# Achievable Downlink Rates of MRC and ZF Precoders in Massive MIMO With Uplink and Downlink Pilot Contamination

Amin Khansefid, *Student Member, IEEE*, and Hlaing Minn, *Senior Member, IEEE*

***Abstract*—Massive multi-input multi-output (MIMO) systems require channel state information (CSI) at base stations. In multi-cell environments, adjacent cells using the same spectrum cause a pilot contamination problem and degrade CSI quality. Under such practical environments, it is important to assess achievable rate of the system for a proper system design. There exist achievable rate bounds of uplink (UL) systems in the literature, but general closed-form rate bounds for downlink (DL) systems are not available yet. This paper studies a practical massive MIMO DL system with or without pilot-aided coherent detection under the scenario of pilot contamination, and derives closed-form approximate achievable DL rate expressions for both maximum ratio combining (MRC) and zero forcing (ZF) precoders. For the case with DL pilot, a corresponding DL effective channel estimator is also developed. Numerical simulation results corroborate accuracy of the closed-form rate expressions and they also show performance characteristics of pilot-contaminated massive MIMO DL systems with MRC and ZF precoders.**

***Index Terms*—Massive MIMO, downlink achievable rate, pilot contamination, channel estimation, imperfect CSI.**

## I. Introduction

MASSIVE MIMO is a promising technology for next generation wireless systems [1], [2] due to its advantages in enhancing spectrum and energy efficiency [3]–[6]. Marzetta shows in [5] that with the use of many antenna elements at the base station (BS), several users can communicate with BS simultaneously on the same spectrum and time. With the use of CSI at BS and by means of asymptotic orthogonality of independent random vectors of large size, the interference among the simultaneous users is canceled and at the same time small-scale channel fading effects are averaged out. These properties and potentials of massive MIMO have motivated substantial further research work [7]–[17].

The very large signal vector dimension at a massive MIMO BS favors low complexity algorithms, e.g., linear detectors based on MRC, ZF or minimum mean squared error (MMSE) [6]–[8] for the UL and linear pre-coding/beamforming such as MRC and ZF [6], [9], [10] for the DL. Nonlinear detectors [18]–[20] and nonlinear pre-coders such as dirty-paper coding and vector perturbation [21] offer better performance but with higher complexity. However, as the number of BS antennas ($M$) becomes large, linear pre-coding and detection techniques such as MRC and ZF become near optimal [6].

In assessing performance of various massive MIMO scenarios, achievable rate expressions or their bounds serve as important metrics. In the UL scenario, [7] derives rate bounds for MRC, ZF and MMSE detectors and analyzes power efficiency mainly considering a single cell system. For multicell systems, UL rate bounds are derived for a ZF receiver in [22] and for MRC and ZF receivers in [12]. The UL rate bounds of massive MIMO systems in Ricean fading channels are investigated in [13] using MRC and ZF receivers under perfect and imperfect CSI. The accuracies of rate bounds for single cell massive MIMO UL are evaluated in [23] by deriving upper and lower bounds.

In the DL scenario, the existing works on achievable rate bounds [8]–[11] consider a particular receiver structure which treats the channel mean as its effective channel for data detection. Single cell DL systems are addressed in [9], [11]. Rate derivation for both UL and DL of multicell systems in [8] is based on asymptotic rate analysis when the numbers of base station antennas and users go to infinity, but our analysis is for a finite number of base station antennas. Another difference is, for downlink transmission we consider both scenarios of with and without downlink pilot. Furthermore, in downlink precoding, [8] uses the average total transmit power normalization, while we use per-user downlink power normalization (which offers user fairness). Multicell DL scenarios with total transmit power normalized precoder are considered in [10] and it does not consider downlink pilot transmission and downlink effective channel estimation. In addition, there is no closed form rate expression, except for a simple scenario of single user in each cell with MRC precoding. In our paper, we also consider rate analysis without downlink pilot and with the assumption of average effective channel gain available at the user. But we derive the closed form rate expressions for this case with MRC and ZF precoders. In addition, we also derive closed form rate expressions for systems with downlink pilot for both MRC and ZF precoders. [24] considers single cell DL transmission for MRC and ZF precoding with average transmit power normalization and DL pilot.

An important issue in massive MIMO as mentioned in [5] is pilot contamination. The issue arises in multicell environments when users in different cells transmit the same pilot set by which the estimated channel is corrupted by channels of users in other cells. To address this issue, [10] proposes an MMSE precoding scheme while a review of available techniques is

Manuscript received December 17, 2014; revised May 9, 2015 and August 3, 2015; accepted September 18, 2015. Date of publication September 28, 2015; date of current version December 15, 2015. The associate editor coordinating the review of this paper and approving it for publication was Z. Zhang.

The authors are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: amin.khansefid@utdallas.edu; hlaing.minn@utdallas.edu).

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

Digital Object Identifier 10.1109/TCOMM.2015.2482965

reported in [4]. Pilot contamination affects both UL and DL performance, and hence achievable rate expressions under pilot contamination are important for providing guidance in system design. Deriving closed-form rate expressions for DL is more challenging since both UL pilot contamination and DL pilot contamination plague the DL system. To the best of our knowledge, simple closed-form rate expressions for massive MIMO DL with an arbitrary number of BS antennas under pilot contamination are not available in the literature.

In this paper, we study massive MIMO DL transmission with MRC or ZF precoding in a multicell scenario with pilot contamination effect. As a common scenario of massive MIMO, we assume time-division duplexing (TDD) which offers channel reciprocity and saves pilot overhead. We consider two scenarios, namely without DL pilot and with DL pilot. For the latter scenario, we also develop a corresponding linear MMSE (LMMSE) estimator for the DL effective channel. Our main contribution is the development of closed-form approximating expressions for the achievable rates of the considered multicell massive MIMO DL system with a finite number of BS antennas under imperfect CSI and pilot contamination.

The paper is organized as follows. In Section II, we describe the system model, UL channel estimation, and DL transmission. Then in Section III, we drive DL rate bounds for systems without DL pilot. In Section IV, DL pilot transmission and DL effective channel estimation are presented, and we derive the corresponding DL rate bounds and their approximations. Section V discusses numerical results and Section VI provides conclusions.

*Notation:* Vectors (matrices) are denoted by bold face small (big) letters. The superscripts $T$, $H$, and $*$ stand for the transpose, conjugate transpose and conjugate of a matrix or vector, respectively. $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. $\mathbb{E}\{\cdot\}$, $\|.\|$, $\Re\{.\}$ and $\Im\{.\}$ denote expectation, Euclidean norm, real part and imaginary part, respectively. $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ means $\mathbf{n}$ has a probability density function (pdf) of the zero-mean complex Gaussian vector with covariance matrix $\boldsymbol{C}$. $\mathcal{N}(0, \sigma^2)$ denotes a zero-mean Gaussian pdf with variance $\sigma^2$. The $\delta_{ki}$ and $\Gamma(\cdot)$ denote the Kronecker delta and the Gamma function, respectively. $\mathrm{diag}\{a_1, \cdots, a_K\}$ means a diagonal matrix with diagonal elements $\{a_1, \cdots, a_K\}$.

## II. System Model

We consider a single-carrier multi-cell system with $L$ cells where each cell has a BS with co-located $M$ antenna elements and $K$ randomly located single antenna users. We assume channel gains are quasi-static within a frame, and channels of different users and antennas are independent. Each user experiences a frequency-flat fading channel. Let $g_{ilkm}$ represent the channel gain from user $k$ in cell $l$ to the antenna $m$ of the BS in cell $i$ (or simply, of BS $i$). We can write $g_{ilkm} = \sqrt{\beta_{ilk}} h_{ilkm}$ where $h_{ilkm} \sim \mathcal{CN}(0, 1)$ is the small scale fading coefficient and $\beta_{ilk}$ is the large scale fading power gain (the same for all channels between user $k$ in cell $l$ and antennas at BS $i$, thus the index $m$ is omitted). The overall $M \times K$ lowpass equivalent UL channel matrix is denoted by $\mathbf{G}_{il}$, whose $k$th column $\boldsymbol{g}_{ilk}$ represents the gains of the channels from user $k$ in cell $l$ to BS $i$ and has a $\mathcal{CN}(\mathbf{0}, \beta_{ilk}\mathbf{I}_M)$ distribution. In the same way as [8], [10], [12], we assume $\{\beta_{ilk}, i, l = 1, \cdots, L, k = 1, \cdots, K\}$ are known at all BSs, and this assumption is justified since $\{\beta_{ilk}\}$ change very slowly compared with $\{h_{ilkm}\}$ and they can be reliably estimated. The same as in the literature, we treat $\{\beta_{ilk}\}$ as deterministic in our channel estimation.

The above assumption of each base station knowing individual $\{\beta_{ilk}\}$ of other cells is in fact not necessary and the required information for the MMSE channel estimator can be easily obtained as in [25] which shows almost the same estimation performance as the MMSE with perfect knowledge of $\{\beta_{ilk}\}$. Thus, information sharing between base stations such as in network MIMO is not necessary. On the other hand, comparison of massive MIMO and network MIMO, or combination of massive MIMO and network MIMO under different levels of information sharing between base stations is an interesting and separate research topic.

We also note that the rate analysis in this paper is developed under equal power allocation among the users. If an appropriate user power allocation strategy is applied based on the knowledge of large scale fading coefficients, the achievable rate would be higher. The problem of power allocation for single cell massive MIMO is studied in [26] while the problem for the considered system is a challenging issue which needs a separate investigation.

### A. Uplink Training

Each user in UL mode transmits $\tau_{\mathrm{u}}$ pilot symbols and then each BS estimates the channels of its users. We assume users of different cells transmit the same set of pilots at the same time (a typical scenario in massive MIMO) and the pilot reuse factor is one. The pilot sequences of $K$ users are represented by a $K \times \tau_{\mathrm{u}}$ matrix $\boldsymbol{\Phi}_{\mathrm{u}}^H$ with orthogonality property $\boldsymbol{\Phi}_{\mathrm{u}}^H \boldsymbol{\Phi}_{\mathrm{u}} = \tau_{\mathrm{u}} \mathbf{I}_K$, $\tau_{\mathrm{u}} \geq K$. The received pilot symbols at BS $i$ are represented by an $M \times \tau_{\mathrm{u}}$ matrix $\mathbf{Y}_i$ as

$$\mathbf{Y}_i = \sum_{l=1}^{L} \sqrt{q_{\mathrm{u}}} \mathbf{G}_{il} \boldsymbol{\Phi}_{\mathrm{u}}^H + \mathbf{N}_i \tag{1}$$

where $q_{\mathrm{u}}$ is the UL pilot power and $\mathbf{N}_i$ is an $M \times \tau_{\mathrm{u}}$ noise matrix with independent and identically distributed (i.i.d.) elements of $\mathcal{CN}(0, 1)$. If $\mathbf{g}_{ilk}$ is to be estimated at BS $i$, a sufficient statistic [27] is

$$\mathbf{z}_{ik} \triangleq \frac{1}{\tau_{\mathrm{u}} \sqrt{q_{\mathrm{u}}}} \mathbf{Y}_i \boldsymbol{\phi}_{\mathrm{u}k} = \sum_{l=1}^{L} \mathbf{g}_{ilk} + \mathcal{CN}\left(\mathbf{0}, \frac{1}{\tau_{\mathrm{u}} q_{\mathrm{u}}} \mathbf{I}_M\right) \tag{2}$$

where $\boldsymbol{\phi}_{\mathrm{u}k}$ denotes the $k$th column of $\boldsymbol{\Phi}_{\mathrm{u}}$. The MMSE estimation is given by [27]

$$\hat{\mathbf{g}}_{ilk} \triangleq b_{ilk} \mathbf{z}_{ik} = b_{ilk} \sum_{l=1}^{L} \mathbf{g}_{ilk} + \mathcal{CN}\left(\mathbf{0}, \frac{b_{ilk}^2}{\tau_{\mathrm{u}} q_{\mathrm{u}}} \mathbf{I}_M\right) \tag{3}$$

where $b_{ilk} \triangleq \frac{\beta_{ilk}}{\sum_{l'=1}^{L} \beta_{il'k} + \frac{1}{\tau_{\mathrm{u}} q_{\mathrm{u}}}}$. We can write $\mathbf{g}_{ilk} = \hat{\mathbf{g}}_{ilk} + \boldsymbol{\varepsilon}_{ilk}$, where the channel estimate $\hat{\mathbf{g}}_{ilk}$ and the estimation error $\boldsymbol{\varepsilon}_{ilk}$ are

independent due to the MMSE property and their Gaussianity. We have $\hat{\mathbf{g}}_{ilk} \sim \mathcal{CN}(\mathbf{0}, \sigma^2_{\hat{\mathbf{g}}_{ilk}} \mathbf{I}_M)$ and $\boldsymbol{\varepsilon}_{ilk} \sim \mathcal{CN}(\mathbf{0}, \sigma^2_{\boldsymbol{\varepsilon}_{ilk}} \mathbf{I}_M)$ with

$$\sigma^2_{\hat{\mathbf{g}}_{ilk}} \triangleq \frac{\beta^2_{ilk}}{\sum_{l'=1}^{L} \beta_{il'k} + \frac{1}{\tau_u q_u}} \tag{4}$$

$$\sigma^2_{\boldsymbol{\varepsilon}_{ilk}} \triangleq \frac{\beta_{ilk} \left( \sum_{l'=1, l' \neq l}^{L} \beta_{il'k} + \frac{1}{\tau_u q_u} \right)}{\sum_{l'=1}^{L} \beta_{il'k} + \frac{1}{\tau_u q_u}}. \tag{5}$$

Increasing $q_u$ decreases $\sigma^2_{\boldsymbol{\varepsilon}_{ilk}}$ but enlarges $\sigma^2_{\hat{\mathbf{g}}_{ilk}}$ while a larger $\beta_{ilk}$ or $\beta_{il'k}$ increases $\sigma^2_{\boldsymbol{\varepsilon}_{ilk}}$. Note that $\sigma^2_{\hat{\mathbf{g}}_{ilk}} + \sigma^2_{\boldsymbol{\varepsilon}_{ilk}} = \beta_{ilk}$. From (5), we see that due to pilot contamination, increasing UL pilot power cannot reduce the Mean Square Error (MSE) to zero. From (3), we can see

$$\hat{\mathbf{g}}_{ilk} = \frac{\beta_{ilk}}{\beta_{iik}} \hat{\mathbf{g}}_{iik} \tag{6}$$

and as $\boldsymbol{\varepsilon}_{ilk}$ is independent of $\hat{\mathbf{g}}_{ilk}$, it is also independent of any function of $\hat{\mathbf{g}}_{ilk}$, such as $\hat{\mathbf{g}}_{iik}$. Note that BS $i$ only needs to estimate $\{\mathbf{g}_{iik}, k = 1, \cdots, K\}$, and the role of (6) is to facilitate the rate analysis by decomposing the inter-cell interference into a correlated term (due to $\hat{\mathbf{g}}_{ilk}$) and an uncorrelated term (due to $\boldsymbol{\varepsilon}_{ilk}$) with respect to the desired user's channel estimate.

### B. Downlink Transmission

BS $l$ linearly precodes data of its users through an $M \times K$ precoding matrix $\mathbf{W}_l$ and transmits them. The $k$th column of $\mathbf{W}_l$ is denoted by $\mathbf{w}_{lk}$ with $\|\mathbf{w}_{lk}\| = 1$. The received signal at user $k$ in cell $i$ is given by

$$r_{ik} = \sqrt{p} \sum_{l=1}^{L} \mathbf{g}_{lik}^T \mathbf{W}_l \mathbf{s}_l + n_{ik} \tag{7}$$

where $\mathbf{s}_l = [s_{l1}, \cdots, s_{lK}]^T$ is the complex data vector of users in cell $l$ with $\mathbb{E}\{\mathbf{s}_l\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{s}_l \mathbf{s}_l^H\} = \mathbf{I}_K$, $n_{ik} \sim \mathcal{CN}(0, 1)$ is the noise term, and $p$ is the transmit data power which also represents the transmit signal-to-noise power ratio (Tx SNR) of each user. So the total transmit power at BS is $Kp$. We expand (7) as

$$r_{ik} = \sqrt{p} \mathbf{g}_{iik}^T \mathbf{w}_{ik} s_{ik} + \underbrace{\sum_{j=1, j \neq k}^{K} \sqrt{p} \mathbf{g}_{iik}^T \mathbf{w}_{ij} s_{ij}}_{\text{intra-cell interference}} + \underbrace{\sum_{l=1, l \neq i}^{L} \sum_{j=1}^{K} \sqrt{p} \mathbf{g}_{lik}^T \mathbf{w}_{lj} s_{lj}}_{\text{inter-cell interference}} + n_{ik}. \tag{8}$$

We define $\mathbf{w}_{ik} \triangleq \frac{\hat{\mathbf{g}}_{iik}^*}{\|\hat{\mathbf{g}}_{iik}\|}$ for MRC precoding and $\mathbf{w}_{ik} \triangleq \frac{\mathbf{a}_{ik}^*}{\|\mathbf{a}_{ik}\|}$ for ZF precoding where $\mathbf{a}_{ik}$ is the $k$th column of $\mathbf{A}_i \triangleq \hat{\mathbf{G}}_{ii} (\hat{\mathbf{G}}_{ii}^H \hat{\mathbf{G}}_{ii})^{-1}$ so that $\mathbf{a}_{ik}^H \hat{\mathbf{g}}_{iij} = \delta_{jk}$. The received signals at user $k$ in cell $i$ with MRC and ZF precoding are respectively given by

$$r_{ik}^{\text{mrc}} = \sqrt{p} \frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|} \mathbf{g}_{iik} s_{ik} + \sum_{j=1, j \neq k}^{K} \sqrt{p} \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|} \mathbf{g}_{iik} s_{ij} + \sum_{l=1, l \neq i}^{L} \sum_{j=1}^{K} \sqrt{p} \frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|} \mathbf{g}_{lik} s_{lj} + n_{ik} \tag{9}$$

$$r_{ik}^{\text{zf}} = \sqrt{p} \frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|} \mathbf{g}_{iik} s_{ik} + \sum_{j=1, j \neq k}^{K} \sqrt{p} \frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|} \mathbf{g}_{iik} s_{ij} + \sum_{l=1, l \neq i}^{L} \sum_{j=1}^{K} \frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|} \mathbf{g}_{lik} s_{lj} + n_{ik}. \tag{10}$$

## III. DL Rate Analysis Without DL Pilot

When the users do not estimate their DL channels, following the signal model given in (8), the achievable rate lower bound of user $k$ in cell $i$ can be given by [10, Theorem 1]

$$R_{ik,\text{np}} = \lambda_{\text{np}} \log_2 \left( 1 + \frac{\left| \mathbb{E}\left\{ \sqrt{p} \mathbf{g}_{iik}^T \mathbf{w}_{ik} \right\} \right|^2}{\mathbb{E}\left\{ \left| \tilde{I}_{ik} \right|^2 \right\}} \right) \tag{11}$$

where $\lambda_{\text{np}}$ accounts for pilot overhead (will be discussed in Remark 7), the expectation is over all random variables, and

$$\tilde{I}_{ik} = \sqrt{p} \left( \mathbf{g}_{iik}^T \mathbf{w}_{ik} - \mathbb{E}\left\{ \mathbf{g}_{iik}^T \mathbf{w}_{ik} \right\} \right) s_{ik} + \sum_{j=1, j \neq k}^{K} \sqrt{p} \mathbf{g}_{iik}^T \mathbf{w}_{ij} s_{ij} + \sum_{l=1, l \neq i}^{L} \sum_{j=1}^{K} \sqrt{p} \mathbf{g}_{lik}^T \mathbf{w}_{lj} s_{lj} + n_{ik}. \tag{12}$$

Here, users treat the mean effective channel gain as the channel knowledge for data detection. Note that [10] does not provide a general closed form expression for (11). In the following, we drive closed form expressions for DL rate bounds.

*Theorem 1:* The closed form of the rate bound in (11) with MRC precoding reads as

$$R_{ik,\text{np}}^{\text{mrc}} = \lambda_{\text{np}} \log_2 \left( 1 + \frac{C_M^2 \sigma^2_{\hat{\mathbf{g}}_{iik}}}{\mathcal{I}_{ik,\text{np}}^{\text{mrc}}} \right) \tag{13}$$

where $\mathcal{I}_{ik,\text{np}}^{\text{mrc}} \triangleq V_M \sigma^2_{\hat{\mathbf{g}}_{iik}} + \sigma^2_{\boldsymbol{\varepsilon}_{iik}} + (K-1)\beta_{iik} + K \sum_{l=1, l \neq i}^{L} \beta_{lik} + \frac{1}{p} + (M-1) \sum_{l=1, l \neq i}^{L} \sigma^2_{\hat{\mathbf{g}}_{lik}} = (V_M - 1)\sigma^2_{\hat{\mathbf{g}}_{iik}} + K \sum_{l=1}^{L} \beta_{lik} + \frac{1}{p} + (M-1) \sum_{l=1, l \neq i}^{L} \sigma^2_{\hat{\mathbf{g}}_{lik}}$, and $C_M$ and $V_M$ are given by

$$C_M \triangleq \frac{\Gamma(M + 1/2)}{\Gamma(M)} \tag{14}$$

$$V_M \triangleq M - C_M^2. \tag{15}$$

For a moderate or large $M$, $C_M \approx \sqrt{M}$ and $V_M \approx 0.25$ [10].

*Proof:* See Appendix A. ■

*Theorem 2:* The closed form of the rate bound in (11) with ZF precoding reads as

$$R_{ik,\mathrm{np}}^{\mathrm{zf}} = \lambda_{\mathrm{np}} \log_2 \left(1 + \frac{C_{M-K+1}^2 \sigma_{\hat{\mathbf{g}}_{iik}}^2}{\mathcal{I}_{ik,\mathrm{np}}^{\mathrm{zf}}}\right) \tag{16}$$

where $\mathcal{I}_{ik,\mathrm{np}}^{\mathrm{zf}} \triangleq V_{M-K+1}\sigma_{\hat{\mathbf{g}}_{iik}}^2 + K\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + (M-K+1)\sum_{l=1,l\neq i}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \frac{1}{p}$ while $C_{M-K+1}$ and $V_{M-K+1}$ are given in (14) and (15).

*Proof:* See Appendix B. ■

*Remark 1:* As $M \to \infty$ with a fixed $K$, the rates in (13) and (16) approach the same rate as

$$R_{ik,\mathrm{np}}^{\mathrm{mrc}},\, R_{ik,\mathrm{np}}^{\mathrm{zf}} \to \lambda_{\mathrm{np}} \underbrace{\log_2 \left(1 + \frac{\sigma_{\hat{\mathbf{g}}_{iik}}^2}{\sum_{l=1,l\neq i}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2}\right)}_{\triangleq R_{ik}^{\infty}}. \tag{17}$$

## IV. DL Rate Analysis With DL Pilot

From (9) for MRC, the effective DL channel gain is $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik}$, but computing interference power conditioned on the estimate of this complex channel gain is analytically intractable. To overcome this, by replacing $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$, we rewrite the effective channel gain as $\|\hat{\mathbf{g}}_{iik}\| + \frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}$, and with observation that the power of the second term is negligible compared to the power of the first one, we propose to treat $\|\hat{\mathbf{g}}_{iik}\|$ as the effective channel gain and treat the $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}$ as part of the interference. This approach is well justified since the power ratio of these two terms is $\frac{M\sigma_{\hat{\mathbf{g}}_{iik}}^2}{\sigma_{\boldsymbol{\varepsilon}_{iik}}^2} = \frac{M\beta_{iik}}{\sum_{l=1,l\neq i}^{L}\beta_{ilk} + \frac{1}{\tau_{\mathrm{u}}q_{\mathrm{u}}}}$ which is large[1] especially for a large $M$. The same reasoning holds for the desired signal term in (10) for ZF precoding. So, we define the effective channel for MRC and ZF precoding scenario as $\alpha_{ik}^{\mathrm{mrc}} \triangleq \|\hat{\mathbf{g}}_{iik}\|$ and $\alpha_{ik}^{\mathrm{zf}} \triangleq \frac{1}{\|\mathbf{a}_{ik}\|}$, respectively.

In systems with DL pilot, each user estimates the effective channel $\alpha_{ik}$ and performs coherent detection. With the channel estimate $\hat{\alpha}_{ik}$ and its corresponding error $\nu_{ik}$ (see Lemma 3 and 6 for their statistics), we can write $\alpha_{ik} = \hat{\alpha}_{ik} + \nu_{ik}$. Then, the received signal at user $k$ in cell $i$ with MRC precoding is

$$r_{ik}^{\mathrm{mrc}} = \sqrt{p}\,\hat{\alpha}_{ik}^{\mathrm{mrc}}\, s_{ik} + I_{ik}^{\mathrm{mrc}} \tag{18}$$

where the interference plus noise term is

$$\begin{aligned} I_{ik}^{\mathrm{mrc}} \triangleq{}& \sqrt{p}\nu_{ik}^{\mathrm{mrc}} s_{ik} + \sqrt{p}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}s_{ik} \\ &+ \sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}s_{ij} \\ &+ \sum_{l=1,l\neq i}^{L}\sum_{j=1}^{K}\sqrt{p}\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}s_{lj} + n_{ik}. \end{aligned} \tag{19}$$

The corresponding received signal for ZF precoding is

$$r_{ik}^{\mathrm{zf}} = \sqrt{p}\,\hat{\alpha}_{ik}^{\mathrm{zf}}\, s_{ik} + I_{ik}^{\mathrm{zf}} \tag{20}$$

[1]e.g., with $\beta_{iik} = 1$, $\beta_{ilk} = 0.1$ $(l \neq i)$, $q_{\mathrm{u}} = 10$, $L = 7$, and $M$ in the range of 30–100, the ratio is in the range of 50–164.

where

$$\begin{aligned} I_{ik}^{\mathrm{zf}} \triangleq{}& \sqrt{p}\nu_{ik}^{\mathrm{zf}} s_{ik} + \sqrt{p}\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik}s_{ik} + \sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\mathbf{g}_{iik}s_{ij} \\ &+ \sum_{l=1,l\neq i}^{L}\sum_{j=1}^{K}\sqrt{p}\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\mathbf{g}_{lik}s_{lj} + n_{ik}. \end{aligned} \tag{21}$$

### A. Downlink Channel Estimation

To estimate the effective DL channel for coherent detection, BSs in different cells transmit the same set of orthogonal pilot sequences over $\tau_{\mathrm{d}}$ symbol intervals. The pilot set for $K$ users is collected in a $K \times \tau_{\mathrm{d}}$ matrix $\boldsymbol{\Phi}_{\mathrm{d}}^H$ which has the orthogonal property $\boldsymbol{\Phi}_{\mathrm{d}}^H\boldsymbol{\Phi}_{\mathrm{d}} = \tau_{\mathrm{d}}\mathbf{I}_K$, $\tau_{\mathrm{d}} \geq K$. The received pilot vector at user $k$ in cell $i$ is

$$\mathbf{y}_{ik}^T = \sqrt{q_{\mathrm{d}}}\sum_{l=1}^{L}\mathbf{g}_{lik}^T\mathbf{W}_l\boldsymbol{\Phi}_{\mathrm{d}}^H + \mathbf{n}_{ik}^T \tag{22}$$

where $q_{\mathrm{d}}$ is the DL pilot power. For estimation of $\alpha_{ik}$, a sufficient statistic is given by

$$z_{ik} \triangleq \frac{\mathbf{y}_{ik}^T\boldsymbol{\phi}_{\mathrm{d}k}}{\tau_{\mathrm{d}}\sqrt{q_{\mathrm{d}}}} = \mathbf{g}_{iik}^T\mathbf{w}_{ik} + \sum_{l=1,l\neq i}^{L}\mathbf{g}_{lik}^T\mathbf{w}_{lk} + \mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}\right) \tag{23}$$

where $\boldsymbol{\phi}_{\mathrm{d}k}$ is the $k$th column of $\boldsymbol{\Phi}_{\mathrm{d}}$. Using (23), we develop effective DL channel estimators.

*1) System with MRC Precoding:* In this case, the signal in (23) is rewritten as

$$z_{ik}^{\mathrm{mrc}} = \alpha_{ik}^{\mathrm{mrc}} + w_{ik}^{\mathrm{mrc}} \tag{24}$$

where $\alpha_{ik}^{\mathrm{mrc}} = \|\hat{\mathbf{g}}_{iik}\|$ is the effective channel gain and

$$w_{ik}^{\mathrm{mrc}} \triangleq \frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik} + \sum_{l=1,l\neq i}^{L}\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik} + \mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}\right) \tag{25}$$

is the interference plus noise term. Note that pilot contamination causes correlation between the precoding matrices of the interfering BSs and their channels to the desired user, which in turn results in higher interference power in the real domain than the imaginary domain. This non-circular symmetric characteristics of inter-cell interference yields different statistics between the real and imaginary parts of $w_{ik}^{\mathrm{mrc}}$. The following lemma describes more details.

*Lemma 1:* The real part $w_{ik}^{\mathrm{mrc,re}}$ and the imaginary part $w_{ik}^{\mathrm{mrc,im}}$ of $w_{ik}^{\mathrm{mrc}}$ are independent and

$$\mathbb{E}\left\{w_{ik}^{\mathrm{mrc,re}}\right\} = C_M\sum_{l=1,l\neq i}^{L}\sqrt{\sigma_{\hat{\mathbf{g}}_{lik}}^2} \tag{26}$$

$$\mathbb{E}\left\{w_{ik}^{\mathrm{mrc,im}}\right\} = 0 \tag{27}$$

$$\mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\} = \frac{1}{2}\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + \frac{1}{2\tau_{\mathrm{d}}q_{\mathrm{d}}} + V_M\sum_{l=1,l\neq i}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2 \quad (28)$$

$$\mathrm{var}\left\{w_{ik}^{\mathrm{mrc,im}}\right\} = \frac{1}{2}\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + \frac{1}{2\tau_{\mathrm{d}}q_{\mathrm{d}}}. \quad (29)$$

Furthermore, $w_{ik}^{\mathrm{mrc,im}}$ is Gaussian and $w_{ik}^{\mathrm{mrc,re}}$ for a large value of $M$ approaches Gaussian. We can rewrite (28) as $\mathrm{var}\{w_{ik}^{\mathrm{mrc,re}}\} = \frac{1}{2}(\sigma_{\boldsymbol{\varepsilon}_{iik}}^2 + \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}) + (\frac{1}{2} - V_M)\sum_{l=1,l\neq i}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + V_M\sum_{l=1,l\neq i}^{L}\beta_{lik}$.

*Proof:* See Appendix C. ■

Lemma 1 implies that the interference power $\mathrm{var}\{w_{ik}^{\mathrm{mrc,re}}\}$ decreases with the increase of $q_{\mathrm{u}}$ or $q_{\mathrm{d}}$ or decrease of $\beta_{ilk}$. The next Lemma describes some statistics of $\alpha_{ik}^{\mathrm{mrc}}$.

*Lemma 2:* Mean and variance of the DL effective channel gain $\alpha_{ik}^{\mathrm{mrc}}$ are respectively given by

$$\mathbb{E}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} = \mathbb{E}\left\{\left\|\hat{\mathbf{g}}_{iik}\right\|\right\} = C_M\sqrt{\sigma_{\hat{\mathbf{g}}_{iik}}^2} \quad (30)$$

$$\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} = V_M\sigma_{\hat{\mathbf{g}}_{iik}}^2 \quad (31)$$

and for a large value of $M$, $\alpha_{ik}^{\mathrm{mrc}}$ approaches Gaussian.

*Proof:* See Appendix D. ■

From Lemma 2, following the approximation of $C_M \approx \sqrt{M}$ and $V_M \approx 0.25$, we can see $\mathbb{E}\{\alpha_{ik}^{\mathrm{mrc}}\}$ increases with $\sqrt{M}$ while $\mathrm{var}\{\alpha_{ik}^{\mathrm{mrc}}\}$ is approximately fixed. So for a large $M$, $\frac{\mathbb{E}\{\alpha_{ik}^{\mathrm{mrc}}\}}{\sqrt{\mathrm{var}\{\alpha_{ik}^{\mathrm{mrc}}\}}} \approx 2\sqrt{M}$ is large and thus, relative fluctuations of $\alpha_{ik}^{\mathrm{mrc}}$ are quite small. In addition, both mean and variance of $\alpha_{ik}^{\mathrm{mrc}}$ increase with $q_{\mathrm{u}}$.

By exploiting Lemma 1, we can estimate $\alpha_{ik}^{\mathrm{mrc}}$ from the real part $z_{ik}^{\mathrm{mrc,re}}$ of $z_{ik}^{\mathrm{mrc}}$ where

$$z_{ik}^{\mathrm{mrc,re}} = \alpha_{ik}^{\mathrm{mrc}} + w_{ik}^{\mathrm{mrc,re}}. \quad (32)$$

*Lemma 3:* For a system with MRC precoding, the LMMSE estimator $\hat{\alpha}_{ik}$ based on (32) and its MSE $\eta_{ik}^{\mathrm{mrc}}$ are respectively given by

$$\hat{\alpha}_{ik}^{\mathrm{mrc}} = c_{ik}^{\mathrm{mrc}} + f_{ik}^{\mathrm{mrc}}z_{ik}^{\mathrm{mrc,re}} \quad (33)$$

$$\eta_{ik}^{\mathrm{mrc}} \triangleq \mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}} - \alpha_{ik}^{\mathrm{mrc}}\right|^2\right\} = \frac{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\}\cdot\mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}}{\mathrm{var}\{\alpha_{ik}^{\mathrm{mrc}}\} + \mathrm{var}\{w_{ik}^{\mathrm{mrc,re}}\}} \quad (34)$$

where

$$c_{ik}^{\mathrm{mrc}} \triangleq \frac{\mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}\mathbb{E}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} - \mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\}\mathbb{E}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}} \quad (35)$$

$$f_{ik}^{\mathrm{mrc}} \triangleq \frac{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}}. \quad (36)$$

Denoting the estimation error by $\nu_{ik}^{\mathrm{mrc}}$, we write $\alpha_{ik}^{\mathrm{mrc}} = \hat{\alpha}_{ik}^{\mathrm{mrc}} + \nu_{ik}^{\mathrm{mrc}}$. Then, for a large $M$, LMMSE estimator becomes MMSE estimator, and consequently, $\alpha_{ik}^{\mathrm{mrc}}$ and $\nu_{ik}^{\mathrm{mrc}}$ are independent, and $\hat{\alpha}_{ik}^{\mathrm{mrc}} \sim \mathcal{N}(\mu_{ik}^{\mathrm{mrc}}, \sigma_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}^2)$ and $\nu_{ik}^{\mathrm{mrc}} \sim \mathcal{N}(0, \eta_{ik}^{\mathrm{mrc}})$ where

$$\mu_{ik}^{\mathrm{mrc}} \triangleq \mathbb{E}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} \quad (37)$$

$$\sigma_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}^2 \triangleq \frac{\left(\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\}\right)^2}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}}. \quad (38)$$

*Proof:* See Appendix E. ■

The quantities in (34)–(38) are given in Lemma 1 and 2. We observe that the MSE $\eta_{ik}^{\mathrm{mrc}}$ cannot be reduced to zero by increasing UL and DL pilot powers due to pilot contamination (c.f. Remark 2).

*2) System with ZF Precoding:* With ZF precoding, the signal in (23) is rewritten as

$$z_{ik}^{\mathrm{zf}} = \alpha_{ik}^{\mathrm{zf}} + w_{ik}^{\mathrm{zf}} \quad (39)$$

where $\alpha_{ik}^{\mathrm{zf}} = \frac{1}{\|\mathbf{a}_{ik}\|}$ is the effective channel gain and

$$w_{ik}^{\mathrm{zf}} \triangleq \frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik} + \sum_{l=1,l\neq i}^{L}\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik} + \mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}\right) \quad (40)$$

is the interference plus noise term. The next Lemma describes statistics of $\alpha_{ik}^{\mathrm{zf}}$.

*Lemma 4:* Mean and variance of the DL effective channel gain $\alpha_{ik}^{\mathrm{zf}}$ are given by

$$\mathbb{E}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} = \mathbb{E}\left\{\frac{1}{\|\mathbf{a}_{ik}\|}\right\} = C_{M-K+1}\sqrt{\sigma_{\hat{\mathbf{g}}_{iik}}^2} \quad (41)$$

$$\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} = V_{M-K+1}\sigma_{\hat{\mathbf{g}}_{iik}}^2 \quad (42)$$

and $\alpha_{ik}^{\mathrm{zf}}$ has a Gaussian pdf for a large value of $M$.

*Proof:* See Appendix F. ■

From Lemma 4, following the approximation of $C_{M-K+1} \approx \sqrt{M-K+1}$ and $V_{M-K+1} \approx 0.25$, we can see $\mathbb{E}\{\alpha_{ik}^{\mathrm{zf}}\}$ increases with $\sqrt{M-K+1}$ while $\mathrm{var}\{\alpha_{ik}^{\mathrm{zf}}\}$ is approximately fixed.

The real and imaginary parts of $w_{ik}^{\mathrm{zf}}$ have different statistics due to the same reason as explained for the MRC case, and the following Lemma describes them.

*Lemma 5:* The real part $w_{ik}^{\mathrm{zf,re}}$ and the imaginary part $w_{ik}^{\mathrm{zf,im}}$ of $w_{ik}^{\mathrm{zf}}$ are independent and

$$\mathbb{E}\left\{w_{ik}^{\mathrm{zf,re}}\right\} = C_{M-K+1}\sum_{l=1,l\neq i}^{L}\sqrt{\sigma_{\hat{\mathbf{g}}_{lik}}^2} \quad (43)$$

$$\mathbb{E}\left\{w_{ik}^{\mathrm{zf,im}}\right\} = 0 \quad (44)$$

$$\mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\} = \frac{1}{2}\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + \frac{1}{2\tau_{\mathrm{d}}q_{\mathrm{d}}} + V_{M-K+1}\sum_{l=1,l\neq i}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2 \quad (45)$$

$$\mathrm{var}\left\{w_{ik}^{\mathrm{zf,im}}\right\} = \frac{1}{2}\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + \frac{1}{2\tau_{\mathrm{d}}q_{\mathrm{d}}}. \quad (46)$$

In addition, $w_{ik}^{\mathrm{zf,im}}$ is Gaussian and $w_{ik}^{\mathrm{zf,re}}$ for a large value of $M-K$ approaches Gaussian. We can rewrite (45) as $\mathrm{var}\{w_{ik}^{\mathrm{zf,re}}\} = \frac{1}{2}(\sigma_{\boldsymbol{\varepsilon}_{iik}}^2 + \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}) + (\frac{1}{2} - V_{M-K+1})\sum_{l=1,l\neq i}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + V_{M-K+1}\sum_{l=1,l\neq i}^{L}\beta_{lik}$.

*Proof:* See Appendix G. ■

Next, by exploiting Lemma 5, we can estimate $\alpha_{ik}^{\mathrm{zf}}$ from the real part $z_{ik}^{\mathrm{zf,re}}$ of $z_{ik}^{\mathrm{zf}}$ where

$$z_{ik}^{\mathrm{zf,re}} = \alpha_{ik}^{\mathrm{zf}} + w_{ik}^{\mathrm{zf,re}}. \tag{47}$$

*Lemma 6:* For a system with ZF precoding, the LMMSE DL channel estimator based on (47) and its MSE are respectively given by

$$\hat{\alpha}_{ik}^{\mathrm{zf}} = c_{ik}^{\mathrm{zf}} + f_{ik}^{\mathrm{zf}} z_{ik}^{\mathrm{zf,re}} \tag{48}$$

$$\eta_{ik}^{\mathrm{zf}} \triangleq \mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{zf}} - \alpha_{ik}^{\mathrm{zf}}\right|^2\right\} = \frac{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} \cdot \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}} \tag{49}$$

where

$$c_{ik}^{\mathrm{zf}} \triangleq \frac{\mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}\mathbb{E}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} - \mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\}\mathbb{E}\left\{w_{ik}^{\mathrm{zf,re}}\right\}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}} \tag{50}$$

$$f_{ik}^{\mathrm{zf}} \triangleq \frac{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}}. \tag{51}$$

Denoting the estimation error by $\nu_{ik}^{\mathrm{zf}}$, we write $\alpha_{ik}^{\mathrm{zf}} = \hat{\alpha}_{ik}^{\mathrm{zf}} + \nu_{ik}^{\mathrm{zf}}$. Then, for a large $M$, LMMSE becomes MMSE estimator and, consequently, $\alpha_{ik}^{\mathrm{zf}}$ and $\nu_{ik}^{\mathrm{zf}}$ are independent, and $\hat{\alpha}_{ik}^{\mathrm{zf}} \sim \mathcal{N}(\mu_{ik}^{\mathrm{zf}}, \sigma_{\hat{\alpha}_{ik}}^2)$ and $\nu_{ik}^{\mathrm{zf}} \sim \mathcal{N}(0, \eta_{ik}^{\mathrm{zf}})$ where

$$\mu_{ik}^{\mathrm{zf}} \triangleq \mathbb{E}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} \tag{52}$$

$$\sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^2 \triangleq \frac{\left(\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\}\right)^2}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}}. \tag{53}$$

*Proof:* See Appendix E. ■

The quantities in (49)–(53) are given in Lemma 4 and 5.

*Remark 2:* Due to the pilot contamination, the normalized MSE of the effective DL channel estimation denoted by $\bar{\eta}_{ik} \triangleq \eta_{ik}/\mathbb{E}\{\alpha_{ik}^2\}$ approaches a floor as the UL and DL pilot SNRs increase. The floors of $\bar{\eta}_{ik}$ for the MRC and ZF precoders are respectively given by

$$\frac{V_M}{M}\frac{\frac{1}{2}\sum_{l=1}^{L}\underline{\sigma}_{\boldsymbol{\varepsilon}_{lik}}^2 + V_M\sum_{l=1,l\neq i}^{L}\underline{\sigma}_{\mathbf{g}_{lik}}^2}{\frac{1}{2}\sum_{l=1}^{L}\underline{\sigma}_{\boldsymbol{\varepsilon}_{lik}}^2 + V_M\sum_{l=1}^{L}\underline{\sigma}_{\mathbf{g}_{lik}}^2} \tag{54}$$

$$\frac{V_{M-K+1}}{M-K+1}\frac{\frac{1}{2}\sum_{l=1}^{L}\underline{\sigma}_{\boldsymbol{\varepsilon}_{lik}}^2 + V_{M-K+1}\sum_{l=1,l\neq i}^{L}\underline{\sigma}_{\mathbf{g}_{lik}}^2}{\frac{1}{2}\sum_{l=1}^{L}\underline{\sigma}_{\boldsymbol{\varepsilon}_{lik}}^2 + V_{M-K+1}\sum_{l=1}^{L}\underline{\sigma}_{\mathbf{g}_{lik}}^2} \tag{55}$$

where $\underline{\sigma}_{\boldsymbol{\varepsilon}_{lik}}^2 \triangleq \frac{\beta_{lik}\sum_{j=1,j\neq i}^{L}\beta_{ljk}}{\sum_{j=1}^{L}\beta_{ljk}}$ and $\underline{\sigma}_{\mathbf{g}_{lik}}^2 \triangleq \frac{\beta_{lik}^2}{\sum_{j=1}^{L}\beta_{ljk}}$. From (54) and (55), it can be seen by increasing $M$ the floor reduces and asymptotically with increase of $M$, approaches zero.

*Remark 3:* $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ and $\Re\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}\}$ are asymptotically independent; and even with a small $M$, they can be considered as independent for practical purposes. To justify this, we expand $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ in (33) by replacing $z_{ik}^{\mathrm{mrc,re}}$ with (32) and $w_{ik}^{\mathrm{mrc,re}}$ with the real part of (25). We note $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ consists of several terms including these two terms with factors $\alpha_{ik}^{\mathrm{mrc}} = \|\hat{\mathbf{g}}_{iik}\|$ and $\Re\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}\}$. While the total power of $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ is summation of powers of individual terms, the power ratio of the two terms is $\frac{2M\sigma_{\hat{\mathbf{g}}_{iik}}^2}{\sigma_{\boldsymbol{\varepsilon}_{iik}}^2} = \frac{2M\beta_{iik}}{\frac{1}{\tau_{\mathrm{u}}q_{\mathrm{u}}}+\sum_{l=1,l\neq i}^{L}\beta_{ilk}}$, which linearly increases with $M$. Even for a small value of $M = 10$, and typical setting of $\beta_{iik} = 1$, $\beta_{ilk} = 0.1$, $L = 7$, $q_{\mathrm{d}} = 10$ dB, that ratio is larger than 28. Thus, discarding the term with $\Re\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}\}$ from $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ does not affect the performance for practical purposes, justifying the above statement. Note that $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ and $\Im\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}\}$ are independent for any $M$. The same statement applies between $\hat{\alpha}_{ik}^{\mathrm{zf}}$ and $\Re\{\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik}\}$.

### B. Achievable Downlink Rates

Since the receiver has knowledge of $r_{ik}$ and the effective channel estimate $\hat{\alpha}_{ik}$, by starting from mutual information, i.e., $\mathcal{I}(s_{ik}; r_{ik}, \hat{\alpha}_{ik})$, and going through the same procedure as mentioned in [28, Appendix I], we can show the achievable DL rate is lower bounded by

$$R_{ik,\mathrm{p}} = \lambda_{\mathrm{p}}\mathbb{E}\left\{\log_2\left(1 + \frac{p\left|\hat{\alpha}_{ik}\right|^2}{\mathbb{E}\left\{\left|r_{ik}\right|^2\left|\hat{\alpha}_{ik}\right\} - p\left|\hat{\alpha}_{ik}\right|^2}\right)\right\} \tag{56}$$

where $\lambda_{\mathrm{p}}$ accounts for the pilot overhead, and we discuss it later in Remark 7. The outer expectation is over $\hat{\alpha}_{ik}$ and the expectation in the denominator is over the random variables in $r_{ik}$ conditioned on $\hat{\alpha}_{ik}$. Now, let us consider $r_{ik}$ in (18)–(21) and compute $\mathbb{E}\{\sqrt{p}s_{ik}I_{ik}^*|\hat{\alpha}_{ik}\}$. Applying the asymptotic independence of $\hat{\alpha}_{ik}$ from $\nu_{ik}$ and $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}$ for the system with MRC precoding and from $\nu_{ik}$ and $\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik}$ for the system with ZF precoding, together with the zero mean property of individual terms of $\mathbb{E}\{\sqrt{p}s_{ik}I_{ik}^*|\hat{\alpha}_{ik}\}$, we can conclude $\mathbb{E}\{\sqrt{p}s_{ik}I_{ik}^*|\hat{\alpha}_{ik}\} \approx 0$. As a result, the rate expression in (56) is approximated to

$$R_{ik,\mathrm{p}} \approx \lambda_{\mathrm{p}}\mathbb{E}\left\{\log_2\left(1 + \frac{p\left|\hat{\alpha}_{ik}\right|^2}{\mathbb{E}\left\{\left|I_{ik}\right|^2\left|\hat{\alpha}_{ik}\right\}}\right)\right\}. \tag{57}$$

*1) System with MRC Precoding:* To find a closed-form expression for the achievable rate in (57), first we find the conditional interference power in the denominator, and then find the approximate value for the outer expectation.

*Lemma 7:* The conditional interference power at user $k$ for a system with MRC precoding is given by

$$\mathbb{E}\left\{\left|I_{ik}^{\mathrm{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right.\right\} \approx p\kappa_{ik}^{\mathrm{mrc}} \tag{58}$$

where

$$\kappa_{ik}^{\mathrm{mrc}} \triangleq \eta_{ik}^{\mathrm{mrc}} + \sigma^2_{\boldsymbol{\varepsilon}_{iik}} + (K-1)\theta_{ik}^{\mathrm{mrc}} + (K-1)\sum_{l=1,l\neq i}^{L}\beta_{lik} + \sum_{l=1,l\neq i}^{L}\zeta_{ilk}^{\mathrm{mrc}} + \frac{1}{p}, \tag{59}$$

$$\theta_{ik}^{\mathrm{mrc}} \triangleq \frac{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right|^2}{M} + \frac{\eta_{ik}^{\mathrm{mrc}}}{M} + \sigma^2_{\boldsymbol{\varepsilon}_{iik}} \tag{60}$$

$$\zeta_{ilk}^{\mathrm{mrc}} \triangleq \left(C_M\sqrt{\sigma^2_{\hat{\mathbf{g}}_{lik}}} + \frac{\gamma_{ilk}^{\mathrm{mrc}}\left(\hat{\alpha}_{ik}^{\mathrm{mrc}} - \mathbb{E}\left\{\hat{\alpha}_{ik}^{\mathrm{mrc}}\right\}\right)}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\}}\right)^2 + \left(1 - \frac{\gamma_{ilk}^{\mathrm{mrc}}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{mrc}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{mrc,re}}\right\}}\right)\gamma_{ilk}^{\mathrm{mrc}} + \frac{1}{2}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} \tag{61}$$

$$\gamma_{ilk}^{\mathrm{mrc}} \triangleq V_M\sigma^2_{\hat{\mathbf{g}}_{lik}} + \frac{1}{2}\sigma^2_{\boldsymbol{\varepsilon}_{lik}}. \tag{62}$$

*Proof:* See Appendix H. ■

For a large value of $M$, due to the Gaussian pdf of $\hat{\alpha}_{ik}^{\mathrm{mrc}}$ in Lemma 3, $|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2$ has a non-central Chi-square pdf with 1 degree of freedom (DoF). Its mean and variance are given by

$$\mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right|^2\right\} = \left(\mu_{ik}^{\mathrm{mrc}}\right)^2 + \sigma^2_{\hat{\alpha}_{ik}^{\mathrm{mrc}}} \tag{63}$$

$$\mathrm{var}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right|^2\right\} = 2\sigma^4_{\hat{\alpha}_{ik}^{\mathrm{mrc}}} + 4\sigma^2_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}\left(\mu_{ik}^{\mathrm{mrc}}\right)^2. \tag{64}$$

From (37) and (38), we can observe that for a large value of $M$, $\sigma^2_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}$ dose not depend on $M$ but $\mu_{ik}^{\mathrm{mrc}}$ increases with $\sqrt{M}$. Thus, both mean and variance of $|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2$ increase linearly with $M$. By substituting (58) in (57), the DL achievable rate bound is given by

$$R_{ik,p}^{\mathrm{mrc}} \approx \lambda_{\mathrm{p}}\mathbb{E}\left\{\log_2\left(1 + \frac{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right|^2}{\kappa_{ik}^{\mathrm{mrc}}}\right)\right\}. \tag{65}$$

Next, we want to find the outer expectation in (65) over the pdf of $\hat{\alpha}_{ik}^{\mathrm{mrc}}$. However, the expectation integral does not lead to a closed-form solution. A common approach adopted in the literature for such a problem is the use of a lower bound by means of Jensen's inequality, i.e., $\log_2(1+\frac{1}{\mathbb{E}\{1/X\}}) \leq \mathbb{E}\{\log_2(1+X)\}$. But such an approach fails in our case because $\mathbb{E}\{\frac{1}{|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2}\}$ is infinity and that lower bound reads as a trivial zero bound. Thus, we use the approximation $\mathbb{E}\{\log_2(1+\frac{X}{Y})\} \approx \log_2(1+\frac{\mathbb{E}\{X\}}{\mathbb{E}\{Y\}})$. Following the proof of the lemma in [13, Appendix I], we can see that if $X$ and $Y$ are two non-negative random variables with properties that $\frac{1}{\mathbb{E}^2\{Y\}}\mathrm{var}\{Y\}$ and $\frac{1}{\mathbb{E}^2\{X+Y\}}\mathrm{var}\{X+Y\}$ approaches zero, then that approximation is asymptotically tight. Here, $X$ and $Y$ represent $|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2$ and $\kappa_{ik}^{\mathrm{mrc}}$, respectively. We can see $\frac{\mathrm{var}\{\kappa_{ik}^{\mathrm{mrc}}\}}{\mathbb{E}^2\{\kappa_{ik}^{\mathrm{mrc}}\}} \to 0$ and $\frac{\mathrm{var}\{\kappa_{ik}^{\mathrm{mrc}}+|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2\}}{\mathbb{E}^2\{\kappa_{ik}^{\mathrm{mrc}}+|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2\}} \to 0$ as $M\to\infty$, since the two numerators increase with order $M$ while the two denominators increase with order $M^2$. Hence, the approximation is asymptotically tight as the number of antennas increases. Then, based on this approximation, we obtain an approximate rate $\tilde{R}_{ik,\mathrm{p}}^{\mathrm{mrc}} \approx R_{ik,\mathrm{p}}^{\mathrm{mrc}}$ given by

$$\tilde{R}_{ik,\mathrm{p}}^{\mathrm{mrc}} \triangleq \lambda_{\mathrm{p}}\log_2\left(1 + \frac{\mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{mrc}}\right|^2\right\}}{\mathbb{E}\left\{\kappa_{ik}^{\mathrm{mrc}}\right\}}\right) \tag{66}$$

where $\mathbb{E}\{|\hat{\alpha}_{ik}^{\mathrm{mrc}}|^2\}$ is given by (63). As will be shown in simulation results, $\tilde{R}_{ik,\mathrm{p}}^{\mathrm{mrc}}$ matches $R_{ik,\mathrm{p}}^{\mathrm{mrc}}$ in (65), even for a small value of $M$.

This subsection is summarized by the following theorem.

*Theorem 3:* For a system with MRC precoding and orthogonal DL pilot, the approximate DL rate bound is given by

$$\tilde{R}_{ik,\mathrm{p}}^{\mathrm{mrc}} = \lambda_{\mathrm{p}}\log_2\left(1 + \frac{C_M^2\sigma^2_{\hat{\mathbf{g}}_{iik}} + \sigma^2_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}}{\jmath_{ik,\mathrm{p}}^{\mathrm{mrc}}}\right) \tag{67}$$

where $\jmath_{ik,\mathrm{p}}^{\mathrm{mrc}} \triangleq \eta_{ik}^{\mathrm{mrc}} + \sigma^2_{\boldsymbol{\varepsilon}_{iik}} + (K-1)\beta_{iik} + K\sum_{l=1,l\neq i}^{L}\beta_{lik} + (M-1)\sum_{l=1,l\neq i}^{L}\sigma^2_{\hat{\mathbf{g}}_{lik}} + \frac{1}{p}$.

*Proof:* It immediately follows from (66), (63), and straightforward calculation of $\mathbb{E}\{\kappa_{ik}^{\mathrm{mrc}}\}$ in (59). ■

*Remark 4:* In (67), the numerator for a large $M$ is approximately $C_M^2\sigma^2_{\hat{\mathbf{g}}_{iik}} \approx M\sigma^2_{\hat{\mathbf{g}}_{iik}}$, which is the same as the numerator in (13), and the denominators are approximately the same (The denominator for the case with DL pilot is always slightly smaller than that without DL pilot due to $\eta_{ik}^{\mathrm{mrc}} < V_M\sigma^2_{\hat{\mathbf{g}}_{iik}} = \mathrm{var}\{\alpha_{ik}^{\mathrm{mrc}}\}$ (see (34)). Hence, comparing (67) and (13), and neglecting pilot overhead, we see DL rates for MRC precoding with DL pilot and without DL pilot are very close for a large $M$.

*2) System with ZF Precoding:* First we present the expectation of the interference power conditioned on $\hat{\alpha}_{ik}^{\mathrm{zf}}$, and then use an approximation for computing the outer expectation.

*Lemma 8:* The conditional interference power at user $k$ for a system with ZF precoding is given by

$$\mathbb{E}\left\{\left|I_{ik}^{\mathrm{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\mathrm{zf}}\right.\right\} \approx p\,\kappa_{ik}^{\mathrm{zf}} \tag{68}$$

where

$$\kappa_{ik}^{\mathrm{zf}} \triangleq \eta_{ik}^{\mathrm{zf}} + K\sigma^2_{\boldsymbol{\varepsilon}_{iik}} + (K-1)\sum_{l=1,l\neq i}^{L}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} + \sum_{l=1,l\neq i}^{L}\zeta_{ilk}^{\mathrm{zf}} + \frac{1}{p} \tag{69}$$

with

$$\zeta_{ilk}^{\mathrm{zf}} \triangleq \left(C_{M-K+1}\sqrt{\sigma^2_{\hat{\mathbf{g}}_{lik}}} + \frac{\gamma_{ilk}^{\mathrm{zf}}\left(\hat{\alpha}_{ik}^{\mathrm{zf}} - \mathbb{E}\left\{\hat{\alpha}_{ik}^{\mathrm{zf}}\right\}\right)}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\}}\right)^2 + \left(1 - \frac{\gamma_{ilk}^{\mathrm{zf}}}{\mathrm{var}\left\{\alpha_{ik}^{\mathrm{zf}}\right\} + \mathrm{var}\left\{w_{ik}^{\mathrm{zf,re}}\right\}}\right)\gamma_{ilk}^{\mathrm{zf}} + \frac{1}{2}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} \tag{70}$$

$$\gamma_{ilk}^{\mathrm{zf}} \triangleq V_{M-K+1}\sigma^2_{\hat{\mathbf{g}}_{lik}} + \frac{1}{2}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} \tag{71}$$

*Proof:* See Appendix I. ■

By substituting (68) in (57), the DL achievable rate bound reads as

$$R_{ik,\mathrm{p}}^{\mathrm{zf}} \approx \lambda_{\mathrm{p}}\mathbb{E}\left\{\log_2\left(1 + \frac{\left|\hat{\alpha}_{ik}^{\mathrm{zf}}\right|^2}{\kappa_{ik}^{\mathrm{zf}}}\right)\right\}. \tag{72}$$

For a large value of $M$, due to the Gaussian pdf of $\hat{\alpha}_{ik}^{\mathrm{zf}}$ in Lemma 6, $|\hat{\alpha}_{ik}^{\mathrm{zf}}|^2$ has a non-central Chi-square pdf with 1 DoF. Its mean and variance are given by

$$\mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{zf}}\right|^2\right\} = \left(\mu_{ik}^{\mathrm{zf}}\right)^2 + \sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^2 \tag{73}$$

$$\mathrm{var}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{zf}}\right|^2\right\} = 2\sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^4 + 4\sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^2\left(\mu_{ik}^{\mathrm{zf}}\right)^2. \tag{74}$$

Next finding the outer expectation in (72) over the pdf of $\hat{\alpha}_{ik}^{\mathrm{zf}}$, similar to the MRC precoding case, does not lead to a simple closed-form solution and finding a lower bound based on Jensen's inequality fails. Now, in the same way as discussed for MRC, we find an asymptotically (with respect to $M$) tight approximation for (72). Simulation results show this approximation accurately matches with the actual value of expectation. Expression of our approximate rate $\tilde{R}_{ik,\mathrm{p}}^{\mathrm{zf}} \approx R_{ik,\mathrm{p}}^{\mathrm{zf}}$ reads as

$$\tilde{R}_{ik,\mathrm{p}}^{\mathrm{zf}} \triangleq \lambda_{\mathrm{p}} \log_2\left(1 + \frac{\mathbb{E}\left\{\left|\hat{\alpha}_{ik}^{\mathrm{zf}}\right|^2\right\}}{\mathbb{E}\left\{\kappa_{ik}^{\mathrm{zf}}\right\}}\right) \tag{75}$$

where $\mathbb{E}\{|\hat{\alpha}_{ik}^{\mathrm{zf}}|^2\}$ is given by (73). This subsection is summarized in the following theorem.

*Theorem 4:* For a DL system with ZF precoding and orthogonal DL pilot, the approximate achievable rate bound is given by

$$\tilde{R}_{ik,\mathrm{p}}^{\mathrm{zf}} = \lambda_{\mathrm{p}} \log_2\left(1 + \frac{C_{M-K+1}^2 \sigma_{\hat{\mathbf{g}}_{iik}}^2 + \sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^2}{\mathfrak{I}_{ik,\mathrm{p}}^{\mathrm{zf}}}\right) \tag{76}$$

where $\mathfrak{I}_{ik,\mathrm{p}}^{\mathrm{zf}} \triangleq \eta_{ik}^{\mathrm{zf}} + K\sum_{l=1}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2 + (M-K+1)\sum_{l=1,l\neq i}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \frac{1}{p}$.

*Proof:* It immediately follows from (75), (73), and straightforward calculation of $\mathbb{E}\{\kappa_{ik}^{\mathrm{zf}}\}$ in (69). ■

*Remark 5:* In (76), the numerator for a large $M$ is approximately $C_{M-K+1}^2\sigma_{\hat{\mathbf{g}}_{iik}}^2$, which is the same as the numerator in (16), and the denominator of (76) is always smaller than the denominator of (16) due to $\eta_{ik}^{\mathrm{zf}} < V_{M-K+1}\sigma_{\hat{\mathbf{g}}_{iik}}^2 = \mathrm{var}\{\alpha_{ik}^{\mathrm{zf}}\}$ (see (49)). However, the difference between the denominators is negligible. So, the DL rates for ZF precoding with DL pilot and without DL pilot are very close for a large $M$.

*Remark 6:* As $M \to \infty$, DL rates for MRC and ZF precoding in (67) and (76) saturate to the same value as $\tilde{R}_{ik,\mathrm{p}}^{\mathrm{mrc}}, \tilde{R}_{ik,\mathrm{p}}^{\mathrm{zf}} \to \lambda_{\mathrm{p}} R_{ik}^{\infty}$, where $R_{ik}^{\infty}$ is defined in (17).

*Remark 7:* The pilot overhead for DL rate expressions can be defined based on DL pilot only or both DL and UL pilots since UL pilot is used for both UL data detection and DL transmit beamforming. In the second definition, UL pilot overhead can be divided between UL and DL according to the proportion of UL and DL frame lengths; with the same frame length of $T$ symbols, the overhead factors read as $\lambda_{\mathrm{p}}^{\mathrm{ud}} = \frac{T-\tau_{\mathrm{d}}-\frac{\tau_{\mathrm{u}}}{2}}{T}$ and $\lambda_{\mathrm{np}}^{\mathrm{ud}} = \frac{T-\frac{\tau_{\mathrm{u}}}{2}}{T}$. In the first definition, the overhead factors are given by $\lambda_{\mathrm{np}}^{\mathrm{d}} = 1$ and $\lambda_{\mathrm{p}}^{\mathrm{d}} = \frac{T-\tau_{\mathrm{d}}}{T}$. If we use the minimum length for orthogonal pilot sequences, $K$ replaces $\tau_u$ and $\tau_d$ in the above overhead factors.

*Remark 8:* Comparing DL rates for systems without DL pilot in (13) and (16), we observe that MRC precoding has a higher rate if $M$ is smaller than a threshold given by

$$M_{\mathrm{th,\,np}} \triangleq \frac{V_M\sigma_{\hat{\mathbf{g}}_{iik}}^2 + \frac{1}{p} + K\sum_{l=1}^{L}\beta_{lik}}{\sum_{l=1}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2} - 1. \tag{77}$$

Similarly, from the DL rates in (67) and (76), the threshold for systems with DL pilots is obtained as

$$M_{\mathrm{th,\,p}} \triangleq \frac{\eta_{ik} + \frac{1}{p} + K\sum_{l=1}^{L}\beta_{lik} - \sigma_{\hat{\alpha}_{ik}}^2}{\sum_{l=1}^{L}\sigma_{\hat{\mathbf{g}}_{lik}}^2} + 2\frac{\sigma_{\hat{\alpha}_{ik}}^2}{\sigma_{\hat{\mathbf{g}}_{iik}}^2} - 1 \tag{78}$$

where in deriving the threshold we have used the approximations $\sigma_{\hat{\alpha}_{ik}}^2 = \sigma_{\hat{\alpha}_{ik}^{\mathrm{mrc}}}^2 \approx \sigma_{\hat{\alpha}_{ik}^{\mathrm{zf}}}^2$ and $\eta_{ik} = \eta_{ik}^{\mathrm{mrc}} \approx \eta_{ik}^{\mathrm{zf}}$, These thresholds are for each user, and they can be useful in developing adaptive beam-forming for each user.

## V. Numerical Results and Discussions

We use a typical multicell structure with $L = 7$ cells and frequency reuse factor of 1. There are $K = 10$ users in each cell, and the numbers of pilot symbols for UL and DL are $\tau_{\mathrm{u}} = \tau_{\mathrm{d}} = K$, and there are $T = 200$ symbols in each frame. We assume the UL transmit pilot SNR is $q_{\mathrm{u}} = 10$dB (unless mentioned otherwise) and the DL pilot and data SNRs are the same $p = q_{\mathrm{d}}$. We evaluate both settings of fixed $\{\beta_{ilk}\}$ and random $\{\beta_{ilk}\}$. In the fixed setting, we set $\beta_{iik} = 1$ and $\beta_{ilk} = a$, $\forall l \neq i$. In the random setting, users in each cell are uniformly located in the disk with radius 1000 m around BS not closer than $d_0 = 100$ m to BS, and the distance of user $k$ in cell $l$ to BS $i$ is denoted by $d_{ilk}$. So, $\{\beta_{ilk}\}$ are independently generated by $\beta_{ilk} = \psi/(\frac{d_{ilk}}{d_0})^{\nu}$ where $\nu = 3.8$ and $10\log_{10}(\psi) \sim \mathcal{N}(0, \sigma_{\mathrm{shadow,\,dB}}^2)$ with $\sigma_{\mathrm{shadow,\,dB}} = 8$. We use the first definition of pilot overhead mentioned in Remark 7 in all the simulations as the difference between the two definitions is just a scaling. The exception is in Fig. 6 where we use both of the definitions to investigate sum-rate versus the number of users. In the figure legends, "NP" refers to the system without DL pilot, "sim" represents Monte Carlo simulation results of non-closed-form rate bounds while "ana" denotes our proposed analytical closed-form rates. For systems with DL pilot, the rate bounds for MRC and ZF and their corresponding approximate closed-form rates are (65), (72), (67) and (76), respectively. For systems without DL pilot, the rate bound is given in (11) and the closed form rate bounds for MRC and ZF are given in (13) and (16).

First, we compare our closed-form DL rates with the Monte Carlo based actual DL rate bounds for systems with and without DL pilots. Fig. 1 shows the sum rate versus Tx SNR for $M = 100$ and $a = 0.2$ while Fig. 2 presents the sum rate versus the number of BS antennas $M$ for $p = q_{\mathrm{d}} = 10$ dB and $a = 0.2$. The closed-form results match with the simulation-based rates for all cases. After verifying the accuracy of the closed-form expressions, we study performance characteristics under various conditions. Fig. 1 illustrates rate saturation of all schemes as the DL Tx SNR increases. This implies an interference dominated operating condition and suggests a low SNR region for the operation of the massive MIMO systems in multicell environments. We have also evaluated for other settings with $M = 30$ and/or $a = 0.1$; except the obvious increase of

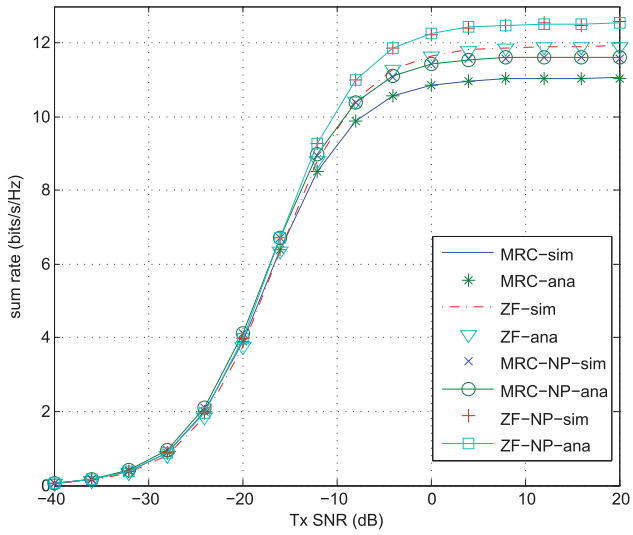



Fig. 1. The DL sum rate performance versus the Tx SNR for systems with and without DL pilot ($M = 100$, $a = 0.2$, $q_{\mathrm{u}} = 10$dB).

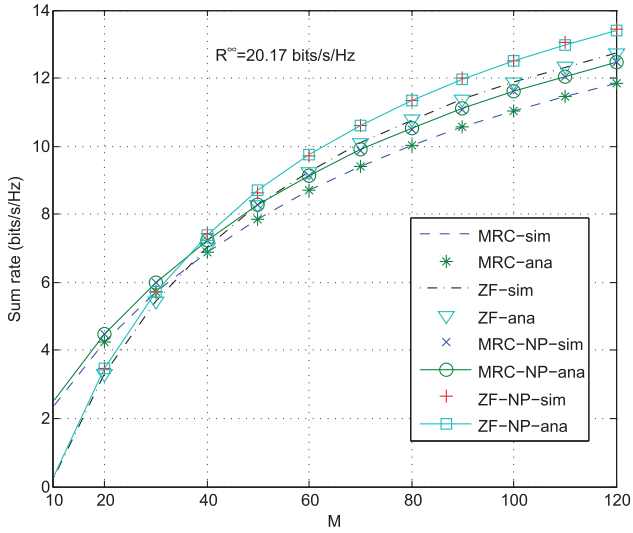



Fig. 2. The effect of the number of BS antennas on the DL sum rates ($p = q_{\mathrm{d}} = 10$ dB, $a = 0.2$).

the saturated rate for a larger $M$ or a smaller $a$, the results show the same trend as in Fig. 1 and hence they are omitted.

From Fig. 2, we observe that MRC scheme performs better (worse) than ZF scheme if the number of BS antennas is small (large). This can be explained as follows. The rate expression has a numerator factor (beam-forming power gain) of $M - K + 1$ for ZF and $M$ for MRC which yields a substantial rate difference between the two schemes at a small $M$ but a negligible rate difference at a large $M$. The denominator in the rate expression is due to interference plus noise, and the ZF scheme has better interference suppression capability (smaller denominator) than the MRC scheme. Thus, at a small $M$, better beam-forming gain of MRC outperforms better interference suppression capability of ZF, yielding performance advantage of MRC. But as $M$ increases, the beam-forming gains of the two schemes become similar and the better interference suppression of ZF yields performance advantage of ZF. Note that

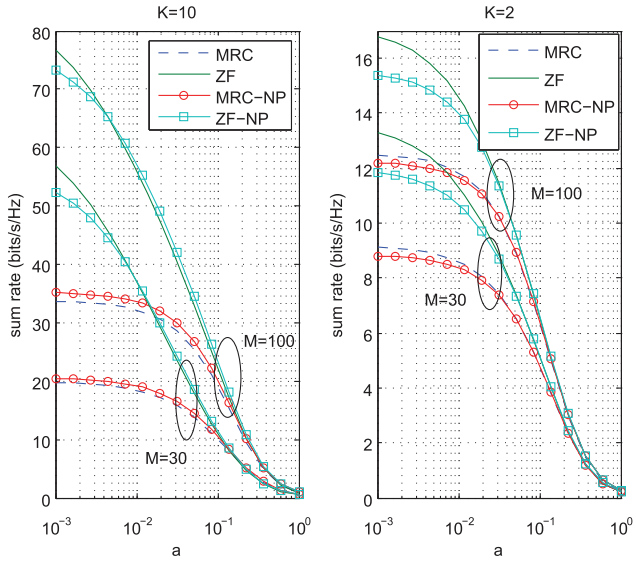



Fig. 3. The impact of the cross-cell interference level $a$ on the sum rates ($p = q_{\mathrm{d}} = q_{\mathrm{u}} = 10$ dB).

as $M \to \infty$, the user channels become orthogonal and hence MRC achieves the same interference suppression capability as ZF, thus yielding the same asymptotic rate (given in (17)).

Next, Fig. 3 elaborates effects of the cross-cell large-scale channel power gain $a$ under the settings of $p = q_{\mathrm{d}} = 10$ dB, $M = 100$ or 30, and $K = 2$ and 10. We observe that as $a$ increases, the sum rates of the two transmission schemes decrease due to their imperfect interference suppression. ZF outperforms MRC at a low cross-cell interference level (i.e., $a \leq 0.2$ for the system with $M = 100$, and $a \leq 0.15$ for $M = 30$), but both of the schemes show almost the same performance as the cross-cell interference level increases. This can be explained as follows. At a low cross-cell interference level, the effect of pilot contamination is insignificant and the ZF scheme can maintain most of its interference suppression capability. However, as the cross-cell interference level increases, the pilot contamination becomes significant which in turn voids the better interference suppression capability of the ZF scheme.

Another observation from Fig. 3 is for a small $a$, the rate with DL pilot is larger than that without DL pilot for ZF, but for MRC the same holds if additionally $K$ is very small. For a large $a$, the system without DL pilot yields a larger rate than that with DL pilot as can be observed in Fig. 1 and Fig. 2. The reasons are as follows. First, the advantage of the system without DL pilot comes from zero DL pilot overhead; at a large $a$ (interference-dominated scenario), rates of the two systems excluding pilot overhead are similar and the overhead cost of the system with DL pilot yields its smaller rate (including overhead cost). Second, when $a$ is small, the pilot contamination and intercell interference are small, and the rate is mainly influenced by the intra-cell interference. The main difference in contributing the intra-cell interference is the variance of the effective DL channel gain for the system without DL pilot versus the MSE of the effective DL channel estimate for the system with DL pilot. When $a$ is small, the former is much larger than the latter, thus giving rate advantage for the system with DL pilot and

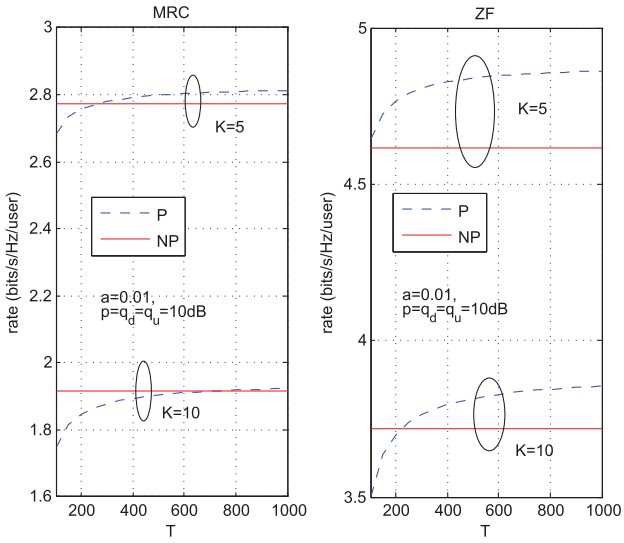



Fig. 4. Per-user rate versus the frame length (with $a = 0.01$, $M = 100$, and $p = q_d = 10$dB).

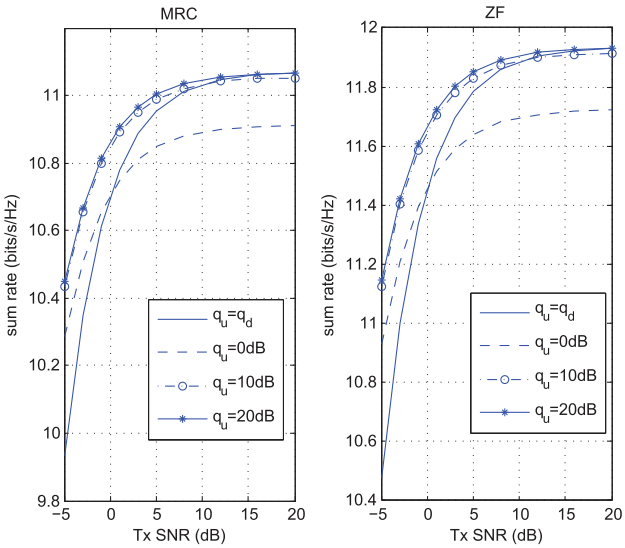



Fig. 5. DL sum rate versus DL Tx SNR ($= q_u = p$) under various settings of UL pilot power (with $M = 100$ and $a = 0.2$).

ZF precoding. For MRC, non-orthogonality of user channels causes additional multi-user interference which is proportional to $K$, thus requiring an additional condition of a small $K$ for the system with DL pilot to outperform the system without DL pilot.

Fig. 4 shows the effects of the frame length in comparing per-user rates between the system with and without DL pilot, for $K = 10$ and $K = 5$ with $a = 0.01$, $M = 100$, and $p = q_d = 10$dB. As we expect, a longer frame length yields a larger rate for systems with DL pilot since the cost of the additional overhead due to DL pilot reduces. The gain of the system with DL pilot over that without DL pilot is larger for ZF than MRC, and for a smaller number of users. This can be explained by the same reasons used for Fig. 3 in the above paragraph.

In the above results, we use a fixed UL pilot SNR $q_u = 10$ dB. In Fig. 5, we present how different UL pilot SNR settings impact the DL rate bounds for systems with $a = 0.2$

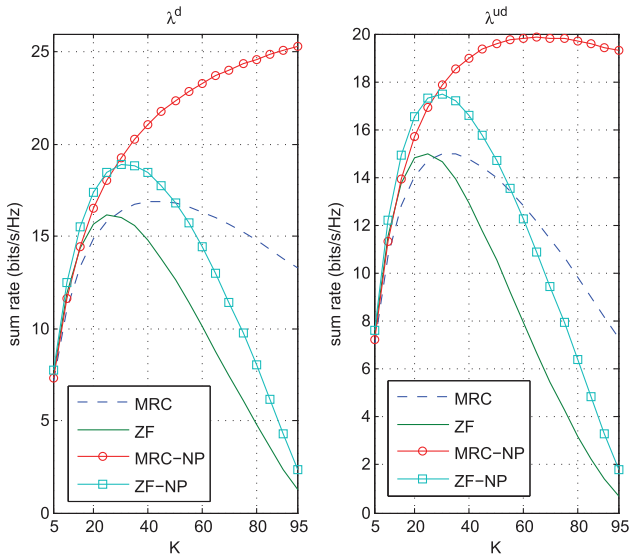



Fig. 6. DL sum rate versus the number of users $K$ (DL Tx SNR ($= q_d = p$) of 10 dB, $M = 100$, and $a = 0.2$).

and DL pilot. Both MRC and ZF precoders show the same trend. All the curves for different $q_u$ show DL rate saturation. For convenience of discussion, let us loosely term the saturation SNR as the DL Tx SNR above which the rate saturation occurs. Then we observe that the UL pilot SNR $q_u$ much lower than the saturation SNR, e.g., $q_u = 0$ dB, yields a drop of the DL rate ceiling. However, $q_u$ above the saturation SNR, e.g., $q_u = 20$ dB, gives very little improvement of the DL rate ceiling. These results imply that both UL and DL pilot Tx SNR should be set around the DL saturation SNR to avoid rate loss or power-inefficient rate saturation.

Next, Fig. 6 shows the sum rate with respect to the number of users $K$ in each cell for systems with and without DL pilot based on the closed form results. Since $K$ affects the overhead, we consider both definitions of overhead factors mentioned in Remark 7. The results show that the sum rate is nonlinearly related to $K$ and there exists an optimal number of users to maximize the sum rate for all cases except the MRC scheme without DL pilot under the overhead definition of $\lambda_{\mathrm{np}}^{\mathrm{d}}$. The reason can be explained by the following factors as $K$ increases: (i) the increasing spatial multiplexing gain, (ii) the increasing interference level and (iii) the (potentially) increasing pilot overhead. For the exception case, the third factor is nulled, the first factor linearly increases the sum rate, and the second factor decreases the sum rate from inside the $\log_2$ function. Thus, the sum rate increases with $K$. For the MRC scheme without DL pilot under the overhead definition of $\lambda_{\mathrm{np}}^{\mathrm{ud}}$, the roles of the first and second factors are the same, but the third factor reduces the sum rate linearly, thus yielding existence of an optimal value of $K$. For all of the ZF cases, the roles of the first and third factors are the same as the corresponding MRC cases. The second factor is less for ZF than MRC due to the interference suppression capability of ZF. But an additional factor for ZF, namely the decreasing beam-forming gain of ZF as $K$ increases, results in an existence of an optimal $K$ for all the ZF cases and their sum rates decay faster than the MRC counterparts as $K$ increases beyond its optimal point.

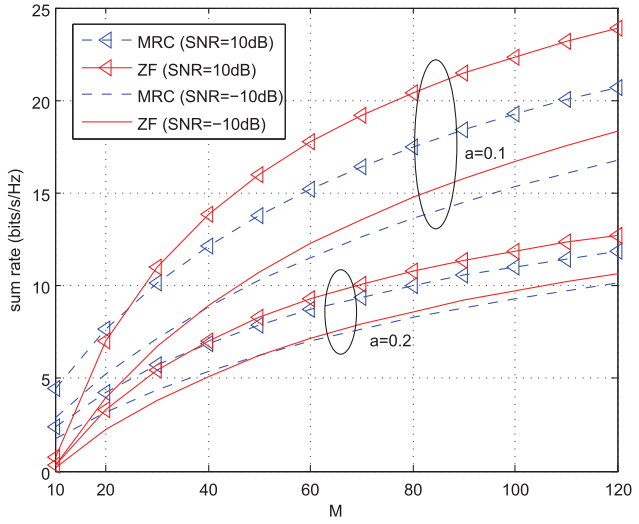



Fig. 7. The DL performance cross-over point between MRC and ZF transmission schemes with DL training in terms of the number of BS antennas $M$ under different conditions of Tx SNR (with $p = q_{\rm d}$) and cross-cell large-scale channel power gain $a$.

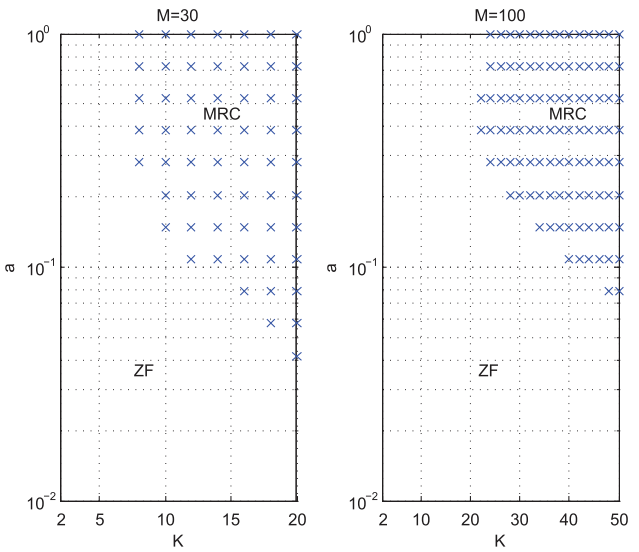



Fig. 8. Preferred regions of system parameters for ZF and MRC precoders (without DL pilot, $p = q_d = 10$dB).

Next, we present in Fig. 7 how different conditions can influence $M_{\rm th}$, the threshold point of $M$ for choosing the precoder. We observe a larger $M_{\rm th}$ at a larger $a$ or a smaller DL Tx SNR. Both conditions increase the interference plus noise level for both precoders which in turns dilutes the interference suppression advantage of ZF scheme. Thus, a larger beam-forming gain (a larger $M$) is required for ZF scheme to outperform MRC scheme.

We further investigate the selection of the precoding scheme based on $M_{\rm th,np}$ as mentioned in Remark 8. As the thresholds for systems with and without DL pilot are quite similar, we just show the results for systems without DL pilot in Fig. 8. The considered discrete points in the parameter space of $a$ and $K$ over which MRC outperforms ZF scheme are shown with cross marks for $M = 30$ and $M = 100$. We observe that

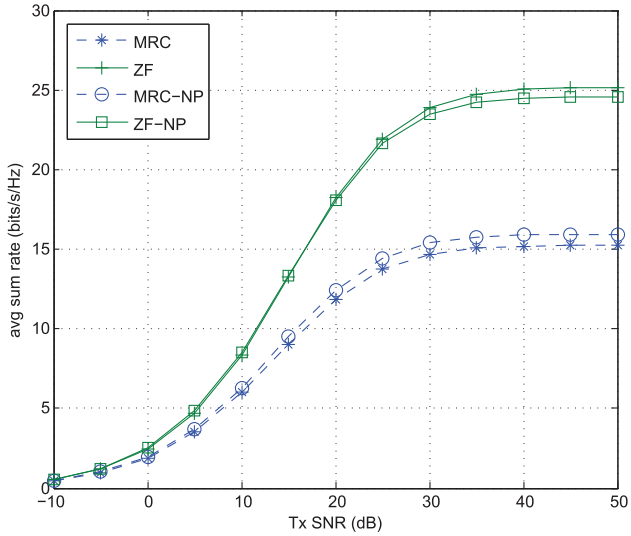



Fig. 9. Average sum rate under random $\{\beta_{ilk}\}$ versus the Tx SNR (with $M = 100$ and $K = 10$).

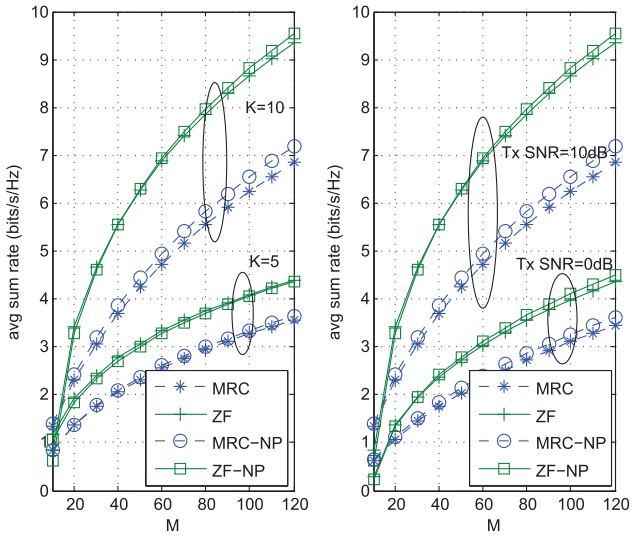



Fig. 10. Average sum rate under random $\{\beta_{ilk}\}$ versus $M$ for different $K$ ($p = q_d = 10$dB) (left) and for different DL Tx SNR ($K = 10$) (right).

for a fixed $M$, smaller $K$ or/and smaller $a$ yield performance advantage of ZF over MRC since under those conditions ZF has much better interference suppression capability and similar beamforming gain if compared to MRC scheme.

Next, we evaluate the sum rate performance under random large-scale fading coefficients. The results are obtained by averaging the sum rates over 1000 large-scale fading coefficients. Fig. 9 presents the average sum rate versus DL Tx SNR which has the same trend as the sum rate for fixed large-scale fading coefficients (see Fig. 1) but with an SNR shift. The reason for the SNR shift can be ascribed to the larger path loss and smaller values of log-normal shadowing coefficients than the fixed coefficient scenario. Fig. 10 shows the average sum rate versus the number of antennas for different values of $K$ in the left and different values of DL Tx SNR in the right. The results are in the same trend as in Fig. 7 except the crossover points of the rate curves between ZF and MRC are shifted to the left.

The reason is due to the larger path loss and smaller values of log-normal shadowing coefficients than the fixed coefficient scenario, and it is also consistent with the characteristics of a smaller $a$ in Fig. 7.

## VI. Conclusions

We have studied a multicell massive MIMO DL system with pilot contamination for MRC and ZF precoders at the base station for two DL scenarios of with and without a pilot-aided coherent detector at users. We have developed an LMMSE DL channel estimator by incorporating the non-circularly-symmetric characteristics of the effective DL channels. We have derived closed-form approximate achievable rate expressions for both of the DL scenarios. Numerical results illustrate that the derived rate expressions are very accurate. Our investigation shows that the considered massive MIMO DL system is interference-dominated in medium and high SNR regions due to imperfect CSI and pilot contamination, and the system is best deployed in a low SNR region. The system with DL pilot may have advantage over the system without DL pilot under scenarios with small cross-cell interference level, small pilot overhead with respect to the frame length, and small number of users. We have also presented criteria for choosing between MRC and ZF precoder, which depends on the number of BS antennas, the number of users, the cross-cell interference level, and the DL transmit SNR. In most of the scenarios investigated, there exists an optimal number of users to maximize the sum rate, which can be easily found from our closed form rate expressions. Thus, our closed form rate analysis offers insights and guidance for an efficient massive MIMO system design.

## Appendix A
## Proof of Theorem 1

First we present a Lemma which we use frequently in this appendix and others.

*Lemma 9:* If $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \sigma_x^2\mathbf{I}_M)$ and $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \sigma_y^2\mathbf{I}_M)$ are independent, then $\frac{\mathbf{x}^H}{\|\mathbf{x}\|}\mathbf{y} \sim \mathcal{CN}(0, \sigma_y^2)$.

*Proof:* Define $z \triangleq \frac{\mathbf{x}^H}{\|\mathbf{x}\|}\mathbf{y}$. Then, by conditioning on $\mathbf{x}$, we see $z|\mathbf{x}$ is a weighted sum of i.i.d. complex Gaussian random variables with pdf $\mathcal{CN}(0, \sigma_y^2)$ where the powers of the weights sum up to 1. Thus, the pdf of $z|\mathbf{x}$ denoted by $f_{z|\mathbf{x}}(z)$ is $\mathcal{CN}(0, \sigma_y^2)$ which is independent of $\mathbf{x}$. Consequently, the pdf of $z$ denoted by $f_z(z)$ equals $f_{z|\mathbf{x}}(z) = \mathcal{CN}(0, \sigma_y^2)$. ■

When users do not estimate their channels, we decompose the received signal in (9) as $r_{ik}^{\text{mrc}} = \mathbb{E}\{\sqrt{p}\,\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik}\}s_{ik} + \tilde{I}_{ik}^{\text{mrc}}$, where $\tilde{I}_{ik}^{\text{mrc}} = \sum_{j=1}^{5} \tilde{I}_{ik,j}^{\text{mrc}}$ and we define these different interference components as $\tilde{I}_{ik,1}^{\text{mrc}} \triangleq \sqrt{p}(\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik} - \mathbb{E}\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik}\})s_{ik}$, $\tilde{I}_{ik,2}^{\text{mrc}} \triangleq \sum_{j=1,j\neq k}^{K} \sqrt{p}\,\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}s_{ij}$, $\tilde{I}_{ik,3}^{\text{mrc}} \triangleq \sum_{l=1,l\neq i}^{L}\sum_{j=1,j\neq k}^{K} \sqrt{p}\,\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}s_{lj}$, $\tilde{I}_{ik,4}^{\text{mrc}} \triangleq \sum_{l=1,l\neq i}^{L} \sqrt{p}\,\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}s_{lk}$, and $\tilde{I}_{ik,5}^{\text{mrc}} \triangleq n_{ik}$. The rate expression is given in (11) and the remaining part is finding powers of the terms in (11). First for the signal term in the numerator, by replacing $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$, we find $\mathbb{E}\{\sqrt{p}\,\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik}\} = \sqrt{p}\;\mathbb{E}\{\|\hat{\mathbf{g}}_{iik}\|\} = \sqrt{p}\sqrt{\sigma_{\hat{\mathbf{g}}_{iik}}^2}C_M$, where we have used the fact that $\|\hat{\mathbf{g}}_{iik}\|$ has a Chi distribution. Then, from independence of user's data, we can see $\mathbb{E}\{|\tilde{I}_{ik}^{\text{mrc}}|^2\} = \sum_{j=1}^{5}\mathbb{E}\{|\tilde{I}_{ik,j}^{\text{mrc}}|^2\}$. Also, $\mathbb{E}\{|\tilde{I}_{ik,1}^{\text{mrc}}|^2\} = p\;\text{var}\{\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\mathbf{g}_{iik}\} = p\quad(\text{var}\{\|\hat{\mathbf{g}}_{iik}\|\} + \sigma_{\boldsymbol{\varepsilon}_{iik}}^2) = p(V_M\sigma_{\hat{\mathbf{g}}_{iik}}^2 + \sigma_{\boldsymbol{\varepsilon}_{iik}}^2)$, where we have replaced $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$ and used the independence between $\hat{\mathbf{g}}_{iik}$ and $\boldsymbol{\varepsilon}_{iik}$ and Lemma 9. For computing power of the other interference terms, due to independence between $\mathbf{g}_{iik}$ and $\hat{\mathbf{g}}_{iij}(j \neq k)$, from Lemma 9 we have $\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik} \sim \mathcal{CN}(0, \beta_{iik})$. Similarly, we have $\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik} \sim \mathcal{CN}(0, \beta_{lik})$, $j \neq k$. But for the terms in $\tilde{I}_{ik,4}$, $\hat{\mathbf{g}}_{llk}$ and $\mathbf{g}_{lik}$ are correlated due to pilot contamination. To proceed further, we replace $\mathbf{g}_{lik}$ with $\hat{\mathbf{g}}_{lik} + \boldsymbol{\varepsilon}_{lik}$ and then $\hat{\mathbf{g}}_{lik}$ with $\frac{\beta_{lik}}{\beta_{llk}}\hat{\mathbf{g}}_{llk}$ from (6). As a result, we have $\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik} \sim (\frac{\beta_{lik}}{\beta_{llk}})\|\hat{\mathbf{g}}_{llk}\| + \mathcal{CN}(0, \sigma_{\boldsymbol{\varepsilon}_{lik}}^2)$. Then we obtain $\mathbb{E}\{|\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}|^2\} = (\frac{\beta_{lik}}{\beta_{llk}})^2 M\sigma_{\hat{\mathbf{g}}_{llk}}^2 + \sigma_{\boldsymbol{\varepsilon}_{lik}}^2 = M\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \sigma_{\boldsymbol{\varepsilon}_{lik}}^2 = (M-1)\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \beta_{lik}$, where we have used the fact that $\|\hat{\mathbf{g}}_{llk}\|^2$ has a Chi square distribution with $2M$ DoF, and $\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \sigma_{\boldsymbol{\varepsilon}_{lik}}^2 = \beta_{lik}$. Now, we can compute the powers of the remaining interference terms as $\mathbb{E}\{|\tilde{I}_{ik,2}^{\text{mrc}}|^2\} = p(K-1)\beta_{iik}$, $\mathbb{E}\{|\tilde{I}_{ik,3}^{\text{mrc}}|^2\} = p(K-1)\sum_{l=1,l\neq i}^{L}\beta_{lik}$, and $\mathbb{E}\{|\tilde{I}_{ik,4}^{\text{mrc}}|^2\} = p\sum_{l=1,l\neq i}^{L}(M-1)\sigma_{\hat{\mathbf{g}}_{lik}}^2 + \beta_{lik}$. By substituting the signal power and the interference power in (11), we arrive at (13).

## Appendix B
## Proof of Theorem 2

When users do not estimate their channels, we decompose the received signal in (9) as $r_{ik}^{\text{zf}}\mathbb{E}\{\sqrt{p}\,\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\mathbf{g}_{iik}\}s_{ik} + \tilde{I}_{ik}^{\text{zf}}$, where $\tilde{I}_{ik}^{\text{zf}} = \sum_{j=1}^{5}\tilde{I}_{ik,j}^{\text{zf}}$ and we define these different interference components as $\tilde{I}_{ik,1}^{\text{zf}} \triangleq \sqrt{p}(\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\mathbf{g}_{iik} - \mathbb{E}\{\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\mathbf{g}_{iik}\})s_{ik}$, $\tilde{I}_{ik,2}^{\text{zf}} \triangleq \sum_{j=1,j\neq k}^{K}\sqrt{p}\,\frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\mathbf{g}_{iik}s_{ij}$, $\tilde{I}_{ik,3}^{\text{zf}} \triangleq \sum_{l=1,l\neq i}^{L}\sum_{j=1,j\neq k}^{K}\sqrt{p}\,\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\mathbf{g}_{lik}s_{lj}$, $\tilde{I}_{ik,4}^{\text{zf}} \triangleq \sum_{l=1,l\neq i}^{L}\sqrt{p}\,\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik}s_{lk}$, and $\tilde{I}_{ik,5}^{\text{zf}} \triangleq n_{ik}$. The rate expression is given in (11) and the remaining part is finding powers of the terms in (11). From Sec. II-B, we recall for ZF precoding $\mathbf{a}_{ik}^H\hat{\mathbf{g}}_{iij} = \delta_{jk}$. First for the signal term in the numerator, by replacing $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$, we find $\mathbb{E}\{\sqrt{p}\,\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\mathbf{g}_{iik}\} = \sqrt{p}\quad\mathbb{E}\{\frac{1}{\|\mathbf{a}_{ik}\|}\} = \sqrt{p}\sqrt{\sigma_{\hat{\mathbf{g}}_{iik}}^2}C_{M-K+1}$, where we have used Lemma 4. Then, from independence of user's data, we can see $\mathbb{E}\{|\tilde{I}_{ik}^{\text{zf}}|^2\} = \sum_{j=1}^{5}\mathbb{E}\{|\tilde{I}_{ik,j}^{\text{zf}}|^2\}$. Also, $\mathbb{E}\{|\tilde{I}_{ik,1}|^2\} = p\quad\text{var}\{\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\mathbf{g}_{iik}\} = p\quad(\text{var}\{\|\frac{1}{\mathbf{a}_{ik}}\|\} + \sigma_{\boldsymbol{\varepsilon}_{iik}}^2) = p\left(V_{M-K+1}\sigma_{\hat{\mathbf{g}}_{iik}}^2 + \sigma_{\boldsymbol{\varepsilon}_{iik}}^2\right)$, where we have replaced $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$ and used the independence between $\hat{\mathbf{g}}_{iik}$ and $\boldsymbol{\varepsilon}_{iik}$ in conjunction with Lemma 9, and for computing variance of the first term we have used Lemma 4. For computing power of the other interference terms, from Lemma 9, we have $\frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\mathbf{g}_{iik} = \frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\boldsymbol{\varepsilon}_{iik} \sim \mathcal{CN}(0, \sigma_{\boldsymbol{\varepsilon}_{iik}}^2)$, $j \neq k$. Similarly, we have

$\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\mathbf{g}_{lik} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{\varepsilon}_{lik}})$, $j \neq k$. But for the terms in $\tilde{I}^{\mathrm{zf}}_{ik,4}$, $\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik} = \frac{\beta_{lik}}{\beta_{llk}}\frac{1}{\|\mathbf{a}_{lk}\|} + \frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\boldsymbol{\varepsilon}_{lik}$, where we have replaced $\mathbf{g}_{lik}$ with $\hat{\mathbf{g}}_{lik} + \boldsymbol{\varepsilon}_{lik}$ and then $\hat{\mathbf{g}}_{lik}$ with $\frac{\beta_{lik}}{\beta_{llk}}\hat{\mathbf{g}}_{llk}$ from (6). So we obtain $\mathbb{E}\{|\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik}|^2\} = (M-K+1)\sigma^2_{\hat{\mathbf{g}}_{lik}} + \sigma^2_{\boldsymbol{\varepsilon}_{lik}}$, where we have used Lemma 10. Now, we can compute the powers of the remaining interference terms as $\mathbb{E}\{|\tilde{I}^{\mathrm{zf}}_{ik,2}|^2\} = p(K-1)\sigma^2_{\boldsymbol{\varepsilon}_{iik}}$, $\mathbb{E}\{|\tilde{I}^{\mathrm{zf}}_{ik,3}|^2\} = p(K-1)\sum_{l=1,l\neq i}^{L}\sigma^2_{\boldsymbol{\varepsilon}_{lik}}$, and $\mathbb{E}\{|\tilde{I}^{\mathrm{zf}}_{ik,4}|^2\} = p\sum_{l=1,l\neq i}^{L}(M-K+1)\sigma^2_{\hat{\mathbf{g}}_{lik}} + \sigma^2_{\boldsymbol{\varepsilon}_{lik}}$. By substituting the signal power and the interference power in (11), we arrive at (16).

## Appendix C
## Proof of Lemma 1

Following (6), we can write $\mathbf{g}_{lik}$ as

$$\mathbf{g}_{lik} = \hat{\mathbf{g}}_{lik} + \boldsymbol{\varepsilon}_{lik} = \frac{\beta_{lik}}{\beta_{llk}}\hat{\mathbf{g}}_{llk} + \boldsymbol{\varepsilon}_{lik}. \tag{79}$$

Then substituting (79) into (25), we have

$$w^{\mathrm{mrc}}_{ik} = \sum_{l=1}^{L}\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\boldsymbol{\varepsilon}_{lik} + \sum_{l=1,l\neq i}^{L}\frac{\beta_{lik}}{\beta_{llk}}\|\hat{\mathbf{g}}_{llk}\| + \mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}\right). \tag{80}$$

Due to the independence between $\hat{\mathbf{g}}_{llk}$ and $\boldsymbol{\varepsilon}_{lik}$ for $l = 1, \cdots, L$, and the Gaussian distribution of $\boldsymbol{\varepsilon}_{lik}$, from Lemma 9 we have $\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\boldsymbol{\varepsilon}_{lik} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{\varepsilon}_{lik}})$. Different terms in (80) are independent, and due to circularly symmetric distribution of complex random variables in (80), the real and imaginary parts are independent. Thus, we have

$$w^{\mathrm{mrc,im}}_{ik} \sim \mathcal{N}\left(0, \frac{1}{2}\sum_{l=1}^{L}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} + \frac{1}{2\tau_{\mathrm{d}}q_{\mathrm{d}}}\right). \tag{81}$$

The $\|\hat{\mathbf{g}}_{llk}\|$ has a Chi pdf with $2M$ DoF, $\mathbb{E}\{\|\hat{\mathbf{g}}_{llk}\|\} = C_M\sqrt{\sigma^2_{\hat{\mathbf{g}}_{llk}}}$ and $\mathrm{var}\{\|\hat{\mathbf{g}}_{llk}\|\} = V_M\sigma^2_{\hat{\mathbf{g}}_{llk}}$. In addition for a large $M$, its pdf approaches a Gaussian pdf [29]. Then, for a large $M$, we have

$$w^{\mathrm{mrc,re}}_{ik} \sim \mathcal{N}\left(\mathbb{E}\{w^{\mathrm{mrc,re}}_{ik}\}, \mathrm{var}\{w^{\mathrm{mrc,re}}_{ik}\}\right) \tag{82}$$

where $\mathbb{E}\{w^{\mathrm{mrc,re}}_{ik}\}$ and $\mathrm{var}\{w^{\mathrm{mrc,re}}_{ik}\}$ are given in (86) and (28).

## Appendix D
## Proof of Lemma 2

As $\hat{\alpha}^{\mathrm{mrc}}_{ik} = \|\hat{\mathbf{g}}_{iik}\|$ follows a Chi distribution with $2M$ DoF, its mean and variance are straightly given by (30) and (31) and its pdf approaches a Gaussian pdf [29].

## Appendix E
## Proof of Lemma 3

Following the independence between $\alpha_{ik}$ and $w^{\mathrm{re}}_{ik}$, and applying the LMMSE principle [27], the estimator of $\alpha_{ik}$ can be derived as

$$\hat{\alpha}_{ik} = \mathbb{E}\{\alpha_{ik}\} + \frac{\mathrm{var}\{\alpha_{ik}\}\left(z^{\mathrm{re}}_{ik} - \mathbb{E}\{\alpha_{ik}\} - \mathbb{E}\left\{w^{\mathrm{re}}_{ik}\right\}\right)}{\mathrm{var}\{\alpha_{ik}\} + \mathrm{var}\left\{w^{\mathrm{re}}_{ik}\right\}}. \tag{83}$$

The MSE of the above estimation reads as

$$\eta_{ik} \triangleq \mathbb{E}\left\{\left|\alpha_{ik} - \hat{\alpha}_{ik}\right|^2\right\} = \frac{\mathrm{var}\{\alpha_{ik}\}\cdot\mathrm{var}\left\{w^{\mathrm{re}}_{ik}\right\}}{\mathrm{var}\{\alpha_{ik}\} + \mathrm{var}\left\{w^{\mathrm{re}}_{ik}\right\}}. \tag{84}$$

So by substituting the mean and variance terms, we arrive at the results in Lemmas 3 and 6.

Due to asymptotically (with respect to $M$) Gaussian distribution of $\alpha_{ik}$ and $w_{ik}$, for a large $M$, LMMSE estimator becomes MMSE estimator and $\hat{\alpha}_{ik}$ has Gaussian distribution, and $\hat{\alpha}_{ik}$ and $v_{ik}$ become independent.

## Appendix F
## Proof of Lemma 4

For the proof of Lemma 4 and 5, we use the following Lemma.

*Lemma 10:* For independent random vectors $\{\mathbf{x}_i\}_{i=1}^K$ with $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_i^2\mathbf{I}_M)$, define $\mathbf{X}$ as an $M \times K$ random matrix $(M \geq K)$ with its $i$th column being $\mathbf{x}_i$ and $\mathbf{y}_i$ as $i$th column of $\mathbf{X}\left(\mathbf{X}^H\mathbf{X}\right)^{-1}$. Then $\mathbb{E}\{\frac{1}{\|\mathbf{y}_i\|^2}\} = \sigma_i^2(M-K+1)$. In addition, the pdf of $\frac{1}{\|\mathbf{y}_i\|}$ is given by

$$f_{\frac{1}{\|\mathbf{y}_i\|}}(t) = \frac{2\,t^{2(M-K)+1}e^{-\frac{t^2}{\sigma_i^2}}}{\left(\sigma_i^2\right)^{M-K+1}\Gamma(M-K+1)},\ t > 0. \tag{85}$$

*Proof:* We have $\|\mathbf{y}_i\|^2 = [(\mathbf{X}^H\mathbf{X})^{-1}]_{ii}$. We can write $\mathbf{X} = \mathbf{H}\Lambda$, where $\mathbf{H}$ is an $M \times K$ matrix of i.i.d. $\mathcal{CN}(0,1)$ elements, and $\Lambda = \mathrm{diag}\{\sigma_1, \cdots, \sigma_K\}$. So $[(\mathbf{X}^H\mathbf{X})^{-1}]_{ii} = \sigma_i^{-2}[(\mathbf{H}^H\mathbf{H})^{-1}]_{ii}$. Let $v_i \triangleq ([\left(\mathbf{H}^H\mathbf{H}\right)^{-1}]_{ii})^{-1}$. Then $v_i$ has a Gamma distribution with pdf $f_{v_i}(v) = \frac{v^{M-K}e^{-v}}{\Gamma(M-K+1)}$, $v > 0$, [30], and $\mathbb{E}\{v_i\} = (M-K+1)$. As $\frac{1}{\|\mathbf{y}_i\|^2} = \sigma_i^2 v_i$, we have $\mathbb{E}\{\frac{1}{\|\mathbf{y}_i\|^2}\} = \sigma_i^2\mathbb{E}\{v_i\} = \sigma_i^2(M-K+1)$. Next, from $\frac{1}{\|\mathbf{y}_i\|} = \sigma_i\sqrt{v_i}$ and the pdf of $v_i$, we can straightly obtain the pdf of $\frac{1}{\|\mathbf{y}_i\|}$ which turns out to be a Chi pdf. ∎

By following Lemma 10 and replacing $\{\mathbf{x}_i\}_{i=1}^K$ with $\{\hat{\mathbf{g}}_{llk}\}_{k=1}^K$, the pdf of $\frac{1}{\|\mathbf{a}_{lk}\|}$ is given by a Chi pdf as

$$f_{\frac{1}{\|\mathbf{a}_{lk}\|}}(t) = \frac{2\,t^{2(M-K)+1}e^{-\frac{t^2}{\sigma^2_{\hat{\mathbf{g}}_{llk}}}}}{\left(\sigma^2_{\hat{\mathbf{g}}_{llk}}\right)^{M-K+1}\Gamma(M-K+1)},\ t > 0. \tag{86}$$

Next, by direct calculation, the mean and variance of $\frac{1}{\|\mathbf{a}_{lk}\|}$ read as in (41) and (42).

## Appendix G
## Proof of Lemma 5

By substituting (79) into (40) and simplifying the expression, we have

$$w^{\mathrm{zf}}_{ik} = \sum_{l=1}^{L}\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\boldsymbol{\varepsilon}_{lik} + \sum_{l=1,l\neq i}^{L}\frac{\beta_{lik}}{\beta_{llk}}\frac{1}{\|\mathbf{a}_{lk}\|} + \mathcal{CN}\left(0, \frac{1}{\tau_{\mathrm{d}}q_{\mathrm{d}}}\right). \tag{87}$$

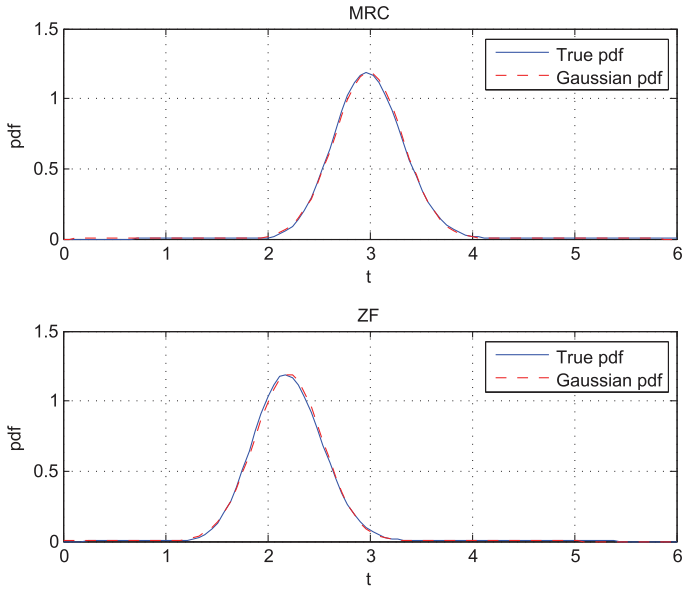



Fig. 11. Comparison of the approximate and actual pdfs of $\alpha_{ik}^{\text{mrc}}$ (upper) and $\alpha_{ik}^{\text{zf}}$ (lower) with $M = 20$ antennas, $K = 10$ users, and pilot powers $q_{\text{u}} = 10$ dB.

$\mathbf{a}_{lk}$ depends on $\hat{\mathbf{G}}_{ll}$ due to its definition but it is independent from $\boldsymbol{\varepsilon}_{lik}$, and due to the Gaussian distribution of $\boldsymbol{\varepsilon}_{lik}$, from Lemma 9 we have $\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\boldsymbol{\varepsilon}_{lik} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{\varepsilon}_{lik}})$ and all different terms are independent. Then the same as in the proof of Lemma 1, the real and imaginary parts of $w_{ik}^{\text{zf}}$ are independent, and we have

$$w_{ik}^{\text{zf,im}} \sim \mathcal{N}\left(0, \frac{1}{2}\sum_{l=1}^{L}\sigma^2_{\boldsymbol{\varepsilon}_{lik}} + \frac{1}{2\tau_{\text{d}}q_{\text{d}}}\right). \tag{88}$$

Next, using (41) and (42), we have $\mathbb{E}\{w_{ik}^{\text{zf,re}}\}$ and $\text{var}\{w_{ik}^{\text{zf,re}}\}$ as given in (43) and (45), respectively. In addition, for a large value of $M - K + 1$, $\frac{1}{\|\mathbf{a}_{lk}\|}$ approaches a Gaussian random variable [29]. So, for a large $M - K + 1$, we have

$$w_{ik}^{\text{zf,re}} \sim \mathcal{N}\left(\mathbb{E}\left\{w_{ik}^{\text{zf,re}}\right\}, \text{var}\{w_{ik}^{\text{zf,re}}\}\right). \tag{89}$$

*Remark 9:* Although the Gaussian approximations in Lemma (2) and (4) are described for a large $M$, we notice in our investigation that they also hold for a small $M$. To illustrate this, we present in Fig. 11 the pdfs of the actual (Chi) pdf and the approximate (Gaussian) pdf of $\alpha_{ik}$ for a system with $M = 20$ antennas, $K = 10$ users and pilot power of $q_{\text{u}} = 10$ dB. The pdfs show a close match even for a small $M$.

## Appendix H
## Proof of Lemma 7

We can decompose the interference term in (19) into six components as

$$I_{ik}^{\text{mrc}} = I_{ik,1}^{\text{mrc}} + I_{ik,2}^{\text{mrc}} + I_{ik,3}^{\text{mrc}} + I_{ik,4}^{\text{mrc}} + I_{ik,5}^{\text{mrc}} + n_{ik} \tag{90}$$

where $I_{ik,1}^{\text{mrc}} \triangleq \sqrt{p}\nu_{ik}^{\text{mrc}}s_{ik}$, $I_{ik,2}^{\text{mrc}} \triangleq \sqrt{p}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}s_{ik}$, $I_{ik,3}^{\text{mrc}} \triangleq \sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}s_{ij}$, $I_{ik,4}^{\text{mrc}} \triangleq \sum_{l=1,l\neq i}^{L}\sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}s_{lj}$, and $I_{ik,5}^{\text{mrc}} \triangleq \sum_{l=1,l\neq i}^{L}\sqrt{p}\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}s_{lk}$.

From Lemma 3 and Remark 3, $\nu_{ik}^{\text{mrc}}$ and $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}$ are asymptotically independent from $\hat{\alpha}_{ik}$. Also $\nu_{ik}^{\text{mrc}}$ and $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}$ are asymptotically independent. As a result, we have $\mathbb{E}\{(I_{ik,1}^{\text{mrc}})(I_{ik,2}^{\text{mrc}})^*|\hat{\alpha}_{ik}\} \approx 0$. The other pairs of interference terms in (90) are also uncorrelated due to uncorrelated $\{s_{lj}\}$. As a result, based on (90), we can write the conditional average interference power as

$$\mathbb{E}\left\{\left|I_{ik}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} \approx \sum_{j=1}^{5}\mathbb{E}\left\{\left|I_{ik,j}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} + 1. \tag{91}$$

In the following, we find the conditional power of each interference component.

For the first component, using the result of Lemma 3 that for a large $M$ the estimation error $\nu_{ik}^{\text{mrc}}$ and the estimate $\hat{\alpha}_{ik}^{\text{mrc}}$ approach to be independent, we obtain

$$\mathbb{E}\left\{\left|I_{ik,1}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} = p\,\mathbb{E}\left\{\left|\nu_{ik}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} \approx p\,\eta_{ik}^{\text{mrc}}. \tag{92}$$

For the second component, from Lemma 9 we use the result that $\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{\varepsilon}_{iik}})$ and obtain

$$\mathbb{E}\left\{\left|I_{ik,2}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} = p\,\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\boldsymbol{\varepsilon}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} \approx p\sigma^2_{\boldsymbol{\varepsilon}_{iik}}. \tag{93}$$

For the third component, due to uncorrelated $\{s_{ij}, j = 1, \cdots, K\}$, we have

$$\mathbb{E}\left\{\left|I_{ik,3}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} = \sum_{j=1,j\neq k}^{K} p\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}\Bigg\}. \tag{94}$$

In the following, we compute $\mathbb{E}\{|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}|^2|\hat{\alpha}_{ik}^{\text{mrc}}\}$. Note that $\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}$ and $\hat{\alpha}_{ik}^{\text{mrc}}$ are dependent. By substituting $\mathbf{g}_{iik}$ with $\hat{\mathbf{g}}_{iik} + \boldsymbol{\varepsilon}_{iik}$, we have

$$\begin{aligned}\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik} &= \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\hat{\mathbf{g}}_{iik} + \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\boldsymbol{\varepsilon}_{iik}\\ &= \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\hat{\alpha}_{ik}^{\text{mrc}} + \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\nu_{ik}^{\text{mrc}}\\ &\quad + \frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\boldsymbol{\varepsilon}_{iik}\end{aligned} \tag{95}$$

where we have used $\|\hat{\mathbf{g}}_{iik}\| = \alpha_{ik}^{\text{mrc}} = \hat{\alpha}_{ik}^{\text{mrc}} + \nu_{ik}^{\text{mrc}}$ in (95). The third term in (95), i.e., $\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\boldsymbol{\varepsilon}_{iik}$, is independent from $\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}$ because $\hat{\mathbf{g}}_{iij}$ and $\boldsymbol{\varepsilon}_{iik}$ are independent and $\boldsymbol{\varepsilon}_{iik}$ has a circularly symmetric distribution [31, Lemma 4]. In addition, conditioned

on $\hat{\alpha}_{ik}^{\text{mrc}}$, the three terms in (95) are mutually asymptotically uncorrelated (see Remark 3, and Lemma 3), and we can write (95) as

$$\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\} \approx \mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\hat{\alpha}_{ik}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}$$
$$+\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}\nu_{ik}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}+\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\boldsymbol{\varepsilon}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}. \tag{96}$$

To proceed further, we introduce the following Lemma.

*Lemma 11:* If $\mathbf{x}\sim\mathcal{CN}(0,\sigma_x^2\mathbf{I}_M)$ and $\mathbf{y}\sim\mathcal{CN}(0,\sigma_y^2\mathbf{I}_M)$, then $z\triangleq|\frac{\mathbf{x}^H}{\|\mathbf{x}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|}|^2$ has a beta distribution with parameter 1 and $M-1$, and its pdf is $f_z(u)=(M-1)(1-u)^{M-2}$ for $u\in[0,1]$ with $\mathbb{E}\{z\}=\frac{1}{M}$.

*Proof:* See [32]. ∎

From Lemma 11 with $\mathbf{x}=\hat{\mathbf{g}}_{iij}$ and $\mathbf{y}=\hat{\mathbf{g}}_{iik}$, we obtain $\mathbb{E}\{|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\frac{\hat{\mathbf{g}}_{iik}^H}{\|\hat{\mathbf{g}}_{iik}\|}|^2\}=\frac{1}{M}$. Then we have

$$\mathbb{E}\left\{\left|\frac{\hat{\mathbf{g}}_{iij}^H}{\|\hat{\mathbf{g}}_{iij}\|}\mathbf{g}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}\approx\frac{|\hat{\alpha}_{ik}^{\text{mrc}}|^2}{M}+\frac{\eta_{ik}^{\text{mrc}}}{M}+\sigma_{\boldsymbol{\varepsilon}_{iik}}^2 \tag{97}$$

and by substituting (97) in (94), we obtain

$$\mathbb{E}\left\{\left|I_{ik,3}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}\approx p\frac{K-1}{M}\left|\hat{\alpha}_{ik}^{\text{mrc}}\right|^2+p\frac{K-1}{M}\hat{\eta}_{ik}^{\text{mrc}}$$
$$+p(K-1)\sigma_{\boldsymbol{\varepsilon}_{iik}}^2. \tag{98}$$

Next, in computing $\mathbb{E}\{|I_{ik,4}^{\text{mrc}}|^2|\hat{\alpha}_{ik}^{\text{mrc}}\}$, we use the fact that $\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}$ is independent from $\hat{\alpha}_{ik}^{\text{mrc}}$. In addition, due to the independence between $\hat{\mathbf{g}}_{llj}$ and $\mathbf{g}_{lik}$ for $j\neq k$, from Lemma 9 we have $\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}\sim\mathcal{CN}(0,\beta_{lik})$. Then we obtain

$$\mathbb{E}\left\{\left|I_{ik,4}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}=p(K-1)\sum_{l=1,l\neq i}^{L}\beta_{lik}. \tag{99}$$

But, in computing $\mathbb{E}\{|I_{ik,5}^{\text{mrc}}|^2|\hat{\alpha}_{ik}^{\text{mrc}}\}$, $\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}$ is dependent on $\hat{\alpha}_{ik}^{\text{mrc}}$. In particular, when we expand $\hat{\alpha}_{ik}^{\text{mrc}}$ in (33) by replacing $w_{ik}^{\text{mrc,re}}$ with the real part of (25), we see $\Re\{\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}|^2\}$ and $\hat{\alpha}_{ik}^{\text{mrc}}$ are dependent while $\Im\{\frac{\hat{\mathbf{g}}_{llk}^H}{\|\hat{\mathbf{g}}_{llk}\|}\mathbf{g}_{lik}|^2\}$ and $\hat{\alpha}_{ik}^{\text{mrc}}$ are independent. Next, considering expression of $\hat{\alpha}_{ik}^{\text{mrc}}$ and using the asymptotically Gaussian distribution of a Chi random variable, we can find $\mathbb{E}\{(\Re\{\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}\})^2|\hat{\alpha}_{ik}^{\text{mrc}}\}$, a conditional second moment, from the sum of square of the conditional mean [27, eqn. 14.6] and conditional variance [27, eqn. 14.8]. After algebraic manipulations and simplification, $\mathbb{E}\{|\frac{\hat{\mathbf{g}}_{llj}^H}{\|\hat{\mathbf{g}}_{llj}\|}\mathbf{g}_{lik}|^2|\hat{\alpha}_{ik}^{\text{mrc}}\}$ can be found as $\zeta_{ilk}^{\text{mrc}}$ as given in (61).

With the above result, we obtain

$$\mathbb{E}\left\{\left|I_{ik,5}^{\text{mrc}}\right|^2\left|\hat{\alpha}_{ik}^{\text{mrc}}\right.\right\}=p\sum_{l=1,l\neq i}^{L}\zeta_{ilk}^{\text{mrc}}. \tag{100}$$

Finally, by substituting (92), (93), (98), (99), (100) into (91), we complete the proof.

## Appendix I
## Proof of Lemma 8

We can decompose the interference term in (21) into six components as

$$I_{ik}^{\text{zf}}=I_{ik,1}^{\text{zf}}+I_{ik,2}^{\text{zf}}+I_{ik,3}^{\text{zf}}+I_{ik,4}^{\text{zf}}+I_{ik,5}^{\text{zf}}+n_{ik} \tag{101}$$

where $I_{ik,1}^{\text{zf}}\triangleq\sqrt{p}\nu_{ik}^{\text{zf}}s_{ik}$, $I_{ik,2}^{\text{zf}}\triangleq\sqrt{p}\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik}s_{ik}$, $I_{ik,3}^{\text{zf}}\triangleq\sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\mathbf{g}_{iik}s_{ij}$, $I_{ik,4}^{\text{zf}}\triangleq\sum_{l=1,l\neq i}^{L}\sum_{j=1,j\neq k}^{K}\sqrt{p}\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\mathbf{g}_{lik}s_{lj}$, and $I_{ik,5}^{\text{zf}}\triangleq\sum_{l=1,l\neq i}^{L}\sqrt{p}\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik}s_{lk}$. In the same way of proving Lemma 7, we have

$$\mathbb{E}\left\{\left|I_{ik}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}\approx\sum_{j=1}^{5}\mathbb{E}\left\{\left|I_{ik,j}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}+1. \tag{102}$$

For the first and second components, similar to the proof of Lemma 7, we have

$$\mathbb{E}\left\{\left|I_{ik,1}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}=p\,\mathbb{E}\left\{\left|\nu_{ik}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}\approx p\,\eta_{ik}^{\text{zf}} \tag{103}$$

$$\mathbb{E}\left\{\left|I_{ik,2}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}=p\mathbb{E}\left\{\left|\frac{\mathbf{a}_{ik}^H}{\|\mathbf{a}_{ik}\|}\boldsymbol{\varepsilon}_{iik}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}\approx p\sigma_{\boldsymbol{\varepsilon}_{iik}}^2. \tag{104}$$

Next, as $\mathbf{a}_{ij}$ and $\boldsymbol{\varepsilon}_{iik}$ are independent, from Lemma 9 we have $\frac{\mathbf{a}_{ij}^H}{\|\mathbf{a}_{ij}\|}\boldsymbol{\varepsilon}_{iik}\sim\mathcal{CN}(0,\sigma_{\boldsymbol{\varepsilon}_{iik}}^2)$, and together with Remark 3, we obtain

$$\mathbb{E}\left\{\left|I_{ik,3}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}\approx p(K-1)\sigma_{\boldsymbol{\varepsilon}_{iik}}^2. \tag{105}$$

For the fourth component, $\mathbf{a}_{lj}$ and $\hat{\alpha}_{ik}$ are independent. We use $\mathbf{g}_{lik}=\hat{\mathbf{g}}_{lik}+\boldsymbol{\varepsilon}_{lik}$ and replace $\hat{\mathbf{g}}_{lik}$ with $\frac{\beta_{lik}}{\beta_{llk}}\hat{\mathbf{g}}_{llk}$. Then, for the fourth component, from Lemma 9 we have $\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\mathbf{g}_{lik}=\frac{\mathbf{a}_{lj}^H}{\|\mathbf{a}_{lj}\|}\boldsymbol{\varepsilon}_{lik}\sim\mathcal{CN}(0,\sigma_{\boldsymbol{\varepsilon}_{lik}}^2)$. As a result, we obtain

$$\mathbb{E}\left\{\left|I_{ik,4}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}=p(K-1)\sum_{l=1,l\neq i}^{L}\sigma_{\boldsymbol{\varepsilon}_{lik}}^2. \tag{106}$$

For the fifth component, expanding $\hat{\alpha}_{ik}^{\text{zf}}$ in (48) by replacing $w_{ik}^{\text{zf,re}}$ with the real part of (40), we find $\Re\{\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik}\}$ and $\hat{\alpha}_{ik}^{\text{zf}}$ are dependent. With the same approach of Lemma 7, we can find the conditional second moment $\mathbb{E}\{|\frac{\mathbf{a}_{lk}^H}{\|\mathbf{a}_{lk}\|}\mathbf{g}_{lik}|^2|\hat{\alpha}_{ik}^{\text{zf}}\}=\zeta_{ilk}^{\text{zf}}$, where $\zeta_{ilk}^{\text{zf}}$ is given in (70) which yields

$$\mathbb{E}\left\{\left|I_{ik,5}^{\text{zf}}\right|^2\left|\hat{\alpha}_{ik}^{\text{zf}}\right.\right\}=p\sum_{l=1,l\neq i}^{L}\zeta_{ilk}^{\text{zf}}. \tag{107}$$

By substituting (103), (104), (105), (106) and (107) into (102), we complete the proof.

## REFERENCES


[1] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[8] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[9] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[10] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[11] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna TDD systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2102–2116, Jun. 2011.

[12] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[13] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[14] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An esprit-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[15] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: New design methods and analysis," *IEEE Access*, vol. 2, pp. 947–959, Aug. 2014.

[16] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[17] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[18] Y.-C. Liang, S. Sun, and C. K. Ho, "Block-iterative generalized decision feedback equalizers for large MIMO systems: Algorithm design and asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2035–2048, Jun. 2006.

[19] N. Srinidhi, T. Datta, A. Chockalingam, and B. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[20] Q. Zhou and X. Ma, "Element-based lattice reduction algorithms for large MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 274–286, Feb. 2013.

[21] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[22] H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multiuser MU-SIMO systems with ZF receivers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4471–4483, Nov. 2013.

[23] A. Khansefid and H. Minn, "Performance bounds for massive MIMO uplink," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 632–636.

[24] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. 51st Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2013, pp. 293–298.

[25] A. Khansefid and H. Minn, "On channel estimation for massive MIMO with pilot contamination," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1660–1663, Sep. 2015.

[26] A. Khansefid and H. Minn, "Asymptotically optimal power allocation for massive MIMO uplink," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 627–631.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ, USA: Pearson Education, 1993, vol. 1.

[28] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

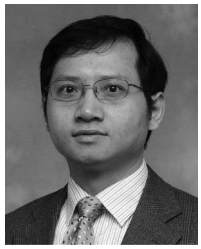
[29] D. Jensen, "Limit properties of noncentral multivariate Rayleigh and Chi-square distributions," *SIAM J. Appl. Math.*, vol. 17, no. 4, pp. 802–814, Jul. 1969.

[30] D. K. Nagar and A. K. Gupta, "Expectations of functions of complex Wishart matrix," *Acta Appl. Math.*, vol. 113, no. 3, pp. 265–288, 2011.

[31] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[32] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.


**Amin Khansefid** (S'06) received the B.S. degrees in electrical engineering and mathematics, both from Isfahan University of Technology, Isfahan, Iran, in 2005, and the M.S. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2008. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Texas at Dallas, Richardson, TX, USA. He was an RF and Electronic Designer with Isfahan University of Technology from 2008 to 2010, and a Research Associate with the Department of Electrical and Computer Engineering, University of Tehran from 2010 to 2012. His research interests include communication systems and signal processing.



**Hlaing Minn** (S'99–M'01–SM'07) received the B.E. degree in electrical engineering (electronics) from Yangon Institute of Technology, Yangon, Myanmar, in 1995, the M.Eng. degree in telecommunications from Asian Institute of Technology, Pathumthani, Thailand, in 1997, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2001. From January to August 2002, he was a Postdoctoral Fellow with the University of Victoria. Since September 2002, he has been with the Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX, USA, where he is currently a Full Professor. His research interests include wireless communications, statistical signal processing, signal design, cross-layer design, cognitive radios, dynamic spectrum access and sharing, energy efficient wireless systems, next-generation wireless technologies, and wireless healthcare applications. He is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the *International Journal of Communications and Networks* (Wiley). He has served as a Technical Program Co-Chair for the Wireless Communications Symposium of the IEEE Global Communications Conference (GLOBECOM 2014) and the Wireless Access Track of the IEEE Vehicular Technology Conference (VTC 2009"Fall), and a Technical Program Committee Member for several IEEE conferences.