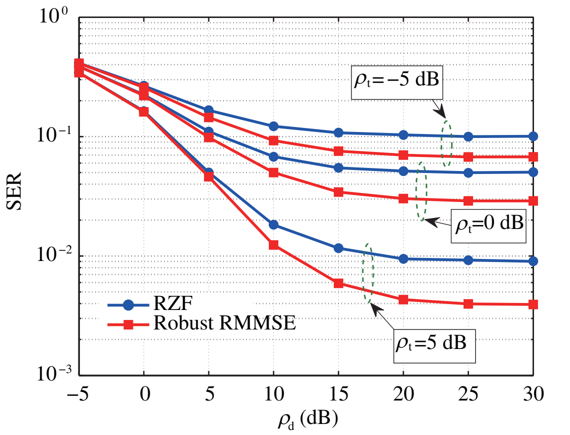

Figure 2 (Color online) Comparison of the SER performances between the robust MMSE precoding and the RZF precoding under QPSK modulation, results are shown versus SNR with $N = 128$, $K = 32$ and $\delta = 0.4$.

5.1 Performance of the proposed robust MMSE precoding

Figure 1 presents a comparison of the ergodic sum rates for robust MMSE precoding and the conventional RZF precoding in terms of the downlink SNR $\rho_{\rm d}$ with $N = 128$, $K = 32$ and $\delta = 0.4$. The ergodic sum rates are obtained by Monte Carlo simulations. As we can see from the figure, for all downlink SNR values, the robust MMSE precoding outperforms the conventional RZF precoding. In the low $\rho_{\rm d}$ regime, where the noise dominates, the sum rate of the robust MMSE precoding is similar to the conventional RZF precoding. However, in the high $\rho_{\rm d}$ regime, where the interference dominates, the improvement is significant, because the channel estimation error is taken account. Moreover, it can be observed that the lager is the channel estimation error, the more significant will be the improvement.

Figure 2 compares the symbol error rate (SER) performances between the robust MMSE precoding and the conventional RZF precoding under the same condition as Figure 1. We assume a quadrature phase-shift-keying (QPSK) modulation. From the figure we observe that the SER achieved by the robust MMSE precoding is lower than the conventional RZF precoding, which means that the robust transmission is more reliable than the conventional RZF.

5.2 Accuracy of the asymptotic approximation

Figure 3 depicts the absolute error of the asymptotic approximation $R_{\rm sum}^{\circ}$ compared to the ergodic sum rate $R_{\rm sum}$, obtained by Monte Carlo simulations for fixed downlink SNR $\rho_{\rm d} = 10$ dB with $N/K = 4$. The result is versus N. We can observe that for different channel estimation errors and hardware mismatch distribution range, the absolute error is small, and becomes smaller with increasing N. Furthermore, it can be seen that for small dimensions, for example $N = 8$, the asymptotic approximation is also accurate.

The ergodic sum rates obtained by Monte Carlo simulations and the asymptotic approximation given by (36) are compared in Figure 4, as functions of SNR with $N = 128$ and $K = 32$. The error bars indicate the standard deviation of the Monte Carlo simulations. From Figure 4, we can see that for different channel estimation errors and hardware mismatch distribution range, the asymptotic approximation lies within one standard deviation of the Monte Carlo simulations, which verifies that the asymptotic approximation is accurate.

6 Conclusion

In this paper, we proposed an MMSE based precoding which is robust to the channel estimation error, in a single cell massive MIMO system with non-reciprocal RF circuits at both the BS and UTs antennas.

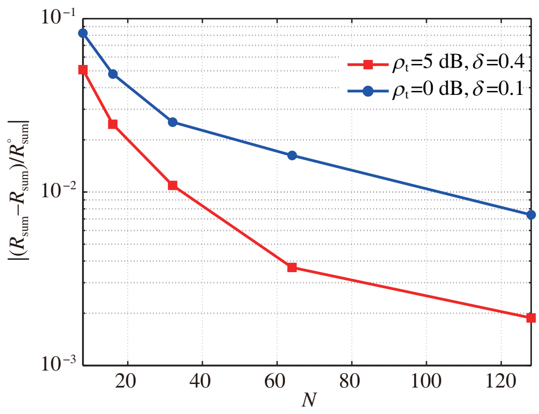

Figure 3 (Color online) Absolute error $|(R_{\rm sum} - R^{\circ}_{\rm sum})/R_{\rm sum}|$ for fixed SNR $\rho_{\rm d} = 10$ dB, result is versus N with $N/K = 4$.

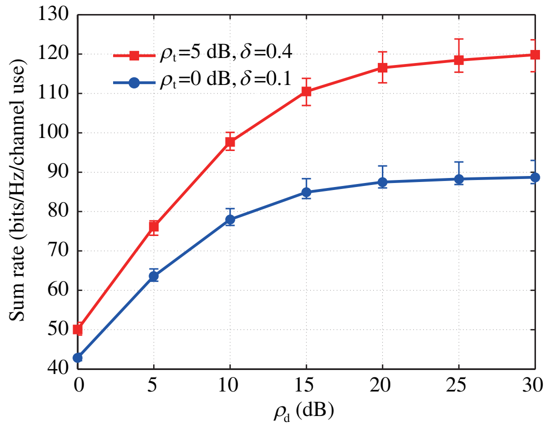

Figure 4 (Color online) Comparison of the ergodic sum rates obtained by Monte Carlo simulations and the asymptotic approximation, results are shown versus the SNR with $N = 128$ and $K = 32$.

Taking the RF circuit gains into account, there are three different approaches for a BS to obtain the DL channel estimate based on the MMSE based estimation theory with hardware mismatch parameters. We show that the estimation results of the three different approaches are the same. Compared to the conventional RZF precoding, the proposed robust MMSE precoding accounts for the channel estimation error, and can achieve an improvement in the sum rate of the system and a lower SER performance for all the SNR values, especially in high SNR regime. Moreover, we derive an asymptotic approximation of the ergodic sum rate of the proposed robust MMSE precoding. The simulation results show that the asymptotic approximation is accurate, even for a small system dimension.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61320106003, 61471113, 61521061, 61631018), National High Technology Research and Development Program of China (863) (Grant Nos. 2015AA01A701, 2014AA01A704), National Science and Technology Major Project of China (Grant No. 2014ZX03003006-003), and Huawei Cooperation Project.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Marzetta T L. Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun, 2010, 9: 3590–3600
2 Rusek F, Persson D, Lau B K, et al. Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process Mag, 2013, 30: 40–60
3 Larsson E G, Edfors O, Tufvesson F, et al. Massive MIMO for next generation wireless systems. IEEE Commun Mag, 2014, 2: 186–195
4 Gao X, Edfors O, Fredrik R, et al. Linear pre-coding performance in measured very-large MIMO channels. In: Proceedings of IEEE Vehicular Technology Conference (VTC Fall), San Francisco, 2011. 1–5
5 Marzetta T L. How much training is required for multiuser MIMO. In: Proceedings of the 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2006. 359–363
6 Wang D M, Zhang Y, Wei H, et al. An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications. Sci China Inf Sci, 2016, 59: 081301
7 Schenk T. RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation. Berlin: Springer, 2008
8 Zhang W C, Ren H, Pan C H, et al. Large-scale antenna systems with UL/DL hardware mismatch: achievable rates analysis and calibration. IEEE Trans Commun, 2015, 63: 1216–1229
9 Björnson E, Hoydis J, Kountouris M, et al. Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits. IEEE Trans Inform Theory, 2014, 60: 7112–7139
10 Bourdoux A. Non-reciprocal transceivers in OFDM/SDMA systems: impact and mitigation. In: Proceedings of Radio and Wireless Conference, Boston, 2003. 183–186
11 Guillaud M, Slock D T M, Knopp R. A practical method for wireless channel reciprocity exploitation through relative calibration. In: Proceedings of the 8th International Symposium on Signal Processing and Its Applications, Sydney,

2005. 403–406
12 Shi J, Luo Q L, You L M. An efficient method for enhancing TDD over the air reciprocity calibration. In: Proceedings of IEEE Wireless Communications and Networking Conference, Cancun, 2011. 339–344
13 Han S Q, Yang C Y, Wang G, et al. Coordinated multipoint transmission strategies for TDD systems with non-ideal channel reciprocity. IEEE Trans Commun, 2012, 10: 4256–4270
14 Guillaud M, Kaltenberger F. Towards practical channel reciprocity exploitation: relative calibration in the presence of frequency offset. In: Proceedings of IEEE Wireless Communications and Networking Conference, Shanghai, 2013. 2525–2530
15 Rogalin R, Bursalioglu O Y, Papadopoulos H, et al. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. IEEE Trans Wirel Commun, 2014, 4: 1815–1831
16 Su L Y, Yang C Y, Wang G, et al. Retrieving channel reciprocity for coordinated multi-point transmission with joint processing. IEEE Trans Commun, 2014, 5: 1541–1553
17 Shepard C, Yu H, Anand N, et al. Argos: practical many-antenna base stations. In: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, Istanburl, 2012. 53–64
18 Kaltenberger F. Relative channel reciprocity calibration in MIMO/TDD systems. In: Proceedings of IEEE Future Network and Mobile Summit, Florence, 2010. 1–10
19 Yang H, Marzetta T. Performance of conjugate and zero-forcing beamforming in large-scale antenna systems. IEEE J Sel Area Commun, 2013, 31: 172–179
20 Peel C B, Hochwald B M, Swindlehurst A L. A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. IEEE Trans Commun, 2005, 53: 195–202
21 Joham M, Utschich W, Nossek J A. Linear transmit processing in MIMO communications systems. IEEE Trans Signal Process, 2005, 53: 2700–2712
22 Costa M H M. Writting on dirty paper. IEEE Trans Inform Theory, 1983, 29: 439–441
23 Hochwald B M, Peel C B, Swindlehurst A L. Vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation. IEEE Trans Commun, 2005, 53: 537–544
24 Björnson E, Larsson E G, Marzetta T L. Massive MIMO: ten myths and one critical question. IEEE Commun Mag, 2016, 54: 114–123
25 Wei H, Wang D M, Wang J Z, et al. Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding. Sci China Inf Sci, 2016, 59: 022302
26 Meng X, Jiang B, Gao X Q. Efficient co-channel interference suppression in MIMO-OFDM systems. Sci China Inf Sci, 2015, 58: 022301
27 Kay S M. Fundamentals of Statistical Signal Processing: Estimation Theory. Englewood Cliffs: Prentice Hall, 1993
28 You L, Gao X Q, Xia X-G, et al. Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels. IEEE Trans Wireless Commun, 2015, 14: 3352–3366
29 Kermoal J P, Schumacher L, Pedersen I K, et al. A stochastic MIMO radio channel model with experimental validation. IEEE Trans Commun, 2002, 20: 1211–1226
30 Sayeed A M. Deconstructing multiantenna fading channels. IEEE Trans Signal Process, 2002, 50: 2563–2579
31 Weichselberger W, Herdin M, Özcelik H, et al. A stochastic MIMO channel model with joint correlation of both link ends. IEEE Trans Wirel Commun, 2006, 5: 90–100
32 Wagner S, Couillet R, Debbah M, et al. Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback. IEEE Trans Inform Theory, 2012, 58: 4509–4537
33 Hoydis J, Brink S T, Debbah M. Massive MIMO in the UL/DL of celluar networks: how many antennas do we need? IEEE J Sel Areas Commun, 2013, 31: 160–171
34 Couillet R, Debbah M. Random Matrix Methods for Wireless Communications. Cambridge: Cambridge University Press, 2011
35 Silverstein J W, Bai Z D. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. J Multivar Anal, 1995, 54: 175–192
36 Jacks W C, Cox D C. Microwave Mobile Communications. New York: Wiley, 1994

Appendix A Proof of Proposition 1

Firstly, we obtain the DL channel estimate via these three approaches in turn. For simplicity, we treat RF circuit gains as known constants at the BS during the channel estimation phase and the concerned data transmission interval [8].

Approach 1. Base on (4), the MMSE estimate of $\boldsymbol{h}_{\rm u}$ is given by [27]

$$\widehat{\boldsymbol{h}}_u = \boldsymbol{C}_{\boldsymbol{h}_{\rm u}\boldsymbol{y}_{\rm t}} \boldsymbol{C}_{\boldsymbol{y}_{\rm t}\boldsymbol{y}_{\rm t}}^{-1} \boldsymbol{y}_{\rm t} = (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u}) \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{v}} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u})^{\rm H} \boldsymbol{C}_{\boldsymbol{y}_{\rm t}\boldsymbol{y}_{\rm t}}^{-1} \boldsymbol{y}_{\rm t}. \tag{A1}$$

According to the orthogonality principle of MMSE estimation [27], the estimation error $\boldsymbol{e}_{\rm u} = \boldsymbol{h}_{\rm u} - \widehat{\boldsymbol{h}}_{\rm u}$ is uncorrelated with $\boldsymbol{h}_{\rm u}$. Furthermore, the covariance matrix of the estimation error is given by

$$\boldsymbol{C}_{\boldsymbol{e}_{\rm u}\boldsymbol{e}_{\rm u}} = (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u}) \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{v}} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u})^{\rm H} - (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u}) \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{v}} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u})^{\rm H} \boldsymbol{C}_{\boldsymbol{y}_{\rm t}\boldsymbol{y}_{\rm t}}^{-1} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u}) \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{v}} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u})^{\rm H}. \tag{A2}$$

Then we obtain the DL channel estimate $\widehat{\boldsymbol{h}}_{\rm d1}$:

$$\begin{aligned} \widehat{\boldsymbol{h}}_{\rm d1} &= \left(\boldsymbol{T}_{\rm b}\boldsymbol{R}_{\rm b}^{-1} \otimes \boldsymbol{R}_{\rm u}\boldsymbol{T}_{\rm u}^{-1}\right) \widehat{\boldsymbol{h}}_{\rm u} \\ &= (\boldsymbol{T}_{\rm b} \otimes \boldsymbol{R}_{\rm u}) \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{v}} (\boldsymbol{R}_{\rm b} \otimes \boldsymbol{T}_{\rm u})^{\rm H} \boldsymbol{C}_{\boldsymbol{y}_{\rm t}\boldsymbol{y}_{\rm t}}^{-1} \boldsymbol{y}_{\rm t}. \end{aligned} \tag{A3}$$

The estimation error $\boldsymbol{e}_{\text{d}1}$, which is expressed as $\boldsymbol{e}_{\text{d}1} = (\boldsymbol{T}_{\text{b}}\boldsymbol{R}_{\text{b}}^{-1} \otimes \boldsymbol{R}_{\text{u}}\boldsymbol{T}_{\text{u}}^{-1})\boldsymbol{e}_{\text{u}}$, is uncorrelated with $\boldsymbol{h}_{\text{d}}$ since $\boldsymbol{e}_{\text{u}}$ is independent on $\boldsymbol{h}_{\text{u}}$. The corresponding estimation error covariance matrix is given by

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{e}_1\boldsymbol{e}_1} &= \left(\boldsymbol{T}_{\text{b}}\boldsymbol{R}_{\text{b}}^{-1} \otimes \boldsymbol{R}_{\text{u}}\boldsymbol{T}_{\text{u}}^{-1}\right)\boldsymbol{C}_{\boldsymbol{e}_{\text{u}}\boldsymbol{e}_{\text{u}}}\left(\boldsymbol{T}_{\text{b}}\boldsymbol{R}_{\text{b}}^{-1} \otimes \boldsymbol{R}_{\text{u}}\boldsymbol{T}_{\text{u}}^{-1}\right)^{\text{H}} \\ &= (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}} - (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}}. \end{aligned} \tag{A4}$$

Approach 2. Base on (5), the MMSE estimate of $\boldsymbol{v}_{\text{w}}$ is given by [27]

$$\widehat{\boldsymbol{v}}_{\text{w}} = \boldsymbol{C}_{\boldsymbol{v}\boldsymbol{y}_{\text{t}}}\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\boldsymbol{y}_{\text{t}} = \boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\boldsymbol{y}_{\text{t}}. \tag{A5}$$

The covariance matrix of the estimation error $\boldsymbol{e}_{\text{w}} = \boldsymbol{v}_{\text{w}} - \widehat{\boldsymbol{v}}_{\text{w}}$ is given by

$$\boldsymbol{C}_{\boldsymbol{e}_{\text{w}}\boldsymbol{e}_{\text{w}}} = \boldsymbol{C}_{\boldsymbol{vv}} - \boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}. \tag{A6}$$

Then we obtain the DL channel estimate $\widehat{\boldsymbol{h}}_{\text{d}2}$:

$$\widehat{\boldsymbol{h}}_{\text{d}2} = (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\widehat{\boldsymbol{v}}_{\text{w}} = (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\boldsymbol{y}_{\text{t}}. \tag{A7}$$

The estimation error $\boldsymbol{e}_{\text{d}2}$, which is expressed as $\boldsymbol{e}_{\text{d}2} = (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{e}_{\text{w}}$, is uncorrelated with $\boldsymbol{h}_{\text{d}2}$ for the same reason as $\boldsymbol{e}_{\text{d}1}$. The covariance matrix is give by

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{e}_2\boldsymbol{e}_2} &= (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{e}_{\text{w}}\boldsymbol{e}_{\text{w}}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}} \\ &= (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}} - (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}}. \end{aligned} \tag{A8}$$

Approach 3. Based on (6), the MMSE estimate of $\boldsymbol{h}_{\text{d}}$ is given by [27]

$$\widehat{\boldsymbol{h}}_{\text{d}} = \boldsymbol{C}_{\boldsymbol{h}_{\text{d}}\boldsymbol{y}_{\text{t}}}\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\boldsymbol{y}_{\text{t}} = (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\boldsymbol{y}_{\text{t}}. \tag{A9}$$

The covariance matrix of the estimation error $\boldsymbol{e}_{\text{d}3} = \boldsymbol{h}_{\text{d}} - \widehat{\boldsymbol{h}}_{\text{d}}$ is given by

$$\boldsymbol{C}_{\boldsymbol{e}_3\boldsymbol{e}_3} = (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}} - (\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})^{\text{H}}\,\boldsymbol{C}_{\boldsymbol{y}_{\text{t}}\boldsymbol{y}_{\text{t}}}^{-1}\,(\boldsymbol{R}_{\text{b}} \otimes \boldsymbol{T}_{\text{u}})\,\boldsymbol{C}_{\boldsymbol{vv}}\,(\boldsymbol{T}_{\text{b}} \otimes \boldsymbol{R}_{\text{u}})^{\text{H}}. \tag{A10}$$

Comparing (A3), (A7) and (A9), it can be seen that $\widehat{\boldsymbol{h}}_{\text{d}1}$, $\widehat{\boldsymbol{h}}_{\text{d}2}$ and $\widehat{\boldsymbol{h}}_{\text{d}3}$ are the same. Moreover, we can see that the covariance of the estimation error is also the same by comparing (A4), (A8), and (A10). This concludes the proof.

Appendix B Proof of Proposition 2

First we simplify the MSE expression as

$$\begin{aligned}\epsilon &= \mathrm{E}\left\{\|\alpha\boldsymbol{y}_{\text{d}} - \boldsymbol{s}\|_2^2\right\} \\ &= \mathrm{E}\left\{\left\|\left[\left(\boldsymbol{R}_{\text{u}}\widehat{\boldsymbol{V}}_{\text{d}}\boldsymbol{T}_{\text{b}} + \boldsymbol{R}_{\text{u}}\boldsymbol{E}_{\text{v}}\boldsymbol{T}_{\text{b}}\right)\boldsymbol{P} + \boldsymbol{I}_K\right]\boldsymbol{s}\right\|_2^2\right\} + \alpha^2 K\sigma_{\text{d}}^2 \\ &= \mathrm{tr}\left\{\alpha^2\boldsymbol{P}^{\text{H}}\left(\widehat{\boldsymbol{H}}_{\text{d}}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{d}} + \boldsymbol{C}_E\right)\boldsymbol{P}\right\} - \mathrm{tr}\left\{\alpha\widehat{\boldsymbol{H}}_{\text{d}}\boldsymbol{P}\right\} - \mathrm{tr}\left\{\alpha\boldsymbol{P}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{d}}^{\text{H}}\right\} + \left(\alpha^2\sigma_{\text{d}}^2 + 1\right)K, \end{aligned} \tag{B1}$$

where $\boldsymbol{C}_{\text{E}} = \mathrm{E}\{(\boldsymbol{R}_{\text{u}}\boldsymbol{E}_{\text{v}}\boldsymbol{T}_{\text{b}})^{\text{H}}\,(\boldsymbol{R}_{\text{u}}\boldsymbol{E}_{\text{v}}\boldsymbol{T}_{\text{b}})\}$ can be obtained from (A10). Furthermore, it can be seen from Proposition 1 that the covariance of the estimation error is abe to be obtained only with the relative calibration matrices.

The objective function in (B1) is non-convex with respect to $(\boldsymbol{P}, \alpha)$, but there exist a global minimum [Lemmas 4, 8]. The global optimal solution must satisfy the Karush-Kunhn-Tucker (KKT) necessary condition[1)]. We will derive all the solutions that satisfy the KKT conditions, and then identify the optimal solution.

The Lagrangian associated with the problem (14) can be formulated as

$$\mathcal{L}\left(\boldsymbol{P}, \alpha, \lambda\right) = \epsilon + \eta\left(\mathrm{tr}\{\boldsymbol{P}\boldsymbol{P}^{\text{H}}\} - P\right), \tag{B2}$$

where η is the Lagrange multiplier associated with the inequality constraint.

The KKT conditions can be obtained as[2)]

$$\frac{\partial}{\partial\boldsymbol{P}}\mathcal{L}(\boldsymbol{P}, \alpha, \lambda) = \alpha^2\left(\widehat{\boldsymbol{H}}_{\text{d}}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{d}} + \boldsymbol{C}_{\text{E}}\right)^{\text{T}}\boldsymbol{P}^* - \alpha\left(\widehat{\boldsymbol{H}}_{\text{d}}\right)^{\text{T}} + \eta\boldsymbol{P}^* = 0, \tag{B3}$$

$$\frac{\partial}{\partial\alpha}\mathcal{L}(\boldsymbol{P}, \alpha, \lambda) = 2\alpha\mathrm{tr}\left\{\boldsymbol{P}^{\text{H}}\left(\widehat{\boldsymbol{H}}_{\text{d}}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{d}} + \boldsymbol{C}_E\right)\boldsymbol{P}\right\} - \mathrm{tr}\left\{\widehat{\boldsymbol{H}}_{\text{d}}\boldsymbol{P}\right\} - \mathrm{tr}\left\{\boldsymbol{P}^{\text{H}}\widehat{\boldsymbol{H}}_{\text{d}}^{\text{H}}\right\} + 2\alpha\sigma_{\text{d}}^2 K = 0, \tag{B4}$$

$$\eta \geqslant 0, \tag{B5}$$

$$\mathrm{tr}\left\{\boldsymbol{P}\boldsymbol{P}^{\text{H}}\right\} \leqslant P, \tag{B6}$$

$$\eta\left(\mathrm{tr}\{\boldsymbol{P}\boldsymbol{P}^{\text{H}}\} - P\right) = 0. \tag{B7}$$

---

1) Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004.

2) Hjørungnes A. Complex-valued Matrix Derivatives with Applications in Signal Processing and Communications. Cambridge: Cambridge University Press, 2011.

An obvious solution is $(\alpha = 0,\ \boldsymbol{P} = 0,\ \eta = 0)$, and the corresponding MSE ϵ equals K. In the case of $\alpha \neq 0$, from (B3) we can obtain

$$\alpha \boldsymbol{P}^{\mathrm{H}} \left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{\eta}{\alpha^2} \boldsymbol{I}_N \right) = \widehat{\boldsymbol{H}}_{\mathrm{d}}. \tag{B8}$$

Combining (B8) with (B4), we obtain

$$\eta \mathrm{tr} \left\{ \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}} \right\} = \alpha^2 \sigma_{\mathrm{d}}^2 K. \tag{B9}$$

Substituting (B9) into (B7), yields

$$\eta = \frac{\alpha^2 K \sigma_{\mathrm{d}}^2}{P} > 0, \quad \mathrm{tr} \left\{ \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}} \right\} = P. \tag{B10}$$

Then we derive the precoding matrix

$$\boldsymbol{P} = \frac{1}{\alpha} \left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}} \boldsymbol{I}_N \right)^{-1} \widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}}, \tag{B11}$$

and α is chosen to satisfy $\mathrm{tr} \left\{ \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}} \right\} = P$, given by

$$\alpha = \sqrt{\frac{\mathrm{tr} \left\{ \widehat{\boldsymbol{H}}_{\mathrm{d}} \left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}} \boldsymbol{I}_N \right)^{-2} \widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \right\}}{P}}. \tag{B12}$$

Substituting (B11) into (B1), the corresponding MSE is expressed as

$$\epsilon = \mathrm{tr} \left\{ \boldsymbol{I}_K - \widehat{\boldsymbol{H}}_{\mathrm{d}} \left(\widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}} \boldsymbol{I}_N \right)^{-1} \widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \right\}. \tag{B13}$$

Since $\boldsymbol{C}_{\mathrm{E}}$ is Hermitian, the second part in the brace is also Hermitian. Then the MSE ϵ can be described as $\epsilon = K - X$, where X is some positive real number, so it is smaller than K. Thus, Eq. (B11) is the optimal precoding matrix. This concludes the proof.

Appendix C Proof of Proposition 3

The SINR in (17) consists of tree terms: (1) the desired signal power $|(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^* |^2$; (2) the interference power $(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}k})^*$; (3) the term ξ. We will derive a deterministic equivalent for each term, and then they together constitute the final expression for γ_k°.

Let us show that the spectral norms of some matrices are uniformly bounded. We take the proof for $\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_k^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}$ where $\widehat{\boldsymbol{C}} = (\frac{1}{N} \widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \frac{1}{N} \boldsymbol{C}_{\mathrm{E}} + \frac{1}{c(N,K)\rho_{\mathrm{d}}} \boldsymbol{I}_N)^{-1}$ ($c\,(N,K) = \frac{N}{K}$) as an example. The proof for the other terms is similar, under the condition in Proposition 3 that $c\,(N,K) \leqslant c$ for all N.

The eigenvalues of $\widehat{\boldsymbol{C}}$ can be written as

$$\delta_k = \frac{1}{f(N)_k + \frac{1}{c(N,K)\rho_{\mathrm{d}}}}, \tag{C1}$$

where $f(N)_k \geqslant 0$ is the eigenvalue of the non-negative definite part $\frac{1}{N} \widehat{\boldsymbol{H}}_{\mathrm{d}}^{\mathrm{H}} \widehat{\boldsymbol{H}}_{\mathrm{d}} + \frac{1}{N} \boldsymbol{C}_{\mathrm{E}}$. Thus $0 < \delta_k \leqslant c\rho_{\mathrm{d}}$ is upper bounded for all N, which means that the spectral norm of $\widehat{\boldsymbol{C}}$ is uniformly bounded.

The spectral norm of $\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_k^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}$ can be written as

$$\|\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_k^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}\| \leqslant \|\boldsymbol{T}_{\mathrm{b}}^*\| \|\boldsymbol{\Phi}_k^{\mathrm{T}}\| \|\boldsymbol{T}_{\mathrm{b}}\| \|\widehat{\boldsymbol{C}}\|. \tag{C2}$$

Since the amplitudes of RF circuit gains are identically and uniformly distributed around 1, and we have assumed that all correlation matrices $\boldsymbol{\Phi}_k$ have uniformly bounded spectral norm on N, we can obtain that $\|\boldsymbol{T}_{\mathrm{b}}^* \boldsymbol{\Phi}_k^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}\|$ is uniformly bounded.

Now we will derive the deterministic equivalent for each term.

(1) Deterministic equivalent for $(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^*$.

By applying Lemma 1 of [20] (matrix inversion lemma) we have

$$(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^* = \frac{(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}}_{[k]} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^*}{1 + (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^{\mathrm{T}} \widehat{\boldsymbol{W}}_{[k]} (\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^*} \tag{C3}$$

$$= \frac{|r_{\mathrm{u}k}|^2 \boldsymbol{z}_k^{\mathrm{T}} \left(\boldsymbol{\Phi}_k^{\frac{1}{2}} \right)^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}_{[k]} \boldsymbol{T}_{\mathrm{b}}^* \left(\boldsymbol{\Phi}_k^{\frac{1}{2}} \right)^* \boldsymbol{z}_k^*}{1 + |r_{\mathrm{u}k}|^2 \left(\boldsymbol{\Phi}_k^{1/2} \boldsymbol{z}_k - \boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2} \boldsymbol{e}_{\mathrm{v}k} \right)^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}_{[k]} \boldsymbol{T}_{\mathrm{b}}^* \left(\boldsymbol{\Phi}_k^{1/2} \boldsymbol{z}_k - \boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2} \boldsymbol{e}_{\mathrm{v}k} \right)^*}$$

$$+ \frac{|r_{\mathrm{u}k}|^2 \boldsymbol{z}_k^{\mathrm{T}} \left(\boldsymbol{\Phi}_k^{\frac{1}{2}} \right)^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}_{[k]} \boldsymbol{T}_{\mathrm{b}}^* \left(\boldsymbol{\Phi}_{\boldsymbol{e}_{\mathrm{v}k}}^{\frac{1}{2}} \right)^* \boldsymbol{e}_k^*}{1 + |r_{\mathrm{u}k}|^2 \left(\boldsymbol{\Phi}_k^{1/2} \boldsymbol{z}_k - \boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2} \boldsymbol{e}_{\mathrm{v}k} \right)^{\mathrm{T}} \boldsymbol{T}_{\mathrm{b}} \widehat{\boldsymbol{C}}_{[k]} \boldsymbol{T}_{\mathrm{b}}^* \left(\boldsymbol{\Phi}_k^{1/2} \boldsymbol{z}_k - \boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2} \boldsymbol{e}_{\mathrm{v}k} \right)^*}, \tag{C4}$$

where $\widehat{\boldsymbol{W}}_{[k]} = [\sum_{j \neq k} (\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_j r_{\mathrm{u}j})^* (\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_j r_{\mathrm{u}j})^{\mathrm{T}} + \boldsymbol{C}_{\mathrm{E}} + \frac{K}{\rho_{\mathrm{d}}} \boldsymbol{I}_N]^{-1}$ and $\widehat{\boldsymbol{C}}_{[k]} = N \widehat{\boldsymbol{W}}_{[k]}$. Eq. (C4) is due to substituting (9) and $\widehat{\boldsymbol{v}}_k = \sqrt{N} (\boldsymbol{\Phi}_k^{1/2} \boldsymbol{z}_k - \boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2} \boldsymbol{e}_{\mathrm{v}k})$ into (C3).

By applying Lemma 14.2 of [34] and Lemma 5 of [32],

$$\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C5}$$

$$\boldsymbol{e}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_{\mathrm{e}_k}^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_{\mathrm{e}_k}^{\frac{1}{2}}\right)^*\boldsymbol{e}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{e}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C6}$$

$$\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{e}_k^*\xrightarrow[N\to\infty]{\text{a.s.}}0. \tag{C7}$$

Furthermore, we have

$$\left(\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k-\boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k-\boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k}\right)^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C8}$$

where $\boldsymbol{\Phi}_{\mathrm{eq}_k}=\boldsymbol{\Phi}_k+\boldsymbol{\Phi}_{\mathrm{e}_k}$.

With Lemma 14.3 of [34] we have

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C9}$$

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0. \tag{C10}$$

By applying Theorem 1 of [33], we obtain

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C11}$$

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C12}$$

where $\boldsymbol{\Psi}$ is defined in (26). Then we have

$$\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C13}$$

$$\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C14}$$

where $\widehat{\boldsymbol{g}}_k=\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k+\boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k}$.

Now we obtain the deterministic equivalent for desired signal power,

$$\left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k}\right)^{\mathrm{T}}\widehat{\boldsymbol{W}}\left(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k}\right)^*-\frac{|r_{\mathrm{u}k}|^2 m_k^{\circ}}{1+|r_{\mathrm{u}k}|^2 m_{\mathrm{eq}_k}^{\circ}}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C15}$$

where $m_k^{\circ}=\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\right\}$ and $m_{\mathrm{eq}_k}^{\circ}=\frac{1}{N}\mathrm{tr}\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}\}$.

(2) Deterministic equivalent for $|\left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{uk}\right)^{\mathrm{T}}\widehat{\boldsymbol{W}}\left(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j}\right)^*|^2$.

The undesired signal power can be rewritten as

$$\left|\left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k}\right)^{\mathrm{T}}\widehat{\boldsymbol{W}}\left(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j}\right)^*\right|^2=\frac{\left(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j}\right)^{\mathrm{T}}\widehat{\boldsymbol{W}}_{[j]}\left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k}\right)^*\left(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{\mathrm{u}k}\right)^{\mathrm{T}}\widehat{\boldsymbol{W}}_{[j]}\left(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j}\right)^*}{\left[1+|r_{\mathrm{u}j}|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[j]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*\right]^2} \tag{C16}$$

$$\xrightarrow[N\to\infty]{\text{a.s.}}\frac{|r_{\mathrm{u}j}|^2|r_{\mathrm{u}k}|^2}{N(1+|r_{\mathrm{u}j}|^2 m_{\mathrm{eq}_j}^{\circ})^2}\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[j]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[j]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*, \tag{C17}$$

where Eq. (C16) is due to applying Lemma 1 of [20] (matrix inversion lemma) for term $\widehat{\boldsymbol{W}}$; Eq. (C17) is obtained by applying (C14) for the denominator.

By applying Lemma 2 of [20] for term $\widehat{\boldsymbol{C}}_{[j]}$, we can write

$$\begin{aligned}
&\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[j]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[j]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*\\
&=\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*\\
&\quad+\frac{|r_{\mathrm{u}k}|^4\left|\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{1/2}\boldsymbol{z}_k-\boldsymbol{\Phi}_{\mathrm{e}_k}^{1/2}\boldsymbol{e}_{\mathrm{v}k}\right)^*\right|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*}{\left[1+|r_{\mathrm{u}k}|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*\right]^2}\\
&\quad-2\Re\left\{\frac{|r_{\mathrm{u}k}|^2\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*}{1+|r_{\mathrm{u}k}|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*}\right\}.
\end{aligned} \tag{C18}$$

By applying Lemma 14.2 of [34] and Theorem 1 of [33] as steps for deriving (C5) similarly, we can obtain

$$\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^*\boldsymbol{z}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C19}$$

$$\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C20}$$

$$\boldsymbol{z}_k^{\mathrm{T}}\left(\boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[jk]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C21}$$

where Eq. (C21) is similar as the numerator in (C15).

By applying Theorem 2 of [33] for which $\boldsymbol{Q}=\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_j^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}$ and $\boldsymbol{D}=\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}$, we have

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}_{\mathrm{T}j}'\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C22}$$

where $\boldsymbol{\Psi}_{\mathrm{T}j}'$ is given by (31). Similarly, we have

$$\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_j}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}\right\}-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}_{\mathrm{T}j}'\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0. \tag{C23}$$

By substituting (C7), (C13), (C19)–(C23) into (C18), then we obtain the deterministic equivalent for undesired signal power,

$$\left|(\boldsymbol{T}_{\mathrm{b}}\boldsymbol{v}_k r_{uk})^{\mathrm{T}}\widehat{\boldsymbol{W}}(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_j r_{\mathrm{u}j})^*\right|^2\xrightarrow[N\to\infty]{\text{a.s.}}\frac{|r_{\mathrm{u}j}|^2|r_{\mathrm{u}k}|^2}{N\left(1+|r_{\mathrm{u}j}|^2 m_{\mathrm{eq}_j}^{\circ}\right)^2}\left[\mu_{\mathrm{T}j}^{\circ}+\frac{|r_{\mathrm{u}k}|^4\left|m_k^{\circ}\right|^2\mu_{\mathrm{Te}j}^{\circ}}{\left(1+|r_{\mathrm{u}k}|^2 m_{\mathrm{eq}_k}^{\circ}\right)^2}-\frac{2|r_{\mathrm{u}k}|^2\Re\left\{m_k^{\circ}\mu_{\mathrm{T}j}^{\circ}\right\}}{\left(1+|r_{\mathrm{u}k}|^2 m_{\mathrm{eq}_k}^{\circ}\right)}\right], \tag{C24}$$

where $\mu_{\mathrm{T}j}^{\circ}$ is given by (28) and $\mu_{\mathrm{Te}j}^{\circ}$ is given by (29).

(3) Deterministic equivalent for ξ.

The term ξ can be rewritten as

$$\xi=\sum_{k=1}^{K}(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^{\mathrm{T}}\widehat{\boldsymbol{W}}^2(\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{v}}_k r_{\mathrm{u}k})^*=\frac{1}{N}\sum_{k=1}^{K}\frac{|r_{\mathrm{u}k}|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}^2\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*}{\left[1+|r_{\mathrm{u}k}|^2\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*\right]^2}, \tag{C25}$$

where Eq. (C25) is due to matrix inversion lemma.

By applying Theorem 2 of [33] as the steps for deriving (C22) for which $\boldsymbol{Q}=\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}$ and $\boldsymbol{D}=\boldsymbol{I}_N$, we have

$$\widehat{\boldsymbol{g}}_k^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\widehat{\boldsymbol{C}}_{[k]}^2\boldsymbol{T}_{\mathrm{b}}^*\widehat{\boldsymbol{g}}_k^*-\frac{1}{N}\mathrm{tr}\left\{\boldsymbol{T}_{\mathrm{b}}^*\boldsymbol{\Phi}_{\mathrm{eq}_k}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{b}}\boldsymbol{\Psi}_I'\right\}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C26}$$

where $\boldsymbol{\Psi}_I'$ is given by (31).

By substituting (C14), (C26) into (C25), we obtain the deterministic equivalent for ξ,

$$\xi-\frac{1}{N}\sum_{k=1}^{K}\frac{|r_{\mathrm{u}k}|^2\mu_{\mathrm{I}k}^{\circ}}{\left[1+|r_{\mathrm{u}k}|^2 m_{\mathrm{eq}_k}^{\circ}\right]^2}\xrightarrow[N\to\infty]{\text{a.s.}}0, \tag{C27}$$

where $\mu_{\mathrm{I}k}^{\circ}$ is given by (30).

Until this point, we have obtained all the deterministic equivalents of the three terms. Combining the results, the deterministic equivalent of the SINR of UT k is given by (20).