Large-Scale Antenna Systems with UL/DL Hardware Mismatch: Achievable Rate Analysis and Calibration

Wence Zhang, *Student Member, IEEE*, Cunhua Pan *Student Member, IEEE*, Ming Chen, *Member, IEEE* and Rodrigo C. de Lamare, *Senior Member, IEEE*, Jianxin Dai, *Member, IEEE*

Abstract

This paper studies the impact of hardware mismatch (HM) between the base station (BS) and the user equipment (UE) in the downlink (DL) of large-scale antenna systems. Analytical expressions to predict the achievable rates are derived for different precoding methods, i.e., matched filter (MF) and regularized zero-forcing (RZF), using large system analysis techniques. Furthermore, the asymptotic downlink signal to interference plus noise ratio (SINR) under HM is investigated, which is only related to the variances of the circuit gains of the mismatched hardware in most practical scenarios. Moreover, we prove that RZF is more sensitive to HM compared with MF in the high SNR region. ~~For~~ HM calibration, we take Zero-Forcing (ZF) precoding as an example to compare two HM calibration schemes, i.e., Pre-precoding Calibration (Pre-Cal) and Post-precoding Calibration (Post-Cal). The analysis shows that Pre-Cal outperforms Post-Cal schemes. Monte-Carlo simulations are carried out, and numerical results demonstrate the correctness of the analysis.

I. Introduction

Large-scale antenna systems or 'Massive MIMO systems' with a large number of antennas deployed at the base station (BS) have drawn great research interest recently [1], [2]. Operating in time-division duplexing (TDD) mode, large-scale antenna systems utilize the channel reciprocity in the uplink (UL) and the downlink (DL), and obtain the channel state information (CSI) through UL pilot training. With an increasing number of antennas, large-scale antenna systems can exploit excess degrees of freedom. As the degree of orthogonality between channels of different

W. Zhang, C. Pan and M. Chen are with National Mobile Communications Research Lab. (NCRL), Southeast University, China. Emails: {wencezhang, cunhuapan, chenming}@seu.edu.cn.

R. C. de Lamare is with Dept. of Electronics, the University of York, UK and CETUC, PUC-Rio, Brazil. Email:rcdl500@york.ac.uk.

J. Dai is with School of Science, Nanjing University of Posts and Telecommunications, China. Email:daijx@njupt.edu.cn.

The work of W. Zhang, C. Pan, and M. Chen was supported in part by China Scholarship Council (CSC), National Nature Science Foundation of China (Nos. 61372106 & 61223001 & 61172077), National 863 High Technology Development Project (No. 2013AA013601), Key Special Project of National Science and Technology (No. 2013ZX03003006), Research Fund of National Mobile Communications Research Laboratory, Southeast University of 2014. The work of R. C. de Lamare was supported by CNPq, PUC-Rio and the University of York.

Part of this work is accepted for publication in IEEE GlobeCom 2014.

users improves, inter-user interference can be suppressed by simple processing [3]. Large-scale antenna systems exhibits advantages in many aspects, such as data rates, error rates, computational complexity of signal processing and energy efficiency [2], [4].

Because different carrier frequency bands are used for UL and DL in frequency-division duplexing (FDD) systems, the DL channel coefficients are estimated at the user equipment (UE) and then sent back to the BS. So it brings heavy training overhead when the number of transmit and receive antennas ~~grows~~ large. Therefore, the benefits brought by large-scale antenna systems should be based on channel reciprocity in TDD mode, which considers the channel coefficients to remain the same in UL and DL within the channel coherent time. This is generally true for wireless propagation, for the electromagnetic waves in the UL and DL undergo the same physical propagation environment, i.e., reflections, refractions, scatters, etc [5]. However, because of the 'hardware mismatch' (HM) between UL and DL, i.e., the base station (BS) and the UEs employ different devices (e.g., antennas, filters and amplifiers), the equivalent channel coefficients that take into account the analog circuit gains differ in UL and DL. Although HM generally does not affect the performance in UL (the channel estimation is carried out in UL), it results in an essential constraint of the DL performance [6].

The HM generates uncertainty in the DL channel. One possible way to deal with this uncertainty is through robust design [7] at the cost of implementation complexity. However, ~~more research propose~~ to handle HM in the DL through estimation and compensation of these HM parameters [5], [8]–[15]. The estimation of HM parameters can be divided into two categories according to whether the users are involved or not. When the estimation of the HM parameters are carried out at the BS without the involvement of a user, we can adjust the circuit gains of all the BS antennas to a reference antenna by either resorting to additional circuitry in the transceivers at the BS [8], or by appointing an antenna as the reference and exchanging pilot signals between the reference antenna and all the other antennas at the BS [11], [12], [14]. This scheme is usually termed as 'self calibration' [5], or 'relative calibration' [11]. In this way, the circuit gains at all the antennas at the BS are calibrated to a constant value and will not affect the the channel reciprocity. The self calibration scheme can also be used to users with multiple antennas, but the performance degrades due to the low cost amplifiers used in users [16]. Another approach to obtain the HM parameters demands the assistance from the users and is often termed as 'over-the-air calibration' because pilot signals are transmitted between the BS and the users through the air interface [5], [9], [10], [13], [15]. ~~Major~~ problems of this approach are the occupation of system resources and the section of supporting users [10].

Once the HM parameters are reliably obtained, channel calibration is carried out to retrieve reciprocity between the UL and the DL. Some work only calibrate the HM parameters at the BS (we refer to these schemes as 'partial calibration'), because they have the dominant impact on the level of inter-user interference in the DL [8], [10]–[12]. However, the quality of service (QoS) of each user is difficult to guarantee without knowledge of the user's HM parameters. We can handle this either by employing additional filters at the users [8], or by compensation [5], [9], [13], [15]. We refer to these approaches as 'full calibration'.

As most research work on HM calibration, to the best of the authors' knowledge, there are few results of the

HM's impact on the DL performance except for [16], in which the problem of different HM calibration schemes is studied for ~~Block-Diagonal~~ (BD) and Zero-Forcing (ZF) precoding. However, few analytical results can be found and the impact of HM on DL performance remains unclear.

In this paper, the impact of HM on different DL precoding methods, i.e., matched filter (MF) and regularized zero-forcing (RZF) is investigated. MF is a potential technique for DL transmission in large-scale antenna systems due to its simplicity, low complexity and scalability [2], [11]. In contrast, RZF has higher computational complexity and inferior scalability, but it performs much better. We give the analytical expression of the achievable rates for both MF and RZF. Unlike prior work~~s~~ in [17], which analyze~~s~~ the DL performance of MF and RZF in large-scale antenna systems without considering HM between UL and DL, in this work we take HM into consideration, and the problem becomes very different especially for RZF. The correlated wireless channel matrix is further corrupted by two diagonal matrices comprising the HM parameters of the system, which demands ~~more mathematical tricks~~ to handle. By utilizing random matrix theory and some useful inequalities, we derive analytical expressions for the achievable rates of MF and RZF with HM in UL and DL for large-scale antenna systems.

As briefly described in [12], the HM calibration can be carried out either before or after the precoding. These two scenarios are referred to as Pre-precoding Calibration (Pre-Cal) and Post-precoding Calibration (Post-Cal). In this work we take ZF as an example to compare the performance of Pre-Cal and Post-Cal under both partial calibration and full calibration.

The main contributions of this paper are summarized as follows:

a) Analytical expressions for the achievable rates per user are given for both MF and RZF with consideration of UL and DL HM, through the evaluation of expectations and asymptotic determinant equivalents (ADE) of a series of random variables;

b) The asymptotic signal to interference plus noise ratio (SINR) of each user under HM is obtained for both MF and RZF, which shows the performance in the downlink under HM converges as the number of transmit antennas at BS increases, and is only related to the statistics of the HM parameters.

c) It is proved that RZF is more sensitive to HM compared with MF in the high SNR region.

d) We ~~prove~~ that the Pre-Cal schemes outperforms the Post-Cal schemes.

The rest of the paper is organized as follows. The system model is described in section II. In section III, the sets of achievable rates for both MF and RZF are derived, and the large system effect on HM is analyzed . The comparison of different HM calibration schemes is carried out in section IV. The simulations and numerical results are shown in section V, and conclusions are drawn in section VI.

Notation: $(\cdot)^*$ is the conjugate operation; $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denotes the transpose and Hermitian transpose of a vector or a matrix , respectively; $\mathbb{E}\{\cdot\}$ denotes the expectation operation; $[v_{nk}]_{N\times K}$ denotes an N by K matrix with the (n,k)-th element being v_{nk}; $\boldsymbol{I}_N$ denotes an N by N identity matrix; $\mathcal{CN}(\boldsymbol{\theta},\boldsymbol{\Sigma})$ denotes circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\theta}$ and covariance $\boldsymbol{\Sigma}$; "$\rightarrow$" means the left hand side term converges to the right hand side term as the number of antennas at BS grows large almost surely (a.s.); $f(x)|_a^b = f(b) - f(a)$; $\mathrm{diag}\{a_1, \cdots, a_N\}$ denotes an N by N diagonal matrix with diagonal entries given by $a_1, \cdots, a_N$.

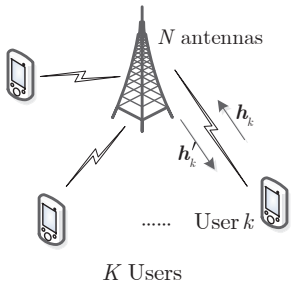

Fig. 1. Single cell MU Massive MIMO systems

II. System Model

Consider the DL of a single cell multi-user large-scale antenna system with N transmit antennas at BS and K single antenna users, where N is very large (tens or hundreds). The system operates in TDD mode, so that the channel reciprocity in the UL and DL can be exploited. An illustration of the system is given in Fig. 1.

Let $\boldsymbol{h}_k = [h_{1k}, \cdots, h_{Nk}]^{\mathrm{T}}$, $\boldsymbol{h}'_k = [h'_{1k}, \cdots, h'_{Nk}]$ be the UL and DL channel vectors of user k, respectively, where h_{nk} is the channel gain from user k to the n-th antenna at BS in the UL, and h'_{nk} is that in the DL. As shown in Fig. 2, h_{nk} and its DL counterpart h'_{nk} are modeled, respectively, as [11], [12]

$$\begin{aligned} h_{nk} &= r_n v_{nk} \tilde{t}_k, \\ h'_{nk} &= t_n v'_{nk} \tilde{r}_k, \end{aligned} \tag{1}$$

where v_{nk} and v'_{nk} are the corresponding UL and DL wireless channel gains, respectively; r_n ,t_n are the equivalent receive and transmit circuit gains of the n-th antenna at BS, and $\tilde{r}_k$,$\tilde{t}_k$ are the equivalent receive and transmit circuit gains of user k, respectively.

In fact, $\boldsymbol{h}_k$ and $\boldsymbol{h}'_k$ *are not exactly the same* due to HM between UL and DL. In TDD mode, because the electromagnetic waves in UL and DL undergo the same physical propagation environment , i.e., reflections, refractions, scatters, etc [5], the wireless channel gains in UL and DL transmission are considered to be unchanged within a channel coherent period, i.e., $v_{nk} = v'_{nk}$. However, the circuit gain is related to the hardware configurations, e.g., transmit and receive filters and amplifiers. Because the BS and user equipments usually have different hardware implementation, so that $r_i \neq \tilde{r}_i$ and $t_i \neq \tilde{t}_i$. These circuit gains may change with the working conditions, e.g., temperature [5]. But they vary slowly compared with the wireless channel, so it is reasonable to *treat them as known constants during the data transmission interval concerned*.

The received signal y_k at user k is given by

$$y_k = \boldsymbol{x}\boldsymbol{W}\boldsymbol{h}'_k + n_k, \quad \forall k = 1, \cdots, K, \tag{2}$$

where $\boldsymbol{x} = [x_1, \cdots, x_K]$ is the transmit signal vector consisting of independent, zero mean and unit energy symbols for respective users, so that $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\} = \boldsymbol{I}_K$; $\boldsymbol{W} \in \mathbb{C}^{K \times N}$ is the transmit precoding matrix; n_k is the Gaussian noise at the receiver of user k with normalized power, i.e., $n_k \sim \mathcal{CN}(0, 1)$.

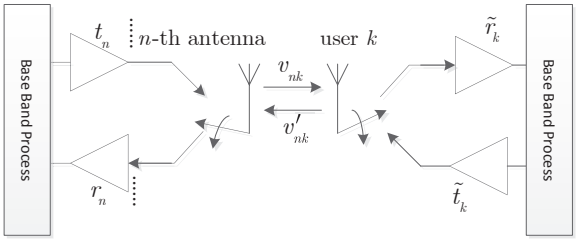

Fig. 2. Hardware Mismatch of UL and DL transmissions

Let the UL combined channel matrix of all users be $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]$, and $\boldsymbol{H} = \boldsymbol{R}\boldsymbol{V}\tilde{\boldsymbol{T}}$, in which $\boldsymbol{R} = \text{diag}(r_1, \cdots, r_N)$ is the receive circuit gain matrix of BS, $\tilde{\boldsymbol{T}} = \text{diag}(\tilde{t}_1, \cdots, \tilde{t}_K)$ is the transmit circuit gain matrix of all the users, and $\boldsymbol{V} = [v_{nk}]_{N\times K}$ is the wireless channel coefficient matrix. Denote the k-th column of $\boldsymbol{V}$ as $\boldsymbol{v}_k$, and $\boldsymbol{v}_k$'s are independent Gaussian vectors with correlated entries, i.e., $\boldsymbol{v}_k \sim \mathcal{CN}(\boldsymbol{0}, \frac{1}{N}\boldsymbol{\Phi}_k)$, $\forall k = 1, \cdots, K$. Note that the channel gain is normalized to $\frac{1}{N}$ for analysis convenience.

The DL channel matrix $\tilde{\boldsymbol{H}} = [\boldsymbol{h}'_1, \cdots, \boldsymbol{h}'_K]$ can be represented as $\tilde{\boldsymbol{H}} = \boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}}$, where $\boldsymbol{T} = \text{diag}(t_1, \cdots, t_N)$ is the transmit circuit gain matrix of BS, and $\tilde{\boldsymbol{R}} = \text{diag}(\tilde{r}_1, \cdots, \tilde{r}_K)$ is the receive circuit gain matrix of all the users.

Let $\boldsymbol{y} = [y_1, \cdots, y_K]$, $\boldsymbol{n} = [n_1, \cdots, n_K]$, and (2) can be reformulated as

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{W}\tilde{\boldsymbol{H}} + \boldsymbol{n}. \tag{3}$$

Remarks: 1) To focus on the effect of HM between UL and DL, we assume pathloss and shadow fading between transmitters and receivers are well estimated and compensated, so that we only consider the simple flat Rayleigh fading wireless channel in this work. 2) For multi-cell scenarios, the analysis in this paper is still valid by treating all the multi-cell interference as noise, which is actually the 'worst case'. Therefore the analysis will be a performance lower bound in multi-cell scenarios.

III. Achievable Rates of MF and RZF with HM

In this section, we first give a introduction on ~~a~~ sets of achievable rates which ~~is~~ upper bounded by the capacity and ~~is~~ very useful for performance analysis [17], [18]. Then we study the effect of HM on DL performance in terms of this achievable rates, and derive analytical expression for both MF and RZF under HM.

A. Achievable Rates

Similarly to [17], [18], in order to study the performance of large antenna systems with HM, we first introduce a set of achievable rates as shown in Lemma 1.

Lemma 1: Consider a point-to-point communication link, the signal model of which is given by

$$y_k = g_{k,k}x_k + \sum_{i=1, i\neq k}^{K} g_{k,i}x_i + n_k, \quad k = 1, \cdots, K, \tag{4}$$

where $\boldsymbol{x} = [x_1, \cdots, x_K]$ and n_k are as described in (2); $g_{k,i}$ is the equivalent channel gain, and $g_{k,i}$'s are correlated. Then the following set of rates is achievable

$$R_k = \log_2\left(1 + \gamma_k\right), \quad \forall k = 1, \cdots, K, \tag{5}$$

where

$$\gamma_k = \frac{\left|\mathbb{E}\left\{g_{k,k}\right\}\right|^2}{1 + \mathrm{var}\left\{g_{k,k}\right\} + \sum\limits_{i=1, i\neq k}^{K} \mathbb{E}\left\{\left|g_{k,i}\right|^2\right\}}. \tag{6}$$

Proof: See Theorem 1 in [18]. ■

By treating the variations of signals and all the correlated interferences as independent noise, this set of achievable rates for each user are actually upper bounded by the capacity, and thus *achievable*.

Note that in order to derive the closed form of the achievable rates of MF and RZF, we first formulate the signal model as in (4), and then find out the closed forms of the mathematical expectations involved in (6).

B. Achievable Rates with MF

When MF is used for DL transmission and assuming perfect CSI at BS, the precoding matrix is given by

$$\boldsymbol{W}_{\text{MF}} = \sqrt{\lambda_{\text{MF}}}\boldsymbol{H}^{\text{H}}, \tag{7}$$

in which

$$\lambda_{\text{MF}} = \frac{P}{\mathbb{E}\text{Tr}\left\{\boldsymbol{H}\boldsymbol{H}^{\text{H}}\right\}}, \tag{8}$$

and P is the total transmit power. Substituting (7) into (3) gives

$$\begin{aligned} \boldsymbol{y} &= \sqrt{\lambda_{\text{MF}}}\boldsymbol{x}\boldsymbol{H}^{\text{H}}\tilde{\boldsymbol{H}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{MF}}}\boldsymbol{x}\tilde{\boldsymbol{T}}^{\text{H}}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}} + \boldsymbol{n}. \end{aligned} \tag{9}$$

Therefore the received signal at user k can be represented as

$$\begin{aligned} y_k &= \sqrt{\lambda_{\text{MF}}}\left(\sum_{i=1}^{K} x_i\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\right)\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{v}_k\tilde{r}_k + n_k \\ &= g_{k,k}x_k + \sum_{i=1, i\neq k}^{K} g_{k,i}x_i + n_k, \end{aligned} \tag{10}$$

where $g_{k,i} = \sqrt{\lambda_{\text{MF}}}\tilde{t}_i^*\tilde{r}_k\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{v}_k$.

In (10), the first term represents the desired signal of user k, whilst the remaining ones are interference and noise. The achievable rates of MF are thus obtained using Lemma 1 and given by Theorem 1.

Theorem 1: When N goes large, the set of achievable rates with MF is $R_{\text{MF},k} = \log_2\left(1 + \gamma_{\text{MF},k}\right), \forall k = 1, \cdots, K$, where $\gamma_{\text{MF},k}$ is given by

$$\gamma_{\text{MF},k} \to \frac{\left|\tilde{t}_k^*\tilde{r}_k\right|^2\left|\frac{1}{N}\text{Tr}\left\{\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\right\}\right|^2}{\frac{1}{P}\sum\limits_{i=1}^{K}\left|\tilde{t}_i\right|^2\frac{1}{N}\text{Tr}\left\{\boldsymbol{R}\boldsymbol{\Phi}_i\boldsymbol{R}^{\text{H}}\right\} + \frac{1}{N}\sum\limits_{i=1, i\neq k}^{K}\left|\tilde{t}_i^*\tilde{r}_k\right|^2\frac{1}{N}\text{Tr}\left\{\boldsymbol{\Phi}_i\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\boldsymbol{T}^{\text{H}}\boldsymbol{R}\right\}}. \tag{11}$$

Proof: See Appendix B. ■

As can be seen in Theorem 1, the achievable rates of each user with MF is affected by HM parameters of both the BS and users.

C. Achievable Rates with RZF

Using RZF for DL transmitting, the precoding matrix is given by $\boldsymbol{W}_{\text{RZF}} = \sqrt{\lambda_{\text{RZF}}}[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \rho\boldsymbol{I}_K]^{-1}\boldsymbol{H}^{\text{H}}$, where $\rho = \frac{K}{P}$ [19], and

$$\lambda_{\text{RZF}} = \frac{P}{\mathbb{E}\text{Tr}\left\{[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \rho\boldsymbol{I}_K]^{-1}\boldsymbol{H}^{\text{H}}\boldsymbol{H}[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \rho\boldsymbol{I}_K]^{-1}\right\}}. \tag{12}$$

The received signal vector $\boldsymbol{y} \in \mathbb{C}^{1\times K}$ for the K users is

$$\begin{aligned}\boldsymbol{y} &= \sqrt{\lambda_{\text{RZF}}}\boldsymbol{x}[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \rho\boldsymbol{I}_K]^{-1}\boldsymbol{H}^{\text{H}}\tilde{\boldsymbol{H}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{RZF}}}\boldsymbol{x}\boldsymbol{H}^{\text{H}}[\boldsymbol{H}\boldsymbol{H}^{\text{H}} + \rho\boldsymbol{I}_N]^{-1}\tilde{\boldsymbol{H}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{RZF}}}\boldsymbol{x}\tilde{\boldsymbol{T}}^{\text{H}}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}\boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}} + \boldsymbol{n},\end{aligned} \tag{13}$$

where $\boldsymbol{Q} = [\boldsymbol{R}\boldsymbol{V}\tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^{\text{H}}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}} + \rho\boldsymbol{I}_N]^{-1}$.

By simple ~~derivations~~, the received signal for user k is given by

$$\begin{aligned}y_k &= \sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\sum_{i=1}^{K} x_i\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}\boldsymbol{T}\boldsymbol{v}_k + n_k \\ &= g_{k,k}x_k + \sum_{i=1,i\neq k}^{K} g_{k,i}x_i + n_k,\end{aligned} \tag{14}$$

where $g_{k,i} = \sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}\boldsymbol{T}\boldsymbol{v}_k$. Expand $\boldsymbol{Q}$ as

$$\boldsymbol{Q} = \left[\boldsymbol{R}\boldsymbol{V}_{-i}\tilde{\boldsymbol{T}}_{-i}\tilde{\boldsymbol{T}}_{-i}^{\text{H}}\boldsymbol{V}_{-i}^{\text{H}}\boldsymbol{R}^{\text{H}} + \left|\tilde{t}_i\right|^2\boldsymbol{R}\boldsymbol{v}_i\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}} + \rho\boldsymbol{I}_N\right]^{-1}, \tag{15}$$

in which

$$\begin{aligned}\boldsymbol{V}_{-i} &= [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_{i+1}, \cdots, \boldsymbol{v}_K], \\ \tilde{\boldsymbol{T}}_{-i} &= \text{diag}\left\{\tilde{t}_1, \cdots, \tilde{t}_{i-1}, \tilde{t}_{i+1}, \cdots, \tilde{t}_K\right\}.\end{aligned}$$

Let $\boldsymbol{Q}_{-i} = [\boldsymbol{R}\boldsymbol{V}_{-i}\tilde{\boldsymbol{T}}_{-i}\tilde{\boldsymbol{T}}_{-i}^{\text{H}}\boldsymbol{V}_{-i}^{\text{H}}\boldsymbol{R}^{\text{H}} + \rho\boldsymbol{I}]^{-1}$ and using Lemma 3 in Appendix, we get

$$g_{k,i} = \frac{\sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-i}\boldsymbol{T}\boldsymbol{v}_k}{1 + \left|\tilde{t}_i\right|^2\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-i}\boldsymbol{R}\boldsymbol{v}_i}. \tag{16}$$

The achievable rates with RZF are obtained using Lemma 1.

To deal with the mathematical expectations in (6), we derive asymptotic determinant equivalents (ADE) of random variables therein. When N and K are large, the random variables in (6) will converge to their ADEs almost surely, so the expectations can be approximated by their corresponding ADEs. For example, if a random variable $x \to x_0$ and x_0 is determinant, then x_0 is the ADE of x. So we can easily prove that $\mathbb{E}\{x\} \to x_0$ and $\text{var}\{x\} \to 0$.

1) ADE of $g_{k,k}$ and $g_{k,i}$ for $i \neq k$: The following proposition gives the ADE of $g_{k,k}$ and $g_{k,i}$ for $i \neq k$.

Proposition 1: When N grows large

$$\begin{aligned} g_{k,i} &\to 0, \forall k \neq i, \\ g_{k,k} &\to \frac{\sqrt{\lambda_{\mathrm{RZF}}}\tilde{r}_k \tilde{t}_k^* \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{T}\boldsymbol{\Phi}_k \boldsymbol{R}^{\mathrm{H}} \boldsymbol{S}(\rho)\right)}{1 + \left|\tilde{t}_k\right|^2 \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}\boldsymbol{\Phi}_k \boldsymbol{R}^{\mathrm{H}} \boldsymbol{S}(\rho)\right)}, \end{aligned} \tag{17}$$

where $\boldsymbol{S}(\rho)$ is given by

$$\boldsymbol{S}(\rho) = \left[\frac{1}{N}\sum_{i\in\Psi_1}\left\{\frac{\left|\tilde{t}_i\right|^2 \boldsymbol{R}\boldsymbol{\Phi}_i\boldsymbol{R}^{\mathrm{H}}}{1+e_i(\rho)}\right\} + \rho\boldsymbol{I}_N\right]^{-1}, \tag{18}$$

in which $\Psi_1 = \{1, \cdots, K\}\backslash\{k\}$ and $e_i(\rho)$ is the unique set of solutions to $K-1$ equations given by

$$e_i(\rho) = \frac{1}{N}\mathrm{Tr}\left\{\left|\tilde{t}_i\right|^2 \boldsymbol{R}\boldsymbol{\Phi}_i\boldsymbol{R}^{\mathrm{H}}\left[\frac{1}{N}\sum_{j\in\Psi_1}\left\{\frac{\left|\tilde{t}_j\right|^2 \boldsymbol{R}\boldsymbol{\Phi}_j\boldsymbol{R}^{\mathrm{H}}}{1+e_j(\rho)}\right\} + \rho\boldsymbol{I}_N\right]^{-1}\right\}, \quad i \in \Psi_1. \tag{19}$$

Proof: See Appendix C. ■

2) Derivation of λ_{RZF}: λ_{RZF} is a function of the combined channel matrix $\boldsymbol{H} = \boldsymbol{R}\boldsymbol{V}\tilde{\boldsymbol{T}}$. The wireless channel matrix $\boldsymbol{V}$ therein has independent columns, while the entries in each column are correlated. Because most results in random matrix theory focus on matrices with independent elements, it is very difficult to calculate λ_{RZF} directly.

The following analysis is based on the assumption that there is *limited correlation* among the transmit antennas. *Limited correlation* here means that there exists some $N \times N$ non-negative definite Hermitian matrix $\boldsymbol{\Phi}$ which satisfies

- AS1: The η transform of $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}$ is $\eta_{\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}}(\gamma) = 1/(1+\gamma)$;
- AS2: $\Delta = \sum_{k=1}^{K} \mathrm{Tr}\{(\tilde{t}_k\boldsymbol{R}\boldsymbol{\Phi}_k^{1/2} - \boldsymbol{\Phi})^{\mathrm{H}}(\tilde{t}_k\boldsymbol{R}\boldsymbol{\Phi}_k^{1/2} - \boldsymbol{\Phi})\} \leq \mathcal{O}(1)$.

Note that the η transform of a nonnegative random variable X is given by

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right],$$

where γ is a nonnegative real number. The η transform of a nonnegative definite random matrix refers to the η transform of its eigenvalues [See Definition 2.11 in [20]].

Because the variance of $\tilde{t}_k$'s and r_i's is not large compared with N, and mean value of these parameters is generally 1, $\boldsymbol{R}$ and $\tilde{\boldsymbol{T}}$ have comparatively small influence on AS2. In fact, AS2 shows that the wireless channels of all the users have ~~more or less~~ the same covariance matrix, which is given by $\boldsymbol{\Phi}$. Moreover, because of AS1, one typical choice of $\boldsymbol{\Phi}$ is the identity matrix $\boldsymbol{I}_N$. In this case, the covariance matrices of all the users' channels are actually a small deviation from $\boldsymbol{I}_N$. Therefore we call this scenario '*limited correlation*'.

These assumptions are generally true if all the transmit antennas at BS are geographically distributed, which results totally independent fading among different antennas, or in another scenario when for all the users, the sum of the squared correlation coefficients of all the antenna pairs does not grow with N, as is shown in the equation of AS2. In fact, these assumptions could be considered very strong. We will show in the simulation that even if these assumptions are not satisfied, the analytical results are still very accurate.

Using the aforementioned two assumptions, we give the following lemma.

Lemma 2: With AS1 and AS2, the asymptotic empirical spectrum distribution (ESD) of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ approximates the M-P Law, which is given by [21] and is described as

$$f^{\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}}(x)=\left(1-\frac{1}{\beta}\right)^{+}\delta(x)+\frac{\sqrt{(x-a)^{+}(b-x)^{+}}}{2\pi\beta x}, \tag{20}$$

where

$$a=(1-\sqrt{\beta})^2, b=(1+\sqrt{\beta})^2, \tag{21}$$

in which $(X)^{+}=\max\{0,X\}$ and $\beta=K/N$.

Proof: See Appendix D. ■

Now we are able to derive λ_{RZF} using Lemma 2.

Applying singular value decomposition (SVD) to $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$, we have $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{U}$ is a unitary matrix and $\boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_1,\cdots,\lambda_K\}$, the diagonal elements of which are the eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$. Substituting this into (12) yields

$$\begin{aligned}\lambda_{\mathrm{RZF}}&=\frac{P}{\mathbb{E}\mathrm{Tr}\left\{[\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}+\rho\boldsymbol{I}_K]^{-1}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}[\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}+\rho\boldsymbol{I}_K]^{-1}\right\}}\\&=\frac{P}{\mathbb{E}\left\{\sum_{i=1}^{K}\frac{\lambda_i}{(\rho+\lambda_i)^2}\right\}}.\end{aligned} \tag{22}$$

According to Lemma 2,

$$\begin{aligned}\mathbb{E}\left\{\sum_{i=1}^{K}\frac{\lambda_i}{(\rho+\lambda_i)^2}\right\}&=K\mathbb{E}\left\{\frac{\lambda_i}{(\rho+\lambda_i)^2}\right\}\\&=K\int_a^b\frac{1}{(\rho+x)^2}\frac{\sqrt{(x-a)(b-x)}}{2\pi\beta}\,\mathrm{d}x.\end{aligned} \tag{23}$$

Let $t=\rho+x$, and then

$$\begin{aligned}&\mathbb{E}\left\{\sum_{i=1}^{K}\frac{\lambda_i}{(\rho+\lambda_i)^2}\right\}\\&=\frac{K}{2\pi\beta}\int_{\rho+a}^{\rho+b}\frac{1}{t^2}\sqrt{(t-\rho-a)(\rho+b-t)}\,\mathrm{d}t.\end{aligned} \tag{24}$$

Denoting $t_1=\rho+b$, $t_2=\rho+a$, $\Delta=(t_1+t_2)^2-4t_1t_2$, and using equations 2.266, 2.261 and 2.267.2 in [22], we have

$$\begin{aligned}&\mathbb{E}\left\{\sum_{i=1}^{K}\frac{\lambda_i}{(\rho+\lambda_i)^2}\right\}\\&=\frac{K}{2\pi\beta}\arcsin\left[\frac{-2t+t_1+t_2}{\sqrt{\Delta}}\right]\Bigg|_{t_2}^{t_1}+\frac{K(t_1+t_2)}{4\pi\beta\sqrt{t_1t_2}}\arcsin\left[\frac{-2t_1t_2+(t_1+t_2)t}{t\sqrt{\Delta}}\right]\Bigg|_{t_2}^{t_1}.\end{aligned} \tag{25}$$

The expression of λ_{RZF} is achieved by substituting (25) into (22).

Until this point, we have obtained all the ingredients required to derive the achievable rates of RZF. Combining the results of (59), (17), (22) and (25), we give the following theorem.

Theorem 2: When N, K goes large, the set of achievable rates with RZF is given by

$$R_{\text{RZF},k} = \log_2\left(1+\gamma_{\text{RZF},k}\right), \qquad \forall k = 1,\cdots,K, \tag{26}$$

where

$$\gamma_{\text{RZF},k} \to \lambda_{\text{RZF}} \left| \frac{\tilde{r}_k \tilde{t}_k^* \frac{1}{N}\text{Tr}\left(\boldsymbol{T}\boldsymbol{\Phi}_k \boldsymbol{R}^{\text{H}} \boldsymbol{S}(\rho)\right)}{1+\left|\tilde{t}_k\right|^2 \frac{1}{N}\text{Tr}\left(\boldsymbol{R}\boldsymbol{\Phi}_k \boldsymbol{R}^{\text{H}} \boldsymbol{S}(\rho)\right)} \right|^2. \tag{27}$$

D. Large-Scale Antenna System Effect on HM

In this subsection we treat the HM parameters at BS as random variables while those at user equipment remain as known constants. In this way, we concentrate on how the number of antennas at BS affects the DL under HM. The following theorem shows that when the number of transmit antennas goes large, the performance in the DL with HM will converge.

Theorem 3: Assume that the HM parameters at BS r_i's are i.i.d., t_i's are i.i.d., r_i's and t_i's are independent. ~~Besides,~~ we assume *no correlation* exists among the N BS antennas. Then for a communication system modeled as (3), as N goes large, the received SINR of user k under MF, RZF precoding, i.e., $\gamma_{\text{MF},k}$, $\gamma_{\text{RZF},k}$, $\forall k = 1,\cdots,K$, will converge to

$$\gamma_{\text{MF},k} \to \frac{\left|\tilde{t}_k^* \tilde{r}_k\right|^2 \left|\mathbb{E}\left\{r_i^* t_i\right\}\right|^2}{\left[\rho\mathbb{E}\{\left|\tilde{t}_i\right|^2\} + \beta|\tilde{r}_k|^2\mathbb{E}\{|t_i^*|^2\}\mathbb{E}\{|t_i|^2\}\right]\mathbb{E}\{|r_i|^2\}}, \tag{28}$$

$$\gamma_{\text{RZF},k} \to \lambda_{\text{RZF}} \left| \frac{\mathbb{E}\left\{r_i^* t_i\right\} \tilde{t}_k^* \tilde{r}_k}{\mathbb{E}\{|r_i|^2\}\left|\tilde{t}_k\right|^2 + \rho} \right|^2. \tag{29}$$

Proof: For MF, according to the Law of Large Numbers (LLN), as N, K goes large

$$\begin{aligned}
&\tfrac{1}{N}\text{Tr}\left\{\boldsymbol{R}^{\text{H}}\boldsymbol{T}\right\} \to \mathbb{E}\left\{r_i^* t_i\right\}, \\
&\tfrac{1}{N}\text{Tr}\left\{\boldsymbol{R}\boldsymbol{R}^{\text{H}}\right\} \to \mathbb{E}\{|r_i|^2\}, \\
&\tfrac{1}{N}\text{Tr}\left\{\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{T}^{\text{H}}\boldsymbol{R}\right\} \to \mathbb{E}\{|r_i|^2\}\mathbb{E}\{|t_i|^2\}, \\
&\frac{1}{K-1}\sum_{i=1,i\neq k}^{K}\left|\tilde{t}_i^*\tilde{r}_k\right|^2 \to |\tilde{r}_k|^2\mathbb{E}\left\{|t_i{}^*|^2\right\}.
\end{aligned} \tag{30}$$

~~Taking~~ (30) into (11) results in (28).

For RZF, let $\bar{\boldsymbol{v}}_i$ be the i-th row of $\boldsymbol{V}$, $\bar{\boldsymbol{v}}_i^{\text{H}} \sim \mathcal{CN}(\boldsymbol{0}, \frac{1}{N}\boldsymbol{I}_K)$, and then using LLN

$$\begin{aligned}
\boldsymbol{y} &= \sqrt{\lambda_{\text{RZF}}}\boldsymbol{x}[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \rho\boldsymbol{I}_K]^{-1}\boldsymbol{H}^{\text{H}}\tilde{\boldsymbol{H}} + \boldsymbol{n} \\
&= \sqrt{\lambda_{\text{RZF}}}\boldsymbol{x}\left[\tilde{\boldsymbol{T}}^{\text{H}}\frac{1}{N}\left(\sum_{i=1}^{N}|r_i|^2\bar{\boldsymbol{v}}_i^{\text{H}}\bar{\boldsymbol{v}}_i\right)\tilde{\boldsymbol{T}} + \frac{1}{N}\rho\boldsymbol{I}_K\right]^{-1}\tilde{\boldsymbol{T}}^{\text{H}}\frac{1}{N}\left(\sum_{i=1}^{N}r_i^* t_i\bar{\boldsymbol{v}}_i^{\text{H}}\bar{\boldsymbol{v}}_i\right)\tilde{\boldsymbol{R}} + \boldsymbol{n} \\
&\to \sqrt{\lambda_{\text{RZF}}}\mathbb{E}\left\{r_i^* t_i\right\}\boldsymbol{x}\left[\mathbb{E}\left\{|r_i|^2\right\}\tilde{\boldsymbol{T}}^{\text{H}}\tilde{\boldsymbol{T}} + \rho\boldsymbol{I}_K\right]^{-1}\tilde{\boldsymbol{T}}^{\text{H}}\tilde{\boldsymbol{R}} + \boldsymbol{n}.
\end{aligned} \tag{31}$$

The received signal of the k-th user is thus given by

$$y_k \to \frac{\sqrt{\lambda_{\text{RZF}}}\mathbb{E}\left\{r_i^* t_i\right\}\tilde{t}_k^*\tilde{r}_k}{\mathbb{E}\left\{|r_i|^2\right\}\left|\tilde{t}_k\right|^2 + \rho}x_k + n_k. \tag{32}$$

and the SINR of user k is obtained as (29). ■

It can be seen from (28) and (29) that when the means of r_i's and t_i's are 1 (in most practical scenarios), the only performance loss of HM compared with symmetric UL/DL channels is due to the receive circuit gains at each user and the variances of circuit gains at BS. In fact, RZF is more sensitive to HM than MF. Because the good performance of RZF relies on inter-user interference control, when HM exists, however, the orthogonality between different users' signals degrades due to the mismatched UL/DL channel. We summarize this in the following corollary.

Corollary 1: For scenarios described in Theorem 3, RZF is more sensitive to HM than MF in the high SNR region.

Proof: When the means of r_i's and t_i's are 1 and the variance is given by σ^2, (28) changes to

$$\gamma_{\mathrm{MF},k} \to \frac{|\tilde{t}_k^* \tilde{r}_k|^2}{[\rho + \beta|\tilde{r}_k|^2(1+\delta^2)](1+\sigma^2)^2}. \tag{33}$$

Assuming $\tilde{t}_k \approx 1$ and $\tilde{r}_k \approx 1$, (33) is approximated by

$$\gamma_{\mathrm{MF},k} \approx \frac{1}{[\rho + \beta(1+\sigma^2)](1+\sigma^2)^2}. \tag{34}$$

Since $(1+x)^2 \approx 1 + 2x$ for small value of x, we have

$$\begin{aligned} \gamma_{\mathrm{MF},k} &\approx \frac{1}{\rho + \beta + (2\rho + 3\beta)\sigma^2 + 2\beta\sigma^4} \\ &\approx \frac{1}{\rho + \beta + (2\rho + 3\beta)\sigma^2}. \end{aligned} \tag{35}$$

Similarly for RZF

$$\begin{aligned} \gamma_{\mathrm{RZF},k} &\approx \frac{\lambda_{\mathrm{RZF}}}{1 + \rho + (2\rho + 2)\sigma^2 + \sigma^4} \\ &\approx \frac{\lambda_{\mathrm{RZF}}}{1 + \rho + (2\rho + 2)\sigma^2}. \end{aligned} \tag{36}$$

Differentiating (35) and (36) with respect to σ^2 yields

$$\begin{aligned} \frac{\partial \gamma_{\mathrm{MF},k}}{\partial(\sigma^2)} &\approx \frac{-(2\rho + 3\beta)}{[\rho + \beta + (2\rho + 3\beta)\sigma^2]^2}, \\ \frac{\partial \gamma_{\mathrm{RZF},k}}{\partial(\sigma^2)} &\approx \frac{-\lambda_{\mathrm{RZF}}(2\rho + 2)}{[1 + \rho + (2\rho + 2)\sigma^2]^2}. \end{aligned} \tag{37}$$

In the high SNR region, ρ is small. In this case the RZF will converge to ZF and $\lambda_{\mathrm{RZF}} \to \lambda_{\mathrm{ZF}} = \frac{1-\beta}{\rho}$ (See (40) in Section IV). Therefore, the variance of HM parameters is small, i.e., $\sigma^2 \to 0$, we have

$$\begin{aligned} \frac{\partial \gamma_{\mathrm{MF},k}}{\partial(\sigma)^2} &\approx -\frac{3}{\beta}, \\ \frac{\partial \gamma_{\mathrm{RZF},k}}{\partial(\sigma)^2} &\approx -\frac{2(1-\beta)}{\rho}. \end{aligned} \tag{38}$$

From (38), it is clear that the sensitivity of RZF to HM is characterised by both β and ρ, while only β has major influence on the sensitivity of MF to HM. When the value of β is moderate (neither too small nor too large), the performance of RZF is far more sensitive to HM compared with MF in the high SNR region ($\rho \to 0$). ■

Note that, in the low SNR region RZF degrades to MF. So it is easy to conclude that RZF and MF have the same sensitivity to HM. In fact, because of the matrix inversion involved in RZF, different choice of ρ greatly affect the sensitivity to RZF.

IV. Comparison of HM Calibration Schemes

With HM, the channels in the UL and DL cannot be reciprocal. So the DL performance will be degraded due to the inaccurate precoding matrix calculated based on UL channel estimation. As analyzed in the previous section, the more severe HM is, the more significant the performance loss becomes. Therefore a channel calibration process is usually preferred. In this section, we first take ZF as an example to show the effects of HM. Then we compare two sorts of HM calibration schemes. The choice of ZF is based on the fact that ZF has simpler expression than RZF and shows more insights of HM's influence than MF, which becomes clear in the following analysis.

When ZF is used for downlink transmission, and assuming perfect CSI at the BS, the precoding matrix is given by

$$\boldsymbol{W}_{\text{ZF}} = \sqrt{\lambda_{\text{ZF}}}\left(\boldsymbol{H}^{\text{H}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\text{H}}, \tag{39}$$

where

$$\begin{aligned}\lambda_{\text{ZF}} &= \frac{P}{\mathbb{E}\text{Tr}\left\{(\boldsymbol{H}^{\text{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\text{H}}\boldsymbol{H}^{\text{H}}(\boldsymbol{H}^{\text{H}}\boldsymbol{H})^{-1}\right\}} \\ &= \frac{P}{\mathbb{E}\text{Tr}\left\{(\boldsymbol{H}^{\text{H}}\boldsymbol{H})^{-1}\right\}} \\ &\to \frac{P(1-\beta)}{K},\end{aligned} \tag{40}$$

in which the last term is achieved using equation (2.104) in [20] with AS1 and AS2 .

Substituting (39) into (3) results in

$$\boldsymbol{y} = \sqrt{\lambda_{\text{ZF}}}\boldsymbol{x}[\boldsymbol{H}^{\text{H}}\boldsymbol{H}]^{-1}\boldsymbol{H}^{\text{H}}\tilde{\boldsymbol{H}} + \boldsymbol{n}. \tag{41}$$

Equation (41) can be expanded as

$$\begin{aligned}\boldsymbol{y} &= \sqrt{\lambda_{\text{ZF}}}\boldsymbol{x}[\boldsymbol{H}^{\text{H}}\boldsymbol{H}]^{-1}\boldsymbol{H}^{\text{H}}\tilde{\boldsymbol{H}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{ZF}}}\boldsymbol{x}[\tilde{\boldsymbol{T}}^{\text{H}}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{R}\boldsymbol{V}\tilde{\boldsymbol{T}}]^{-1}\tilde{\boldsymbol{T}}^{\text{H}}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{ZF}}}\boldsymbol{x}\tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{R}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}} + \boldsymbol{n}.\end{aligned} \tag{42}$$

As can be seen from (42), if $\tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{R}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{T}\boldsymbol{V}\tilde{\boldsymbol{R}}$ is not diagonal, inter-user interference cannot be completely eliminated by ZF. This is caused by differences between $\boldsymbol{R}$ and $\boldsymbol{T}$ and the resulting biased beam directions in the DL. Hence a calibration scheme can be introduced to adapt ZF to HM.

A. An Introduction of Partial and Full Calibration Schemes

In this subsection, we give an introduction of partial calibration and full calibration, and describe the two different ways to implement such calibration schemes.

1) Partial Calibration:

Partial calibration is used to eliminate HM's impact on DL beam directions without considering the HM parameters at the users. It adjusts the relative circuit gains of all the antennas at BS to a reference antenna. Approaches of such partial calibration can be found in [8], [10], [11].

When the HM parameters of the BS are obtained, there are basically two ways to carry out the HM calibration as mentioned briefly in [12]. One is to calibrate the DL channel after precoding, namely 'Post-precoding Calibration' (Post-Cal); the other one is to calibrate the estimated UL channel and use it for precoding, namely 'Pre-precoding Calibration' (Pre-Cal). For partial calibration, we refer to the two methods as Partial Post-Cal (P-Post-Cal) and Partial Pre-Cal (P-Pre-Cal), respectively.

For P-Post-Cal, let us suppose linear calibration is used as

$$\boldsymbol{W}_{\text{P-Post}} = \sqrt{\lambda_{\text{P-Post}}}[\boldsymbol{H}^{\text{H}}\boldsymbol{H}]^{-1}\boldsymbol{H}^{\text{H}}\boldsymbol{A}, \tag{43}$$

where the calibration matrix $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ is invertible. For P-Pre-Cal methods, let us suppose $\boldsymbol{A}^{-1}\boldsymbol{H}$ is used for precoding, and

$$\boldsymbol{W}_{\text{P-Pre}} = \sqrt{\lambda_{\text{P-Pre}}}[\boldsymbol{H}^{\text{H}}(\boldsymbol{A}^{-1})^{\text{H}}\boldsymbol{A}^{-1}\boldsymbol{H}]^{-1}\boldsymbol{H}^{\text{H}}(\boldsymbol{A}^{-1})^{\text{H}}. \tag{44}$$

The reason ~~of~~ introducing $\boldsymbol{A}^{-1}$ in (44) is to unify the expression of sufficient conditions for nulling inter-user interference in (43) and (44).

One sufficient condition for zero downlink inter-user interference is given by $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{T}^{-1}$ [11], [12]. Substituting $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{T}^{-1}$ and (43) into (3) gives

$$\boldsymbol{y} = \sqrt{\lambda_{\text{P-Post}}}\boldsymbol{x}\tilde{\boldsymbol{T}}^{-1}\tilde{\boldsymbol{R}} + \boldsymbol{n}, \tag{45}$$

The same holds for the P-Pre-Cal case. By choosing such $\boldsymbol{A}$, both P-Post-Cal and P-Pre-Cal methods are able to eliminate inter-user interference in the downlink.

2) Full Calibration:

It can be seen from (45), the HM parameters at the users will change the average received signal power. Although ~~as pointed out in [14],~~ the entries in $\tilde{\boldsymbol{T}}^{-1}\tilde{\boldsymbol{R}}$ can be estimated at each users as part of the DL channel, it could be important regarding QoS concerns, for the received signal power of every user is unpredicted by the BS, and the QoS concerns cannot be guaranteed. Another consideration would be that the the partial calibration algorithm in [11] is not suitable for RZF and some non-linear precoding schemes which needs accurate information of the gains of each user's sub-channels. Therefore full calibration is necessary, which exchanges pilot signals between the BS and the users through the air interface and estimates the HM parameters at both the BS and the users.

Many different versions of full calibration can be found in previous work, e.g., [5] and [15], and detailed discussion of these schemes are beyond the scope of this paper. Here for convenience, an demonstration of such calibration is described as follows.

For the Post-Cal methods, an additional calibration matrix $\boldsymbol{B} = \tilde{\boldsymbol{R}}^{-1}\tilde{\boldsymbol{T}}$ is introduced to (45). The uplink channel from j-th user to a reference antenna at BS is represented as $h_{1j} = r_1 v_{1j}\tilde{t}_j$, and its downlink counterpart is

$h'_{1j} = t_1 v_{1j} \tilde{r}_j$. h_{1j} can be obtained at BS using uplink pilots, while h'_{1j} must be reported to BS by user j. Then we have

$$\frac{h_{1j}}{h'_{1j}} = \frac{r_1 v_{1j} \tilde{t}_j}{t_1 v_{1j} \tilde{r}_j} = \frac{r_1 \tilde{t}_j}{t_1 \tilde{r}_j} \Rightarrow \frac{\tilde{t}_j}{\tilde{r}_j} = \frac{h_{1j}}{h'_{1j}} \frac{t_1}{r_1} \triangleq \frac{h_{1j}}{h'_{1j}} \mu. \tag{46}$$

Thus the calibration matrix $\boldsymbol{B}$ is constructed as $\boldsymbol{B} = \mathrm{diag}\left\{\mu h_{11}/h'_{11}, \cdots, \mu h_{1K}/h'_{1K}\right\}$.

Once $\boldsymbol{A}$ and $\boldsymbol{B}$ are acquired, the Post-Cal scheme is carried out as

$$\boldsymbol{W}_{\text{F-Post}} = \sqrt{\lambda_{\text{F-Post}}} \boldsymbol{B} [\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H}]^{-1} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{A}, \tag{47}$$

and then the received signal vector at K users is given by

$$\begin{aligned} \boldsymbol{y} &= \sqrt{\lambda_{\text{F-Post}}} \boldsymbol{x} \boldsymbol{B} \tilde{\boldsymbol{T}}^{-1} \tilde{\boldsymbol{R}} + \boldsymbol{n} \\ &= \sqrt{\lambda_{\text{F-Post}}} \boldsymbol{x} + \boldsymbol{n}. \end{aligned}$$

which indicates no inter-user interference due to HM exists in the system. Therefore, by using calibration matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, all the impact of HM in the DL is compensated for ZF. We refer to the calibration scheme based on (47) as Full Post Calibration (F-Post-Cal).

For the Pre-Cal scheme, with the same $\boldsymbol{A}$ and $\boldsymbol{B}$ as F-Post-cal, we have

$$\begin{aligned} \boldsymbol{A}^{-1} \boldsymbol{H} \boldsymbol{B}^{-1} &= \boldsymbol{T} \boldsymbol{R}^{-1} \boldsymbol{V} \tilde{\boldsymbol{T}}^{-1} \tilde{\boldsymbol{R}} \\ &= \boldsymbol{T} \boldsymbol{V} \tilde{\boldsymbol{R}} = \tilde{\boldsymbol{H}}. \end{aligned}$$

Thus the accurate values of DL channel coefficients are obtained by multiplying $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ to the UL channel matrix $\boldsymbol{H}$. Any precoding scheme can be carried out based on the calibrated UL channel matrix $\boldsymbol{A}^{-1} \boldsymbol{H} \boldsymbol{B}^{-1}$. We call this method Full Pre-Cal (F-Pre-Cal). After calibration, the equivalent ZF precoding matrix for F-Pre-Cal becomes

$$\boldsymbol{W}_{\text{F-Pre}} = \sqrt{\lambda_{\text{F-Pre}}} [\tilde{\boldsymbol{H}}^{\mathrm{H}} \tilde{\boldsymbol{H}}]^{-1} \tilde{\boldsymbol{H}}^{\mathrm{H}}.$$

Remarks: 1) ~~In comparison of~~ partial and full calibration approaches, there is a tradeoff between complexity and performance. It is apparent that full calibration demands assistance from the users and thus will cost more system resources. 2) In general full calibration has superior performance and the F-Pre-cal scheme is able to suit any precoding scheme.

B. Comparison between Post-Cal and Pre-Cal Schemes

After partial calibration, the equivalent ZF Precoding matrices for P-Post-Cal and P-Pre-Cal become, respectively

$$\begin{aligned} \boldsymbol{W}_{\text{P-Post}} &= \sqrt{\lambda_{\text{P-Post}}} [\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H}]^{-1} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{A} \\ &= \sqrt{\lambda_{\text{P-Post}}} \tilde{\boldsymbol{T}}^{-1} [\boldsymbol{V}^{\mathrm{H}} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{R} \boldsymbol{V}]^{-1} \boldsymbol{V}^{\mathrm{H}} \boldsymbol{R}^{\mathrm{H}} \boldsymbol{R} \boldsymbol{T}^{-1}, \\ \boldsymbol{W}_{\text{P-Pre}} &= \sqrt{\lambda_{\text{P-Pre}}} [\boldsymbol{H}^{\mathrm{H}} (\boldsymbol{A}^{-1})^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{H}]^{-1} \boldsymbol{H}^{\mathrm{H}} (\boldsymbol{A}^{-1})^{\mathrm{H}} \\ &= \sqrt{\lambda_{\text{P-Pre}}} \tilde{\boldsymbol{T}}^{-1} [\boldsymbol{V}^{\mathrm{H}} \boldsymbol{T}^{\mathrm{H}} \boldsymbol{T} \boldsymbol{V}]^{-1} \boldsymbol{V}^{\mathrm{H}} \boldsymbol{T}^{\mathrm{H}}. \end{aligned}$$

From the above equations, it can be seen that the only difference between $\boldsymbol{W}_{\text{P-Post}}$ and $\boldsymbol{W}_{\text{P-Pre}}$ is the additional term $\boldsymbol{A} = \boldsymbol{R}\boldsymbol{T}^{-1}$ in $\boldsymbol{W}_{\text{P-Post}}$ if $\boldsymbol{R}$ and $\boldsymbol{T}$ have the same distributions. However, the performance ~~would not be~~ the same because of $\lambda_{\text{P-Post}}$ and $\lambda_{\text{P-Pre}}$. Actually when a transmit power constraint exists, the performance of P-Pre-Cal methods is better than the P-Post-Cal schemes, which is shown in the following proposition.

Proposition 2: Provided that the HM parameters at both BS and uesers are independently and uniformly distributed in the range of $(1-0.5\delta, 1+0.5\delta)$ with $0 < \delta < 2$, $\lambda_{\text{P-Pre}} \geq \lambda_{\text{P-Post}}$ under AS1 and AS2 in Section III.

Proof: See Appendix E. ■

For full calibration, the equivalent ZF precoding matrices for F-Post-Cal and F-Pre-Cal become, respectively

$$\boldsymbol{W}_{\text{F-Post}} = \sqrt{\lambda_{\text{F-Post}}}\boldsymbol{B}[\boldsymbol{H}^{\text{H}}\boldsymbol{H}]^{-1}\boldsymbol{H}^{\text{H}}\boldsymbol{A},$$

$$\boldsymbol{W}_{\text{F-Pre}} = \sqrt{\lambda_{\text{F-Pre}}}[\tilde{\boldsymbol{H}}^{\text{H}}\tilde{\boldsymbol{H}}]^{-1}\tilde{\boldsymbol{H}}^{\text{H}}.$$

Note that the difference between the F-Pre-Cal and F-Post-Cal methods for ZF precoding lies in the equivalent channel used for data transmission. F-Post-Cal calibrates the DL channel according to the UL channel. The equivalent channel used for data transmission is actually the UL one. However, the F-Pre-Cal use the UL channel to predict the real DL channel and therefore the equivalent channel is the DL channel. Similar to the partial calibration schemes, F-Pre-Cal outperforms F-Post-Cal as stated in the following Proposition.

Proposition 3: Provided that the HM parameters are independently and uniformly distributed in the range of $(1-0.5\delta, 1+0.5\delta)$ with $0 < \delta < 2$, $\lambda_{\text{F-Pre}} \geq \lambda_{\text{F-Post}}$ under AS1 and AS2 in Section III.

Proof: The proof is similar to that of Proposition 2 and thus omitted. ■

The additional term $\boldsymbol{A}$ introduced in $\boldsymbol{W}_{\text{P-Post}}$ generates a smaller power factor and thus degrades the overall performance. The main reason is that when the diagonal entries of $\boldsymbol{A}$ has a mean value greater than 1, it causes a relatively larger power consumption. When a power constraint exists, the power factor used to normalize the transmit power becomes smaller, and the average received SNR decreases consequently. For the same reason that the diagonal entries of $\boldsymbol{B}$ and $\boldsymbol{A}$ have a mean value greater than 1, $\lambda_{\text{F-Pre}} > \lambda_{\text{F-Post}}$.

Remarks: 1) Although Proposition 2 and Proposition 3 are proposed assuming a certain uniform distribution for the HM parameters, the results are generally true as long as the mean of the diagonal entries of $\boldsymbol{A}$ is larger than 1. We choose uniform distribution in these propositions because this is one typical distribution for HM parameters and is widely used in simulations [12]. 2) These results can also be extended to MF and RZF. The proofs and discussions follow similar procedure as ZF and are thus omitted here.

V. Numerical Results

Similar to [12], the simulations are carried out by assuming that t_i, r_i, $\tilde{t}_i$, $\tilde{r}_i$ are identically and uniformly distributed in the range of $(1-0.5\delta, 1+0.5\delta)$. The larger δ is, the more severer HM is. The correlation matrix of each user's channel, i.e., $\boldsymbol{\Phi}_{\mathbf{k}}$, $k = 1, \cdots, K$, is also randomly generated for each simulation.

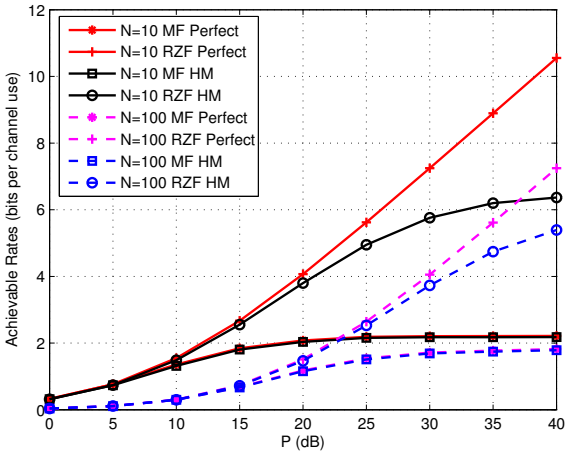

Fig. 3. Achievable rates per user with and without HM when $\beta = 0.4$, $\delta = 0.4$ for HM.

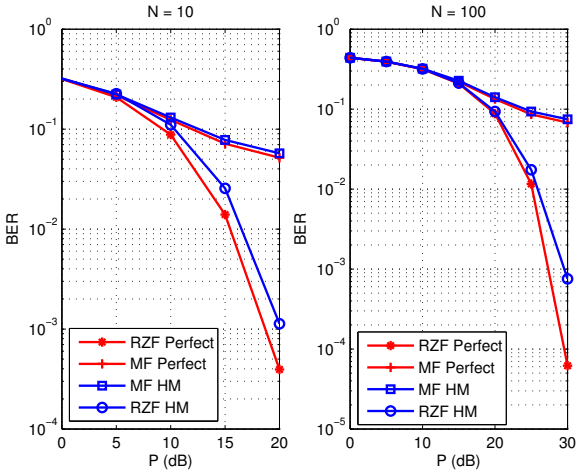

Fig. 4. Bit Error Rates of QPSK with and without HM when $\beta = 0.4$, $\delta = 0.4$ for HM.

A. Performance Loss due to HM

In this subsection, the performance loss due to HM is evaluated through simulation. All the simulation results are averaged over randomly generated HM parameters and channel samples. The loss caused by HM in terms of achievable rates is illustrated in Fig. 3 for both MF and RZF, and the simulation results of empirical cumulative distribution function (CDF) of data rates per user, the bit error rate (BER) of QPSK are shown in Fig. 4 and Fig. 5, respectively. In all the figures, 'Perfect' means perfect UL/DL hardware with no mismatch and 'HM' stands for scenario with HM. It is worth to mention that the simulation results of $N = 10$ are better than those of $N = 100$, because the channel gain is normalized to $\frac{1}{N}$ (please see description of $\boldsymbol{v}_k$ in Section II), and 10 times more users are served when $N = 100$.

As can be seen in Fig. 3, Fig. 4 and Fig. 5, the performance of RZF is much better than MF in terms of achievable

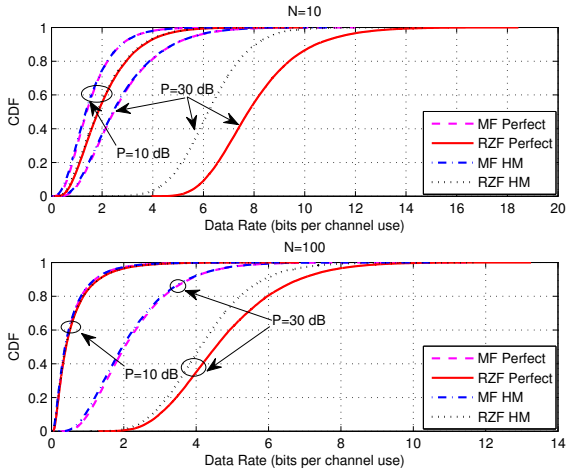

Fig. 5. Empirical CDF of data rates per user with and without HM when $\beta = 0.4$, $\delta = 0.4$ for HM.

rates, BER and CDF of data rates per user. The achievable rates of RZF grow almost linearly in the SNR range, while those of MF level out when SNR is high. Besides, a significant performance loss (as much as 5-10 dB) is observed for RZF when $N = 100$, $\delta = 0.4$ in high SNR region. For BER results of QPSK modulation, MF sees an error floor around 10^{-1}-10^{-2} in high SNR region, and RZF experiences a performance loss of 2 dB due to HM. When it comes to empirical CDF of data rates per user and when SNR is large, a remarkable performance gap caused by HM can be seen in simulation results of RZF, while the results of MF with HM almost coincide with those with perfect UL/DL hardware.

This ..

From the simulation results, we can tell that the performance of MF with and without HM is comparable, while there are significant performance loss in high SNR region for RZF when HM exists. This is due to the fact that the performance of MF is limited by the severe inter-user interference, so it is less sensitive to HM as compared with RZF. In order to improve the performance of MF, a smaller value of β is required.

Another viewpoint to interpret RZF's significant performance loss in high SNR comes from (32). When N is large and the mean of HM parameter is 1, the asymptotic SINR of received signal of each user with RZF is proportionally to $\frac{1}{K/P+\delta^2/12}$. When P is large or K is small, the impact of HM becomes more severe.Therefore taking HM into consideration can lead to remarkably better downlink performance for RZF.

B. Achievable rates of MF and RZF with HM

Fig. 6 illustrates the analysis and Monte-Carlo simulation results on achievable rates of MF and RZF, where $\delta = 0.2$, $N = 100$, $\beta = 0.2, 0.4$. The analytical achievable rates with MF and RZF are given by Theorem 1 and Theorem 2, respectively. As the transmit power P increases, the achievable rates of both RZF and MF increases, and higher achievable rates per user, can be obtained for smaller value of β. In the high SNR region, RZF sees

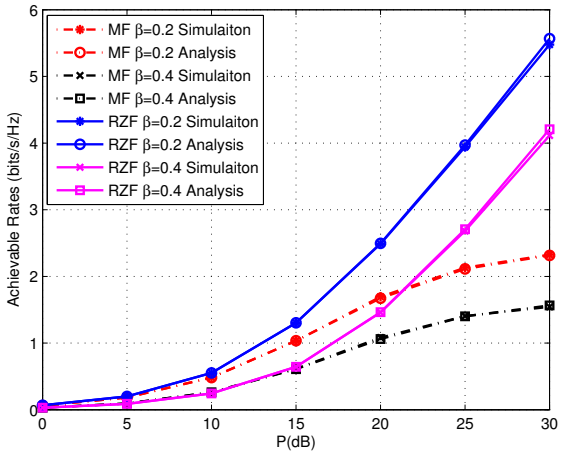

Fig. 6. Achievable rates of MF, RZF when $N = 100$, $\delta = 0.2$.

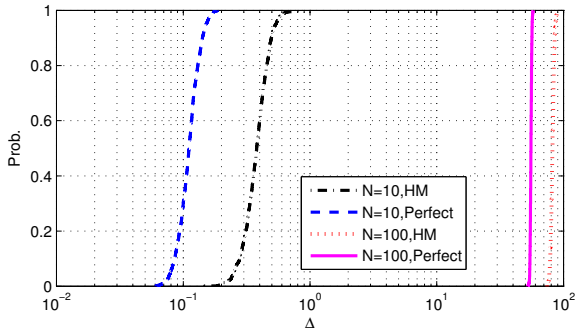

Fig. 7. Empirical CDF of Δ in AS2 with and without HM when $\beta = 0.4$, $\delta = 0.2$.

a linear growth, while MF levels out. This is due to RZF's superior suppression of inter-user interference. The analysis results of MF and RZF match well with the simulation results.

As mentioned in Section III, AS2 could be considered a very strong condition for the analysis of RZF to hold. In order to show the values of Δ in simulations, the empirical CDF of Δ is illustrated in Fig. 7 for $\beta = 0.4$ and $\delta = 0.2$. We can see from Fig. 7 that the average value of Δ is around 0.4 with HM and 0.1 without HM when $N = 10$. When $N = 100$, the average value becomes around 50 and 80 for scenarios with and without HM, respectively. Although Δ is comparable to the number of antennas, the simulation results show that the analysis in Theorem 2 is still valid even if AS2 is not satisfied.

C. Large System Effect on HM

Fig. 8 shows the accuracy of Theorem 3. The analysis results for MF match well with those of simulations. However, it is not very tight for RZF when $N = 100$. There are two reasons for this: 1) (29) is derived directly from the signal model, while the achievable rates in (5) are lower bound of rates calculated with (29); 2) the accuracy

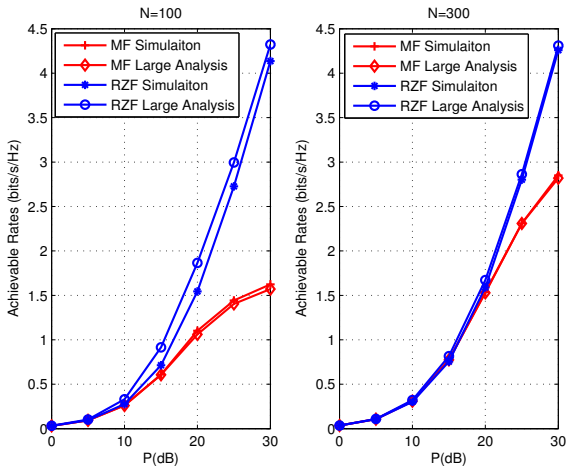

Fig. 8. Achievable rates of large system analysis compared with simulation when $K = 40$, $N = 100, 300$, $\delta = 0.4$.

of (29) relies on $\sum_{i=1}^{N} r_i^* t_i \bar{\boldsymbol{v}}_i^{\mathrm{H}} \bar{\boldsymbol{v}}_i \rightarrow \mathbb{E}\left\{r_i^* t_i \bar{\boldsymbol{v}}_i^{\mathrm{H}} \bar{\boldsymbol{v}}_i\right\}$. Only when β is small and N is large enough it converges well. As Fig. 8 shows, when $N = 300$, the curves obtained with the analytical expressions derived almost coincide with those of simulations for both MF and RZF.

Note that with the same number of users, when N gets larger, the achievable rates per user of MF improve while those of RZF remain the same. This is because RZF is able to suppress inter-user interference with lower number of antennas than MF when serving the same number of users. Since the channel gain is normalized, the extra antennas do not increase the SNR condition, so the performance of RZF remains the same. For MF, however, the extra degrees of freedom brought by additional antennas help improve the channel orthogonality between users. Therefore, the inter-user interference is reduced and the achievable rates per user of MF increase.

D. Calibration for HM

In order to compare the performance of the four calibration schemes discussed in Section IV, i.e., P-Pre-Cal, P-Post-Cal, F-Pre-Cal and F-Post-Cal, we show simulation results in terms of the empirical PDF of the achievable rates per user, which is useful to show the overall calibration performance.

As illustrated in Fig. 9, the PDF curves of the full calibration schemes (F-Post-Cal and F-Pre-Cal) is shaper than those of the partial calibration approaches (F-Post-Cal and F-Pre-Cal). Therefore, the full calibration provides a much more stable performance. When the QoS demands a small data rates, the outage probability of each user (when the achievable rates is below the data rates required) with full calibration is smaller than those with partial calibration. On the other hand, it can be seen that there is performance loss with post-precoding calibration schemes (F-Post-Cal and P-Post-Cal) compared with the pre-precoding counterparts. As described in Proposition 2 and Proposition 3, the loss comes from the additional terms $\boldsymbol{A}$ and $\boldsymbol{B}$ introduced in the precoding matrix, which decreases the value of the power factor λ_{ZF}, and the received SNR degrades consequently.

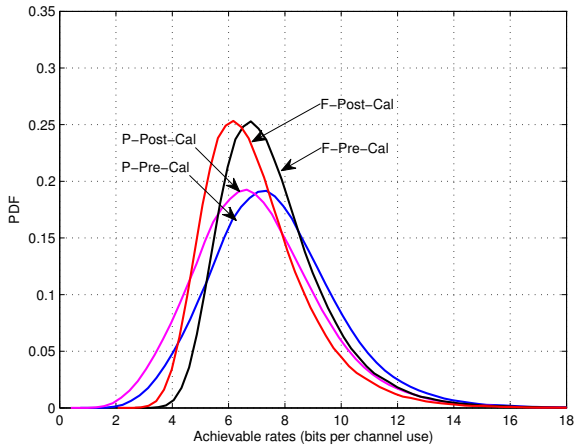

Fig. 9. Empirical PDF of Achievable rates (bits per channel use) with ZF when $N = 100$, $K = 40$, $\delta = 1.0$ and $P = 40$ dB.

VI. Conclusion

In this paper, the problem of HM in large-scale antenna systems has been studied. By handling expectations and asymptotic determinant equivalents of a series of random variables, we have derived analytical expressions that provide sets of achievable rates for both MF and RZF. In addition, the large system effect on HM has been investigated. The results show that when N and K are large, the downlink SINR of each user will converge to a constant value, which is only related to the statistics of random HM parameters. ~~It is also proved~~ that RZF is more sensitive to HM in the high SNR region compared with MF. Moreover, we have compared different HM calibration schemes and proved that Pre-Cal outperforms Post-Cal schemes. The analytical results are verified by the simulations.

Appendix A
Useful Lemmas

Lemma 3: Let $\boldsymbol{A}$ be Hermitian invertible. Then for any vector $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ and any scalar $\rho \in \mathbb{C}$, such that $\boldsymbol{A} + \rho \boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}$ is invertible, and

$$\begin{aligned}(\boldsymbol{A} + \boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})^{-1} &= \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}} \\ &= \boldsymbol{A}^{-1}\left(\boldsymbol{I} - \frac{\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}}\right) \\ &= \left(\boldsymbol{I} - \frac{\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}}{1 + \boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}}\right)\boldsymbol{A}^{-1},\end{aligned}$$

and

$$\boldsymbol{y}^{\mathrm{H}}(\boldsymbol{A}+\delta\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})^{-1}=\frac{\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}}{1+\delta\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}},$$
$$(\boldsymbol{A}+\delta\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})^{-1}\boldsymbol{x}=\frac{\boldsymbol{A}^{-1}\boldsymbol{x}}{1+\delta\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}}.$$

Proof: The proof of the first part can be found in [23], and the second part is obtained accordingly. ■

Lemma 4: Let $\boldsymbol{A}\in\mathbb{C}^{N\times N}$ and $\boldsymbol{x}\sim\mathcal{CN}(\boldsymbol{0},\frac{1}{N}\Phi_{\boldsymbol{x}})$, $\boldsymbol{y}\sim\mathcal{CN}(\boldsymbol{0},\frac{1}{N}\Phi_{\boldsymbol{y}})$. Assume that $\boldsymbol{A}$ has uniformly bounded spectral norm (with respect to N) and that $\boldsymbol{x}$ and $\boldsymbol{y}$ are mutually independent and independent of $\boldsymbol{A}$. Then for all $p\geq 1$,

a) $\mathbb{E}\left\{\left|\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}-\frac{1}{N}\mathrm{Tr}\left(\boldsymbol{A}\Phi_x\right)\right|^p\right\}=\mathcal{O}\left(N^{-\frac{p}{2}}\right)$

b) $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}-\frac{1}{N}\mathrm{Tr}\left(\boldsymbol{A}\right)\to 0$

c) $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{y}\to 0$

d) $\mathbb{E}\left\{\left|\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\right)^2-\left(\frac{1}{N}\mathrm{Tr}\left(\boldsymbol{A}\Phi_x\right)\right)^2\right|\right\}\to 0.$

Proof: This lemma can be proved by simple derivations from Lemma 4 in [17]. ■

Lemma 5: Let $\boldsymbol{T}=\mathrm{diag}\left\{t_1,\cdots,t_K\right\}$, $\boldsymbol{R}=\mathrm{diag}\left\{r_1,\cdots,r_N\right\}$ be non-negative diagonal matrices, and $\boldsymbol{V}$ is as described in (3), then when N and K grow large with fixed ratio, $\forall\boldsymbol{A}\in\mathbb{C}^{N\times N}$ and $\rho>0$,

$$\frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{A}\left[\boldsymbol{R}\boldsymbol{V}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}+\rho\boldsymbol{I}_N\right]^{-1}\right\}\to\frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{A}\boldsymbol{S}(\rho)\right\},$$

where

$$\boldsymbol{S}(\rho)=\left[\frac{1}{N}\sum_{i=1}^{K}\left\{\frac{|t_i|^2\boldsymbol{R}\Phi_i\boldsymbol{R}^{\mathrm{H}}}{1+e_{N,i}(\rho)}\right\}+\rho\boldsymbol{I}_N\right]^{-1},$$

in which

$$e_{N,i}(\rho)=\frac{1}{N}\mathrm{Tr}\left\{|t_i|^2\boldsymbol{R}\Phi_i\boldsymbol{R}^{\mathrm{H}}\left[\frac{1}{N}\sum_{j=1}^{K}\left\{\frac{|t_j|^2\boldsymbol{R}\Phi_j\boldsymbol{R}^{\mathrm{H}}}{1+e_{N,j}(\rho)}\right\}+\rho\boldsymbol{I}_N\right]^{-1}\right\},i=1,\cdots,K.$$

Proof: See Appendix I in [24]. ■

APPENDIX B
PROOF OF THEOREM 1

In order to prove Theorem 1, we need to handle λ_{MF}, $\mathbb{E}\left\{|g_{k,i}|^2\right\}$ for $i\neq k$, $|\mathbb{E}\left\{g_{k,k}\right\}|$ and $\mathrm{var}\left\{g_{k,k}\right\}$.

1) Through simple derivations, we get

$$\mathbb{E}\left\{\boldsymbol{V}\tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^{\mathrm{H}}\boldsymbol{V}^{\mathrm{H}}\right\}=\frac{1}{N}\sum_{i=1}^{K}\left|\tilde{t}_i\right|^2\boldsymbol{\Phi}_i, \tag{48}$$

Substituting (48) into (8) yields

$$\lambda_{\mathrm{MF}} = \frac{P}{\sum\limits_{i=1}^{K} \left|\tilde{t}_i\right|^2 \frac{1}{N} \mathrm{Tr}\left\{\boldsymbol{R}\boldsymbol{\Phi}_i\boldsymbol{R}^{\mathrm{H}}\right\}}. \tag{49}$$

2) Because $\boldsymbol{v}_k$ and $\boldsymbol{v_i}$ are independent for $i \neq k$, $\mathbb{E}\{\left|\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{v}_k\right|^2\} = \frac{1}{N^2}\mathrm{Tr}\left\{\boldsymbol{\Phi}_i\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\boldsymbol{T}^{\mathrm{H}}\boldsymbol{R}\right\}$. Therefore

$$\mathbb{E}\left\{\left|g_{k,i}\right|^2\right\} = \frac{\lambda_{\mathrm{MF}}}{N^2}\left|\tilde{t}_i^*\tilde{r}_k\right|^2 \mathrm{Tr}\left\{\boldsymbol{\Phi}_i\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\boldsymbol{T}^{\mathrm{H}}\boldsymbol{R}\right\}. \tag{50}$$

3) To handle $\left|\mathbb{E}\left\{g_{k,k}\right\}\right|$ and $\mathrm{var}\left\{g_{k,k}\right\}$, we first introduce $a_k = \boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{v}_k$, $b_k = \boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{v}_k$, and

$$\begin{aligned}
\mathbb{E}\left(a_k\right) &= \mathbb{E}\left\{\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{v}_k\right\} = \frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\right\}, \\
\mathbb{E}\left(b_k\right) &= \mathbb{E}\left\{\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{v}_k\right\} = \frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{T}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{\Phi}_k\right\}.
\end{aligned} \tag{51}$$

Using the Cauchy-Schwarz Inequality and Lemma 4 in Appendix, we get

$$\begin{aligned}
&\mathrm{var}\left\{\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{v}_k\right\} \\
&= \mathbb{E}\left\{\left[a_k - \mathbb{E}\left(a_k\right)\right]\left[b_k - \mathbb{E}\left(b_k\right)\right]\right\} \\
&\leqslant \sqrt{\mathrm{var}\left[a_k - \mathbb{E}\left(a_k\right)\right]\mathrm{var}\left[b_k - \mathbb{E}\left(b_k\right)\right]} \\
&= \sqrt{\mathbb{E}\left\{\left|a_k - \mathbb{E}\left(a_k\right)\right|^2\right\}\mathbb{E}\left\{\left|b_k - \mathbb{E}\left(b_k\right)\right|^2\right\}} \\
&= \mathcal{O}\left(N^{-1}\right).
\end{aligned} \tag{52}$$

Taking advantage of (51) and (52), we get

$$\begin{aligned}
\left|\mathbb{E}\left\{g_{k,k}\right\}\right| &= \left|\tilde{t}_k^*\tilde{r}_k\right|\left|\frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{R}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{\Phi}_k\right\}\right|, \\
\mathrm{var}\left\{g_{k,k}\right\} &\to \mathcal{O}\left(N^{-1}\right).
\end{aligned} \tag{53}$$

Substituting (49), (50) and (53) into (6) gives (11).

APPENDIX C
PROOF OF PROPOSITION C

Let us start from the ADE of $g_{k,k}$.

From (16), we get

$$g_{k,k} = \frac{\sqrt{\lambda_{\mathrm{RZF}}}\tilde{r}_k\tilde{t}_k^*\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{T}\boldsymbol{v}_k}{1 + \left|\tilde{t}_k\right|^2\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{R}\boldsymbol{v}_k}. \tag{54}$$

According to Lemma 4,

$$\begin{aligned}
\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{R}\boldsymbol{v}_k &\to \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{R}\boldsymbol{\Phi}_k\right), \\
\boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{T}\boldsymbol{v}_k &\to \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{T}\boldsymbol{\Phi}_k\right).
\end{aligned} \tag{55}$$

Because $\boldsymbol{Q}_{-k}$ is still random, using Lemma 5 gives,

$$\begin{aligned}
\frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{R}\boldsymbol{\Phi}_k\right) &\to \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}\boldsymbol{\Phi}_k\boldsymbol{R}^{\mathrm{H}}\boldsymbol{S}(\rho)\right), \\
\frac{1}{N}\mathrm{Tr}\left(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Q}_{-k}\boldsymbol{T}\boldsymbol{\Phi}_k\right) &\to \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{T}\boldsymbol{\Phi}_k\boldsymbol{R}^{\mathrm{H}}\boldsymbol{S}(\rho)\right),
\end{aligned} \tag{56}$$

where $\boldsymbol{S}(\rho)$ is given by (18).

Substituting (56) and (55) into (54), we have

$$g_{k,k} \to \frac{\sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_k^* \frac{1}{N}\text{Tr}\left(\boldsymbol{T}\boldsymbol{\Phi}_k\boldsymbol{R}^{\text{H}}\boldsymbol{S}(\rho)\right)}{1+\left|\tilde{t}_k\right|^2\frac{1}{N}\text{Tr}\left(\boldsymbol{R}\boldsymbol{\Phi}_k\boldsymbol{R}^{\text{H}}\boldsymbol{S}(\rho)\right)}. \tag{57}$$

Now let us handle the ADE of $g_{k,i}$ for $k \neq i$.

In (16), although $\boldsymbol{Q}_{-i}$ is independent of $\boldsymbol{v}_i$, it is correlated with $\boldsymbol{v}_k$ for $i \neq k$. Using Lemma 3, we get

$$\begin{aligned}\boldsymbol{Q}_{-i} &= \left(\boldsymbol{Q}_{-ki}^{-1} + \left|\tilde{t}_k\right|^2\boldsymbol{R}\boldsymbol{v}_k\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\right)^{-1} \\ &= \boldsymbol{Q}_{-ki} - \frac{\left|\tilde{t}_k\right|^2\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}}{1+\left|\tilde{t}_k\right|^2\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k},\end{aligned} \tag{58}$$

where $\boldsymbol{Q}_{-ki}$ is similar as $\boldsymbol{Q}_{-i}$ by removing $\boldsymbol{v}_k$, $\tilde{t}_k$ from $\boldsymbol{V}_{-i}$ and $\tilde{\boldsymbol{T}}_{-i}$, respectively.

Substituting (58) into (16) gives

$$\begin{aligned} g_{k,i} &= \frac{\sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\left[\boldsymbol{Q}_{-ki} - \frac{\left|\tilde{t}_k\right|^2\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}}{1+\left|\tilde{t}_k\right|^2\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k}\right]\boldsymbol{T}\boldsymbol{v}_k}{1+\left|\tilde{t}_i\right|^2\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-i}\boldsymbol{R}\boldsymbol{v}_i} \\ &\xrightarrow{(a)} \frac{\sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_i^*\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\left|\tilde{t}_k\right|^2\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{T}\boldsymbol{v}_k}{\left[1+\left|\tilde{t}_i\right|^2\frac{1}{N}\text{Tr}\left(\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-i}\boldsymbol{R}\boldsymbol{\Phi}_i\right)\right]\left[1+\left|\tilde{t}_k\right|^2\frac{1}{N}\text{Tr}\left(\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{\Phi}_k\right)\right]} \\ &\xrightarrow{(b)} \frac{\sqrt{\lambda_{\text{RZF}}}\tilde{r}_k\tilde{t}_i^*\left|\tilde{t}_k\right|^2\left(\boldsymbol{v}_i^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{v}_k\right)\left(\boldsymbol{v}_k^{\text{H}}\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{T}\boldsymbol{v}_k\right)}{\left[1+\left|\tilde{t}_i\right|^2\frac{1}{N}\text{Tr}\left(\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-i}\boldsymbol{R}\boldsymbol{\Phi}_i\right)\right]\left[1+\left|\tilde{t}_k\right|^2\frac{1}{N}\text{Tr}\left(\boldsymbol{R}^{\text{H}}\boldsymbol{Q}_{-ki}\boldsymbol{R}\boldsymbol{\Phi}_k\right)\right]} = \mathcal{O}\left(N^{-1}\right). \end{aligned} \tag{59}$$

where (a), (b) follow Lemma 4. Therefore we have $g_{k,i} \to 0, \forall k \neq i$.

APPENDIX D
PROOF OF LEMMA 2

Denote $\boldsymbol{\Phi}_k' = \tilde{t}_k\boldsymbol{R}\boldsymbol{\Phi}_k^{1/2}$ and rewrite $\boldsymbol{H} = [\boldsymbol{\Phi}_1'\boldsymbol{z}_1, \cdots, \boldsymbol{\Phi}_K'\boldsymbol{z}_K]$, such that $\boldsymbol{Z} = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_K] \in \mathcal{CN}(\boldsymbol{0}, \frac{1}{N}\boldsymbol{I}_N)$. $\boldsymbol{H}$ can be represented as

$$\begin{aligned}\boldsymbol{H} &= \boldsymbol{H} - \boldsymbol{\Phi}\boldsymbol{Z} + \boldsymbol{\Phi}\boldsymbol{Z} \\ &= [(\boldsymbol{\Phi}_1' - \boldsymbol{\Phi})\boldsymbol{z}_1, \cdots, (\boldsymbol{\Phi}_K' - \boldsymbol{\Phi})\boldsymbol{z}_K] + \boldsymbol{\Phi}\boldsymbol{Z} \\ &\triangleq \boldsymbol{D}_1 + \boldsymbol{D}_0.\end{aligned} \tag{60}$$

Let $d_i, \boldsymbol{r}_i, \boldsymbol{l}_i$ be the i-th singular value and the corresponding normalized right, left singular vector of $\boldsymbol{D}_0$. Multiplying $\boldsymbol{r}_i$ to the right of both sides of (60) results in

$$\begin{aligned}\boldsymbol{H}\boldsymbol{r}_i &= \boldsymbol{D}_1\boldsymbol{r}_i + \boldsymbol{D}_0\boldsymbol{r}_i \\ &= \boldsymbol{D}_1\boldsymbol{r}_i + d_i\boldsymbol{l}_i,\end{aligned} \tag{61}$$

in which

$$\begin{aligned}\|\boldsymbol{D}_1\boldsymbol{r}_i\|_{\mathrm{F}}^2 &= \boldsymbol{r}_i^{\mathrm{H}}\boldsymbol{D}_1^{\mathrm{H}}\boldsymbol{D}_1\boldsymbol{r}_i \\ &\le \mathrm{Tr}(\boldsymbol{D}_1^{\mathrm{H}}\boldsymbol{D}_1) \\ &= \mathrm{Tr}\left\{\sum_{k=1}^{K}\boldsymbol{z}_i^{\mathrm{H}}(\boldsymbol{\Phi}_k'-\boldsymbol{\Phi})^{\mathrm{H}}(\boldsymbol{\Phi}_k'-\boldsymbol{\Phi})\boldsymbol{z}_i\right\} \\ &\overset{(a)}{\to} \frac{1}{N}\sum_{k=1}^{K}\mathrm{Tr}\{(\boldsymbol{\Phi}_k'-\boldsymbol{\Phi})^{\mathrm{H}}(\boldsymbol{\Phi}_k'-\boldsymbol{\Phi})\} \\ &\overset{(b)}{\le} \mathcal{O}(N^{-1}),\end{aligned} \tag{62}$$

where (a) follows Lemma 4 and (b) is achieved using AS2.

Substituting (62) into (61) gives $\boldsymbol{H}\boldsymbol{r}_i \approx d_i\boldsymbol{l}_i$ when N is large. Therefore $\boldsymbol{H}$ has almost the same singular values as $\boldsymbol{D}_0$ and the ESD of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ can be accurately approximated by $\boldsymbol{D}_0^{\mathrm{H}}\boldsymbol{D}_0$. According to Theorem 2.39 in [20], the η transform of $\boldsymbol{D}_0$ satisfies

$$\beta = \frac{1-\eta_{\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}}(\gamma\eta_{\boldsymbol{D}_0}(\gamma))}{1-\eta_{\boldsymbol{D}_0}(\gamma)}. \tag{63}$$

According to AS1, the ESD of $\boldsymbol{D}_0$ can be obtained as in (20).

APPENDIX E
PROOF OF PROPOSITION 2

Before the proof of this proposition, two useful lemmas are introduced first.

Lemma 6 (Lemma 11, [25]): For any semi-positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{E}$ with dimension $N \times N$, the following inequality holds

$$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{E}) \ge \sum_{i=1}^{N}\lambda_i(\boldsymbol{A})\lambda_{N-i+1}(\boldsymbol{E}),$$

where $\lambda_1(\boldsymbol{A}) \ge \lambda_2(\boldsymbol{A}) \ge \cdots \ge \lambda_N(\boldsymbol{A})$ are the ordered eigenvalues of $\boldsymbol{A}$.

Lemma 7: If a is a random variable, uniformly distributed in the range of $(1-0.5\delta, 1+0.5\delta)$, with $0<\delta<2$, then $\mathbb{E}\left(\frac{1}{a}\right) \ge 1$.

Proof: It is easy to obtain the PDF of $\frac{1}{a}$ as $p(x) = \frac{1}{\delta x^2}, \frac{1}{1+0.5\delta} < x < \frac{1}{1-0.5\delta}$. Then the expected value of $\frac{1}{a}$ is given by

$$\mathbb{E}\left\{\frac{1}{a}\right\} = \int_{\frac{1}{1+0.5\delta}}^{\frac{1}{1-0.5\delta}} xp(x)\mathrm{d}x = \frac{1}{\delta}\ln\left(\frac{1+0.5\delta}{1-0.5\delta}\right).$$

Denote $f(\omega) = \omega\ln\left(\frac{\omega+0.5}{\omega-0.5}\right), \omega > 0.5$. The first and seconder order differential of $f(\omega)$ are given by, respectively

$$\begin{aligned}f'(\omega) &= \ln\left(\frac{\omega+0.5}{\omega-0.5}\right) - \frac{\omega}{\omega^2-0.25}, \\ f''(\omega) &= \frac{0.5}{(\omega^2-0.25)^2}.\end{aligned}$$

Because $f''(\omega) > 0$, $f'(\omega) \leq \lim_{\omega\to\infty} f'(\omega) = 0$, $f(\omega)$ decreases monotonously and $f(\omega) \geq \lim_{\omega\to\infty} f(\omega) = 1$. By taking advantage of this result, we have

$$\mathbb{E}\left\{\frac{1}{a}\right\} = \frac{1}{\delta}\ln\left(\frac{1+0.5\delta}{1-0.5\delta}\right) = f(1/\delta) \geq 1.$$

■

Now we are able to prove the proposition. Rewrite the two P-Precoding matrices for P-Pre-Cal and P-Post-Cal as follows

$$\begin{aligned}\boldsymbol{W}_{\text{P-Post}} &= \sqrt{\lambda_{\text{P-Post}}}\tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{T}^{-1},\\ \boldsymbol{W}_{\text{P-Pre}} &= \sqrt{\lambda_{\text{P-Pre}}}\tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}.\end{aligned} \tag{64}$$

Denote $\boldsymbol{G}_{\text{P-Post}} = \tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}^{\mathrm{H}}$ and $\boldsymbol{G}_{\text{P-Pre}} = \tilde{\boldsymbol{T}}^{-1}[\boldsymbol{V}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T}\boldsymbol{V}]^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}$. Therefore according to (40), we have

$$\begin{aligned}\lambda_{\text{P-Post}} &= \frac{P}{\mathbb{E}\mathrm{Tr}\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\}},\\ \lambda_{\text{P-Pre}} &= \frac{P}{\mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Pre}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Pre}}\}}.\end{aligned} \tag{65}$$

Note that if we replace $\boldsymbol{T}$ with $\boldsymbol{R}$ in $\boldsymbol{G}_{\text{P-Pre}}$, $\boldsymbol{G}_{\text{P-Pre}}$ with be exactly the same as $\boldsymbol{G}_{\text{P-Post}}$. Because $\boldsymbol{T}$ and $\boldsymbol{R}$ have the same distribution, we have $\mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Pre}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Pre}}\} = \mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\} = \mathbb{E}\mathrm{Tr}\{[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}]^{-1}\}$. According to Lemma 2, $\mathbb{E}\mathrm{Tr}\{[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}]^{-1}\}$ converges to some constant value given in equation (2.104) in [20] when N, K are large no matter what the exact values of $\boldsymbol{R}$ and $\tilde{\boldsymbol{T}}$ are. Therefore, to compare $\lambda_{\text{P-Post}}$ and $\lambda_{\text{P-Pre}}$ is to compare $\mathbb{E}\mathrm{Tr}\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\}$ and $\mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\}$.

Using Lemma 6, we achieve

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\} &= \mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\}\\ &\geq \mathbb{E}\left\{\sum_{i=1}^{N} a_{N-i+1}^2\lambda_i(\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}})\right\},\end{aligned}$$

where $a_j = \frac{r_j}{t_j}$ is the j-th diagonal entry of $\boldsymbol{A}$. Because $\boldsymbol{G}_{\text{P-Post}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}$ have the same non-zero eigenvalues as $\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}$, we have

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\} &\geq \mathbb{E}\left\{\sum_{i=1}^{K} a_{N-i+1}^2\lambda_i[(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}]\right\}\\ &= \sum_{i=1}^{K}\mathbb{E}\{a_{N-i+1}^2\}\mathbb{E}\lambda_i[(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}]\\ &= \mathbb{E}\{a_{N-i+1}^2\}\mathbb{E}\mathrm{Tr}[(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}]\\ &\geq |\mathbb{E}\{a_{N-i+1}\}|^2\mathbb{E}\mathrm{Tr}[(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}].\end{aligned} \tag{66}$$

According to Lemma 7, $\mathbb{E}\{a_j\} = \mathbb{E}\{t_j\}\mathbb{E}\{\frac{1}{r_j}\} \geq 1$, substituting this into (66) finally gives

$$\mathbb{E}\mathrm{Tr}\{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\boldsymbol{A}\} \geq \mathbb{E}\mathrm{Tr}[(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}] = \mathbb{E}\mathrm{Tr}\{\boldsymbol{G}_{\text{P-Post}}^{\mathrm{H}}\boldsymbol{G}_{\text{P-Post}}\},$$

and $\lambda_{\text{P-Post}} < \lambda_{\text{P-Pre}}$. Thus complete the proof.

REFERENCES

[1] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] F. Rusek, D. Persson, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear Pre-Coding Performance in Measured Very-Large MIMO Channels," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*. IEEE, Sep. 2011, pp. 1–5.

[4] W. Zhang, B. Du, C. Pan, and M. Chen, "Downlink SINR Distribution in Multiuser Large Scale Antenna Systems with Conjugate Beamforming," in *IEEE GlobeCom*, 2013.

[5] M. Guillaud, D. T. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *ISSPA*, 2005, pp. 403–406.

[6] R. C. de Lamare, "Massive MIMO Systems: Signal Processing Challenges and Future Trends," *URSI Radio Science Bulletin*, 2013.

[7] S. Han, C. Yang, G. Wang, D. Zhu, and M. Lei, "Coordinated Multi-Point Transmission Strategies for TDD Systems with Non-Ideal Channel Reciprocity," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4256–4270, Oct. 2013.

[8] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *IEEE Radio and Wireless Conference, 2003*, 2003, pp. 183–186.

[9] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative Channel Reciprocity Calibration in MIMO/TDD Systems," in *Future Network and Mobile Summit*, 2010, pp. 1–10.

[10] J. Shi, Q. Luo, and M. You, "An efficient method for enhancing TDD over the air reciprocity calibration," *2011 IEEE Wireless Communications and Networking Conference*, pp. 339–344, Mar. 2011.

[11] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical Many-Antenna Base Stations," in *Mobicom '12*. ACM Press, 2012, p. 53.

[12] R. Rogalin, O. Y. Bursalioglu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale MIMO," *2013 Information Theory and Applications Workshop (ITA)*, pp. 1–10, Feb. 2013.

[13] M. Guillaud and F. Kaltenberger, "Towards practical channel reciprocity exploitation: Relative calibration in the presence of frequency offset," *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2525–2530, Apr. 2013.

[14] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.

[15] L. Su, C. Yang, G. Wang, and M. Lei, "Retrieving Channel Reciprocity for Coordinated Multi-Point Transmission with Joint Processing," *IEEE Transactions on Communications*, vol. 62, no. 5, pp. 1541–1553, May 2014.

[16] F. Huang, Y. Wang, J. Geng, and D. Yang, "Antenna mismatch and calibration problem in coordinated multi-point transmission system," *IET Communications*, vol. 6, no. 3, p. 289, 2012.

[17] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[18] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[19] C. Peel, B. Hochwald, and A. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication‐Part I: Channel Inversion and Regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[20] A. M. Tulino and S. Verd, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004, vol. 1.

[21] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.

[22] I. R. I.S. Gradshteyn, *Tables of Integrals, Series, and Products*, 7th ed. Elsevier, 2007.

[23] X. Zhang, *Matrix Analysis and Applications*. Beijing: Tsinghua University Press, 2004.

[24] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[25] D. Palomar, J. Cioffi, and M. Lagunas, "Joint TX-RX Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.