Performance Analysis of Massive MIMO Networks with Random Unitary Pilot Matrices

Rusdha Muharar

Department of Electrical and Computer Engineering
Universitas Syiah Kuala, Indonesia

Jamie Evans

Department of Electrical and Electronic Engineering,
The University of Melbourne, Australia

Abstract—A common approach to obtain channel state information for massive MIMO networks is to use the same orthogonal training sequences in each cell. We call this the full-pilot reuse (FPR) scheme. In this paper, we study an alternative approach where each cell uses different sets of orthogonal pilot (DOP) sequences. Considering uplink communications with matched filter (MF) receivers, we first derive the SINR in the large system regime where the number of antennas at the base station, the number of users in each cell, and training duration grow large with fixed ratios. For tractability in the analysis, the orthogonal pilots are drawn from Haar distributed random unitary matrices. The resulting expression is simple and easy to compute. As shown by the numerical simulations, the asymptotic SINR approximates the finite-size systems accurately. Secondly, we derive the user capacity of the DOP scheme under a simple power control and show that it is generally better than that of the FPR scheme.

I. Introduction

One of the candidates for 5G technology is the massive Multiple-Input and Multiple-Output (MIMO) system (see e.g., [1], [2]) introduced by Marzetta in [3]. In massive MIMO cellular networks, a large number of small and low-cost antennas, in the order of hundreds, is employed at base stations (BSs). This enables an aggressive spatial multiplexing which can lead to a ten times capacity increase compared to conventional MIMO systems [2], [3].

A common approach in the uplink training of massive MIMO systems is the full-pilot reuse (FPR) where the same orthogonal training sequences are used in each cell, see e.g., [3], [4], [5]. The other approach proposed in [3] is to use different orthogonal pilots in different cells and we denote this approach as the DOP scheme. To the best of our knowledge, this scheme is largely unexplored.

It is argued in [3] that this scheme gives little difference in terms of the achieved SINRs compared to the FPR scheme. However, in the DOP, we get a small amount of contamination from all inter-cell users rather than a (potentially) large amount from a few users (those with the same pilot sequence) and this can lead to a better user capacity (see [6]). Note that the analysis in [3], [6] is performed in the regime where the number of antennas (N) tends to infinity and the number users (K) is finite ($K \ll N$). This implies that the cell-loading (K/N) is close to zero. It should be noted that the analysis in this regime gives a loose approximation for finite-size systems and can also converge slowly [4], [7].

In this paper, we generalize the performance analysis of the DOP scheme for arbitrary numbers of cell-loading. We obtain the approximation of the SINR by performing the analysis in the large system regime where the number of antennas at the base station, the number of users in each cell, and the number of training symbols go to infinity with fixed ratios. For analytical tractability, we choose the training pilots from *Haar*-distributed random unitary matrices. This approach has been used previously in CDMA systems, see for example [8] and [9]. Our numerical simulations show that the asymptotic results approximate the finite-size systems accurately. In the analysis, we show that the pilot contamination in the asymptotic SINR expression is the *average* of the square of received powers of *all users* from the interfering cells (see also [3] for a similar conclusion). This result differs from that obtained in the FPR case [3], [4], [5], where the pilot contamination is the sum of the received power of users from the interfering cells that use the *same* training sequence.

In this paper, we also consider another performance criterion, i.e., the user capacity. In the downlink with maximum ratio transmission (MRT) precoders, the user capacities of massive MIMO networks for single and multi-cell scenarios have been characterized in [10], [11], respectively. Recent work [6] studies the uplink user capacity when the cell-loading approaches zero. Here, we derive the uplink user capacity for arbitrary numbers of cell-loading under a simple power control where the uplink transmit power of a user is the inverse of the slow path-gain of that user (see also [12]). Our numerical simulations show that even though all users from the interfering cells contribute to the pilot contamination in the DOP scheme, its user capacity can be larger compared to that of FPR. Other related work is [12] that investigates the optimal number of users that maximizes the spectral efficiency of massive MIMO networks.

The following notations are used in this paper. The boldface lower and upper case letters denote vectors and matrices, respectively. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\mathbb{E}[\cdot]$ and $\xrightarrow{a.s.}$ denotes respectively the statistical expectation and the almost sure convergence. The circularly symmetric complex Gaussian (CSCG) vector with zero mean and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$. $|a|$ and $\Re[a]$ denote the magnitude and the real part of the complex variable a, respectively. $\|\cdot\|$ represents the Euclidean norm. $\mathrm{Tr}\,(\cdot)$, $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ refer to the trace, transpose and Hermitian transpose, of a matrix respectively.

II. System Model

We consider a multi-cell communication system with L-cells. Each cell has a base station equipped with N antennas and the number of users in cell i is denoted by K_i. The channel between user k in cell j and the BS in cell i is denoted by the column vector $\mathbf{g}_{kji}$ that can be modeled as

$$\mathbf{g}_{kji} = \sqrt{\ell_{kji}}\mathbf{h}_{kji} \qquad (1)$$

where $\mathbf{h}_{kji} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ represents the fast-fading channel coefficients. It is assumed that the slow-fading coefficient ℓ_{kji} is distance-dependent and the shadowing effect is ignored. The channel variations follow the block channel fading model and one block duration is equal to the channel coherence time.

We also consider the time-division duplex (TDD) protocol with perfect channel reciprocity between the uplink and downlink channels. In this paper, we focus on the uplink transmission where all scheduled users transmit simultaneously to their base station (BS). In the pilot-based TDD training, the BS estimates each user channel from the pilot symbols sent by each user in the uplink transmission. The BS uses the channel information to decode the transmitted symbols from its users. We should note that the uplink training and the uplink data transmissions should occur in one coherence block time.

A. Uplink Training

Let T be the length of the channel coherence time (in symbols). Let $\tau \leq T$ be the uplink training interval or the number of training symbols. In this training phase, each user in each cell sends the pilot symbols to their BS. We assume a synchronized training, where all cells perform the training at the same time and with the same training period τ. Let $h_{kji}^{[n]}$ be the n-th element of $\mathbf{h}_{kji}$ that represents the channel from the corresponding user to the n-th antenna of BS i. Also, let $q_{ki}^{(t)}$ be the pilot symbol sent by user k in cell i at time t and ϱ_{ki} be the corresponding average training power. Note that, in the uplink training, BS i will estimate the channels of its users $\mathbf{h}_{kii}$. Since our assumption that the elements of $\mathbf{h}_{kii}$ are independent, therefore we can estimate each element of the channel vector $\mathbf{h}_{kii}$ independently. The received signal vector at the n-th antenna of BS for τ training symbols is

$$\mathbf{y}_i^{[n]} = \sum_{j=1}^{L}\sum_{k=1}^{K_j} \sqrt{\rho_{kji}}h_{kji}^{[n]}\mathbf{q}_{kj} + \mathbf{n}_i^{[n]} \qquad (2)$$

where $\mathbf{y}_i^{[n]} = [y_{1i}^{[n]},\ldots,y_{\tau i}^{[n]}]^\mathsf{T}$, $\mathbf{q}_{kj} = [q_{kj}^{(1)},\ldots,q_{kj}^{(\tau)}]^\mathsf{T}$, and $\mathbf{n}_i^{[n]} = [n_{1i}^{[n]},\ldots,n_{\tau i}^{[n]}]^\mathsf{T} \sim \mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I}_\tau)$ is the receiver noise vector. We also denote $\rho_{kji} = \varrho_{kj}\ell_{kji}$ as the received power of user k from cell j at BS i. From (2), we can see that BS i also receives the training transmissions from other cells. It is assumed that the BS knows all the path-gains and training sequences of all users perfectly. We also assume that the training matrix for cell j, $\mathbf{Q}_j = [\mathbf{q}_{1j},\mathbf{q}_{2j},\cdots,\mathbf{q}_{K_jj}]$ is obtained by extracting $K_j \leq \tau$ columns of a $\tau \times \tau$ *Haar*-distributed random unitary matrix $\mathfrak{U}_j$ (see also [8], [9]). Thus, $\mathbf{Q}_j^\mathsf{H}\mathbf{Q}_j = \mathbf{I}_K$ which implies that the training sequences are orthogonal (orthonormal) across the users in the same cell. It is assumed that $\mathfrak{U}_j, j = 1,\ldots,L$ are independent [9]. In other words, different cells employs different sets (independent) of orthogonal training sequences. It is in contrast to the majority of works in massive MIMO where the same orthogonal training sequences are used in each cell.

Let us focus on obtaining the estimate for $h_{kji}^{[n]}$. By correlating the observation vector $\mathbf{y}_i^{[n]}$ with the training vector $\mathbf{q}_{ki}$, we have

$$\mathbf{q}_{ki}^\mathsf{H}\mathbf{y}_i^{[n]} = \sqrt{\rho_{kii}}h_{kii}^{[n]}+\sum_{j\neq i}^{L}\sum_{m=1}^{K_j}\sqrt{\rho_{mji}}h_{mji}^{[n]}\mathbf{q}_{ki}^\mathsf{H}\mathbf{q}_{mj}+\bar{n}_{ki}^{[n]} \quad (3)$$

where $\bar{n}_{ki}^{[n]} = \mathbf{q}_{ki}^\mathsf{H}\mathbf{n}_i^{[n]} \sim \mathcal{CN}(0,\sigma^2)$. The minimum mean-square estimation (MMSE) is employed based on the observation (3). Note that since BS i knows ρ_{kji}, and $\mathbf{q}_{kj}, \forall k,j$, the vector $\mathbf{y}_i^{[n]}$ is Gaussian. Moreover, the scalars $\mathbf{q}_{ki}^\mathsf{H}\mathbf{y}_i^{[n]}$ and $h_{kii}^{[n]}$ are jointly Gaussian. The MMSE estimate for $h_{kii}^{[n]}$ is given by

$$\hat{h}_{kii}^{[n]} = \frac{\upsilon_{ki}}{\sqrt{\rho_{kii}}}\mathbf{q}_{ki}^\mathsf{H}\mathbf{y}_i^{[n]} \qquad (4)$$

where

$$\upsilon_{ki} = \frac{\rho_{kii}}{\rho_{kii} + \displaystyle\sum_{j\neq i}^{L}\sum_{m=1}^{K_j}\rho_{mji}|\mathbf{q}_{ki}^\mathsf{H}\mathbf{q}_{mj}|^2 + \sigma^2}. \qquad (5)$$

Note that $\frac{\upsilon_{ki}}{\sqrt{\rho_{kii}}}$ is the scalar estimator for $h_{kii}^{[n]}$ based on the observation (3) and υ_{ki} is the variance of the channel estimate. It can be checked that the channel estimates for different users (in the same cell) are correlated. In contrast, they are uncorrelated in the FPR scheme. From (4), we can model $\hat{h}_{kii}^{[n]} \sim \mathcal{CN}(0,\upsilon_{ki})$. Note that we have removed the index n in the notation for the channel estimation variance because it does not depend on n. The channel estimate vector, $\widehat{\mathbf{h}}_{kii} = \left[\hat{h}_{kii}^{[1]}, \hat{h}_{kii}^{[2]},\ldots, \hat{h}_{kii}^{[N]}\right]^\mathsf{T}$ can be expressed as follows

$$\widehat{\mathbf{h}}_{kii} = \upsilon_{ki}\left(\mathbf{h}_{kii} + \sum_{j\neq i}^{L}\sum_{m=1}^{K_j}\sqrt{\frac{\rho_{mji}}{\rho_{kii}}}\mathbf{h}_{mji}\mathbf{q}_{mj}^\mathsf{H}\mathbf{q}_{ki} + \frac{\hat{\mathbf{n}}_{ki}}{\sqrt{\rho_{kii}}}\right) \tag{6}$$

where $\hat{\mathbf{n}}_{ki} = [\mathbf{n}_i^{[1]},\cdots,\mathbf{n}_i^{[N]}]^\mathsf{T}\mathbf{q}_{ki}^* \sim \mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I}_N)$. Note that $\widehat{\mathbf{h}}_{kii} \sim \mathcal{CN}(\mathbf{0},\upsilon_{ki}\mathbf{I}_N)$. By the property of MMSE, we can model $\mathbf{h}_{kii} = \widehat{\mathbf{h}}_{kii} + \widetilde{\mathbf{h}}_{kii}$, where $\widetilde{\mathbf{h}}_{kii} \sim \mathcal{CN}(\mathbf{0},(1-\upsilon_{ki})\mathbf{I}_N)$ is the channel estimation error vector. Moreover, $\widehat{\mathbf{h}}_{kii}$ and $\widetilde{\mathbf{h}}_{kii}$ are independent. Note that the second term of (6) is the pilot contamination term which includes the channels of all users from the interfering cells. In the FPR scheme, only users that have the same training sequence contribute to this term.

B. Uplink Data Transmission

The uplink data transmission will take the duration of $T-\tau$ symbols. Let s_{kj} be the transmitted symbol from user k in cell j with the average power p_{kj}. The received signal at BS i is

$$\mathbf{u}_i = \sum_{j=1}^{L}\sum_{k=1}^{K_j}\sqrt{p_{kj}\ell_{kji}}\mathbf{h}_{kji}s_{kj}+\mathbf{n}_i \tag{7}$$

where $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I}_N)$ is the receiver noise at BS i. It is assumed that $s_{kj} \sim \mathcal{CN}(0,1)$ and is independent of other users' data symbols. For the rest of the paper, we assume $p_{kj}=\varrho_{kj}$ and thus, $\rho_{kji}=p_{kj}\ell_{kji}$.

Let us consider user k at cell i. A linear receiver, denoted by $\mathbf{c}_{ki}$, is used by BS i to decode the received data symbols. The estimate for symbol s_{ki} is

$$\hat{s}_{ki}=\mathbf{c}_{ki}^{\mathsf{H}}\mathbf{u}_i=\mathbf{c}_{ki}^{\mathsf{H}}\sum_{j=1}^{L}\sum_{k=1}^{K_j}\sqrt{\rho_{kji}}\mathbf{h}_{kji}s_{kj}+\mathbf{c}_{ki}^{\mathsf{H}}\mathbf{n}_i.$$

In this paper, we consider the MF receiver due to its low complexity. A more sophisticated LMMSE receiver is left for future works. The receiver is constructed by using the channel estimate, i.e., $\mathbf{c}_{ki}=\widehat{\mathbf{h}}_{kii}$. The resulting SINR, denoted by γ_{ki}, can be written as follows

$$\gamma_{ki}=\frac{\rho_{kii}\|\widehat{\mathbf{h}}_{kii}\|^4}{\xi_{ki}+\displaystyle\sum_{(j,m)\neq(i,k)}^{L\ \ K_j}\rho_{mji}|\widehat{\mathbf{h}}_{kii}^{\mathsf{H}}\mathbf{h}_{mji}|^2+\sigma^2\|\widehat{\mathbf{h}}_{kii}\|^2} \tag{8}$$

where $\xi_{ki}=\rho_{kii}|\widehat{\mathbf{h}}_{kii}^{\mathsf{H}}\widetilde{\mathbf{h}}_{kii}|^2$ is the self interference noise.

III. ASYMPTOTIC ANALYSIS

In this section, we derive the asymptotic uplink SINR (8) by analyzing it in the large system regime, i.e., when K_j, N and τ tend to infinity with K_j/N and K_j/τ being finite constants. The derivation relies on results in random matrix theory [13], [14], as well as the results on Haar-distributed random matrices (see e.g., [8], [9], [14]).

Theorem 1: Let $\mathbf{\Gamma}_{ji}=\mathrm{diag}\left(\rho_{1ji},\rho_{2ji},\cdots,\rho_{K_jji}\right)$. Suppose that the empirical spectral distribution (e.s.d.) of $\mathbf{\Gamma}_{ji}$ converges to a non-random distribution $F_{\mathbf{\Gamma}_{ji}}$. As $K_j, N, \tau \to \infty$ with $\frac{K_j}{N}\to\alpha_j$ and $\frac{K_j}{\tau}\to\kappa_j$, the uplink SINR γ_{ki} converges almost surely to $\bar{\gamma}_{ki}$ which is given by

$$\bar{\gamma}_{ki}=\frac{\rho_{kii}^2\bar{v}_{ki}}{\rho_{kii}\displaystyle\sum_{j=1}^{L}\alpha_j\mathbb{E}\left[\Gamma_{ji}\right]+\bar{v}_{ki}\displaystyle\sum_{j\neq i}^{L}\kappa_j\mathbb{E}\left[\Gamma_{ji}^2\right]} \tag{9}$$

where

$$\bar{v}_{ki}=\frac{\rho_{kii}}{\rho_{kii}+\displaystyle\sum_{j\neq i}^{L}\kappa_j\mathbb{E}\left[\Gamma_{ji}\right]+\sigma^2} \tag{10}$$

and Γ_{ji} is a random variable with distribution $F_{\mathbf{\Gamma}_{ji}}$.

Proof: See Appendix A. ∎

Remark 1: In practice, i.e. when the distribution of $F_{\mathbf{\Gamma}_{ji}}$ is difficult to obtain or K_j is relatively small, we can replace $\mathbb{E}\left[\Gamma_{ji}\right]$ with its empirical value $\frac{1}{K_j}\sum_k\rho_{kji}$.

As mentioned previously, the large system analysis is performed to obtain the approximation for the uplink SINR. In practice, the parameters K_j, N and τ are finite and the approximation (9) uses the ratios of those parameters (α_j and κ_j). As shown in Section V, (9) can approximate the finite-size systems accurately. Moreover, we can infer from (9) that the uncertainty due to the fast fading has been removed and only that from the large-scale fading (ρ_{kii}) remains.

Now, let us compare (9) with the limiting SINR in the FPR scheme, denoted by $\bar{\gamma}_{ki}^s$, given by (see also [4, eq. (23)])

$$\bar{\gamma}_{ki}^s=\frac{\rho_{kii}^2\bar{v}_{ki}^s}{\rho_{kii}\displaystyle\sum_{j=1}^{L}\alpha_j\mathbb{E}\left[\Gamma_{ji}\right]+\bar{v}_{ki}^s\displaystyle\sum_{j\neq i}^{L}\rho_{kji}^2} \tag{11}$$

where

$$\bar{v}_{ki}^s=\frac{\rho_{kii}}{\rho_{kii}+\displaystyle\sum_{j\neq i}^{L}\rho_{kji}+\sigma^2}=\frac{\rho_{kii}}{\displaystyle\sum_{j=1}^{L}\rho_{kji}+\sigma^2}. \tag{12}$$

Observing the denominator of (9), the first term is the total interference (the intra- and inter-cell interference). This term also appears in the denominator of (11). Thus the limiting interference power is the same for both FPR and DOP schemes. The second term in the denominators of (9) and (11) is the pilot contamination contributed by the users from the interfering cells. However, this term takes different forms in both schemes. In DOP, the pilot contamination is caused by all users in each interfering cell in the form of the second moment of the interference powers. In contrast, the pilot contamination in the FPR scheme is caused only by the users with the same training pilot sequence.

Remark 2: In common massive MIMO setups, i.e., when $N\to\infty$ and K_j is finite ($K_j \ll N$), or equivalently $\alpha_j \to 0$, (9) will reduce to

$$\lim_{\alpha\to 0}\bar{\gamma}_{ki}=\rho_{kii}^2\left(\frac{1}{\tau}\sum_{j\neq i}^{L}\sum_{m=1}^{K_j}\rho_{mji}^2\right)^{-1} \tag{13}$$

where we use $\frac{1}{K_j}\sum_k\rho_{kji}^2$ to represent $\mathbb{E}\left[\Gamma_{ji}^2\right]$ in finite system dimensions. This agrees with the result in [3].

Now, let us consider a simple power control scheme where $p_{kj}=\frac{P_u}{\ell_{kjj}}$. Thus, the received power $\rho_{kjj}=P_u$ is the same for all users in cell j (see also [12]). Consequently, $\mathbf{\Gamma}_{ji}$ becomes

$$\mathbf{\Gamma}_{ji}=\begin{cases}P_u\mathbf{I}_{K_j}, & j=i\\P_u\bar{\mathbf{L}}_{ji}, & j\neq i\end{cases},$$

where $\bar{\mathbf{L}}_{ji}=\mathrm{diag}\left(\frac{\ell_{1ji}}{\ell_{1jj}},\frac{\ell_{2ji}}{\ell_{2jj}},\cdots,\frac{\ell_{K_jji}}{\ell_{K_jjj}}\right)$. Suppose that the e.s.d. of $\bar{\mathbf{L}}_{ji}$ converges almost surely to a deterministic distribution $F_{\bar{\mathbf{L}}_{ji}}$. Then, it is straightforward from Theorem 1 to

show that the limiting SINR under the power control, denoted by $\bar{\gamma}_{ki}^p$, is

$$\bar{\gamma}_{ki}^p = \bar{\gamma}_i^p = \frac{\bar{v}_i^p}{\sum_{j=1}^{L} \alpha_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] + \bar{v}_i^p \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]} \qquad (14)$$

where

$$\bar{v}_i^p = \frac{1}{1 + \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] + \sigma^2/P_u} \qquad (15)$$

and $\bar{\mathsf{L}}_{ji}$ is a random variable whose distribution $F_{\bar{\mathsf{L}}_{ji}}$.

We can see from (14) and (15) that the limiting SINR and channel estimation variance under power control are the same for all users in cell i. Moreover, $\bar{\gamma}_i^p$ is deterministic since it does not depend on the random locations (or particular realizations of the locations) of the users. Under the same power control, the limiting SINR for the FPR scheme is

$$\bar{\gamma}_{ki}^{s,p} = \frac{\bar{v}_{ki}^{s,p}}{\sum_{j=1}^{L} \alpha_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] + \bar{v}_{ki}^{s,p} \sum_{j \neq i}^{L} \frac{\ell_{kji}^2}{\ell_{kjj}^2}} \qquad (16)$$

where $\bar{v}_{ki}^{s,p} = \left(1 + \sum_{j \neq 1}^{L} \ell_{kji}/\ell_{kjj} + \sigma^2/P_u\right)^{-1}$. It is obvious that $\bar{\gamma}_{ki}^{s,p}$ still depends on the particular realizations of $\bar{\ell}_{kji} = \ell_{kji}/\ell_{kjj}$. Note that, for a particular realization of $\bar{\ell}_{kji}$, $\sum_{j \neq 1}^{L} \bar{\ell}_{kji}$ and $\sum_{j \neq i}^{L} \bar{\ell}_{kji}^2$ can be larger or smaller than $\sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]$ and $\sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]$, respectively. Consequently, $\bar{\gamma}_{ki}^{s,p}$ can be smaller or larger than $\bar{\gamma}_i^p$.

IV. USER CAPACITY UNDER POWER CONTROL

Here we will characterize the user capacity based on the limiting SINR under power control (14). Recall that the limiting SINR, $\bar{\gamma}_i^p$, is deterministic and is the same for all users in cell i. Now, let us assume that all users in cell i have the quality of service (QoS) requirement,

$$\bar{\gamma}_i^p \geq \gamma_i^{\text{th}} \qquad (17)$$

where γ_i^{th} is the minimum required SINR for users in cell i. Then, the user capacity is defined as the number of users per degree of freedom (N), or α_i, that satisfies the QoS (17), see also [10], [11]. From (17), we have

$$\alpha_i + \sum_{j \neq i}^{L} \alpha_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] \leq \bar{v}_i^p \left(\frac{1}{\gamma_i^{\text{th}}} - \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]\right). \qquad (18)$$

To get more insights on the user capacity, let us consider the simplest case where $K_j = K$, so that $\alpha_j = \alpha$, $\kappa_j = \kappa$. Note that K affects both α and κ and we can write $\alpha = \frac{K}{\tau} \frac{\tau}{N} = \kappa\theta$ with $\theta = \frac{\tau}{N}$. For a given degree of freedom, there are two ways of defining α that satisfies the QoS in (17), either by fixing τ or κ.

A. Fixed τ

Recall that $\alpha = \kappa\theta$. By substituting the expression for $\bar{v}_i^p$ in (15) into (18), we obtain

$$A_i \kappa^2 + B_i \kappa - \frac{1}{\gamma_i^{\text{th}}} \leq 0 \qquad (19)$$

where

$$A_i = \theta \left(\sum_{j=1}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]\right) \left(\sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]\right)$$

$$B_i = \theta \left(1 + \sigma^2/P_u\right) \sum_{j=1}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] + \sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right].$$

Now we need to find the range of κ (or α) such that (19) holds with $A_i, B_i, \gamma_i^{\text{th}} > 0$ and $0 \leq \kappa \leq 1$. The left hand side of (19) is a quadratic equation with one positive root and one negative root. The positive root (solution) is

$$\kappa_\tau = \frac{-B_i + \sqrt{B_i^2 + \frac{4A_i}{\gamma_i^{\text{th}}}}}{2A_i}. \qquad (20)$$

It is easy to check that the left hand side of (19) is a convex function and has a negative value at $\kappa = 0$. Thus, (19) will hold when

$$0 \leq \kappa \leq \kappa_\tau.$$

Equivalently, the cell loading that satisfies (17) is

$$0 \leq \alpha \leq \alpha_\tau, \quad \text{with } \alpha_\tau = \kappa_\tau \theta. \qquad (21)$$

Now, we will check the condition where $\kappa_\tau \leq 1$ holds. It is equivalent to $\sqrt{B_i^2 + \frac{4A_i}{\gamma_i^{\text{th}}}} \leq 2A_i + B_i$. By squaring both sides and performing some algebraic manipulations, $\kappa_\tau \leq 1$ holds when

$$\frac{1}{\gamma_i^{\text{th}}} \leq A_i + B_i.$$

Otherwise, we set $\kappa_\tau = 1$ (or $\alpha_\tau = \theta$).

Considering (20) and (21), it can be easily checked that α_τ is an increasing function of τ. Thus, a larger τ will increase the user capacity.

B. Fixed κ

In this case, κ is fixed and thus τ varies as K varies. The corresponding user capacity is simply given by

$$\alpha \leq \alpha_\kappa = \bar{v}_i^p \frac{\frac{1}{\gamma_i^{\text{th}}} - \kappa \sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]}{1 + \sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]}. \qquad (22)$$

Since $\alpha_\kappa \geq 0$, the right-hand side of the above equality must satisfy

$$\frac{1}{\gamma_i^{\text{th}}} \geq \kappa \sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right].$$

It is obvious that the choice of κ affects the user capacity. It can be checked that $\bar{v}_i^p \left(\frac{1}{\gamma_i^{\text{th}}} - \kappa \sum_{j \neq i}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]\right)$ is increasing as κ decreases. Note that by decreasing κ, we have a better

quality of channel estimates that obviously increases the user capacity.

Remark 3: As discussed in both subsections above, increasing τ (or decreasing κ) will result in a higher user capacity. On the other side, this will reduce the uplink transmission time and consequently the cell sum rate. This trade-off is also a concern raised in [3]. One way to handle this trade-off is to find an optimal τ, for $K \leq \tau \leq T$, that maximizes the cell sum rate. It is obvious that $\tau = T$ is never an optimal solution since it gives a zero sum-rate. Thus, the optimal τ is either at $\tau = K$ or at $\tau \in (K, T)$. Let us consider the worst case in terms of the user capacity, i.e., $\tau = K$ (or $\kappa = 1$). The user capacity in this case is easily obtained from (22) (or (21)) and is given by

$$\alpha_\kappa = \frac{\frac{1}{\gamma_i^{\text{th}}} - \sum_{j \neq i}^L \mathbb{E}\left[\bar{\mathsf{L}}_{ji}^2\right]}{\left(1 + \sum_{j \neq i}^L \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]\right)\left(1 + \sum_{j \neq i}^L \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right] + \frac{\sigma^2}{P_u}\right)}.$$

The numerical simulation for this case is shown in Figure 4 in the following section. We can see that the user capacity of the DOP scheme is considerably larger compared to that of the FPR scheme.

V. NUMERICAL SIMULATIONS

In this section, we will validate the approximation of our large system results for finite K_j, N, and τ. We consider a 7–cell hexagonal layout where the inner cell radius is normalized to 0.9 and the distance between the base stations is normalized to 2. The users are uniformly distributed across the cell. We adopt the bounded path-loss model where the slow path-gain is modeled as $\ell_{kji} = (1 + d_{kji}^\zeta)^{-1}$ where ζ is the path-loss exponent and d_{kji} is the distance between user k in cell j and BS i. We set $\zeta = 3.7$ and ignore the shadowing effect in the model. The experiment takes parameters $P_u = 0$ dB, $K_j = K = 20$, $\kappa = \frac{2}{3}$ and the number of antennas varies from 50 to 500 with interval 50. We generated 500 channel realizations according to the Rayleigh distribution. Figure 1 shows the SINR (in decibels) as the number of antennas at the BS increases. We can see that the SINR obtained from the large system analysis or LSA (9) can approximate the simulation results (finite-size systems, i.e., with $K = 20$) accurately. Observe that the SINR obtained by using the conventional approach in massive MIMO ($K < \infty$ but $N \to \infty$), acts as the upper-bound for the finite-size systems. The figure also implies that this approach gives quite a loose approximation compared to the large system approach (see also [7] for the convergence behavior of massive MIMO systems). Observe that for $N = 100$, the SINR gap between the massive MIMO approximation and the finite size result is about 9 dB. This gap reduces as N increases and it is approximately 2.3 dB when $N = 500$.

Figure 2 compares the cdf of $\bar{\gamma}_{ki}^{s,p}$ and $\bar{\gamma}_i^p$. Note that $\bar{\gamma}_i^p$ is a deterministic quantity. Hence, its cdf is a unit impulse located at $\bar{\gamma}_i^p$. To produce the curves, we set $\alpha = 0.1$ and $\kappa = \{\frac{1}{3}, \frac{2}{3}, 1\}$. Note that, $\bar{\gamma}_i^p$ depends on κ while $\bar{\gamma}_{ki}^{s,p}$ does not. It is obvious from the curves for $\bar{\gamma}_i^p$ that smaller κ
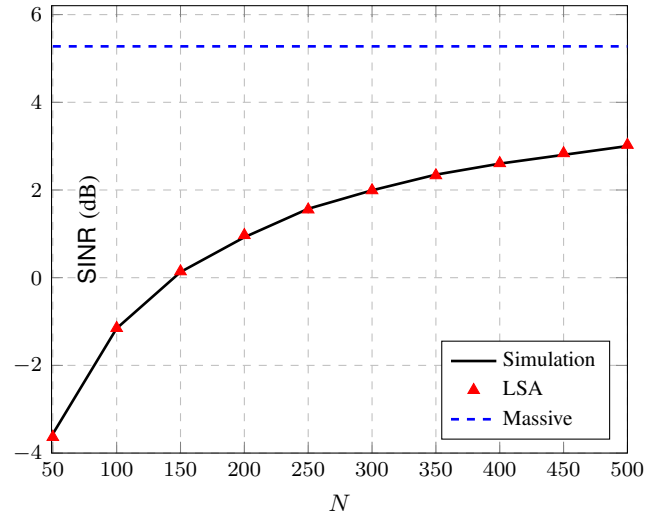

Fig. 1. The asymptotic SINR obtained from the simulation, large system analysis (LSA) and the conventional approach ($N \to \infty$ but K is finite) in massive MIMO systems.
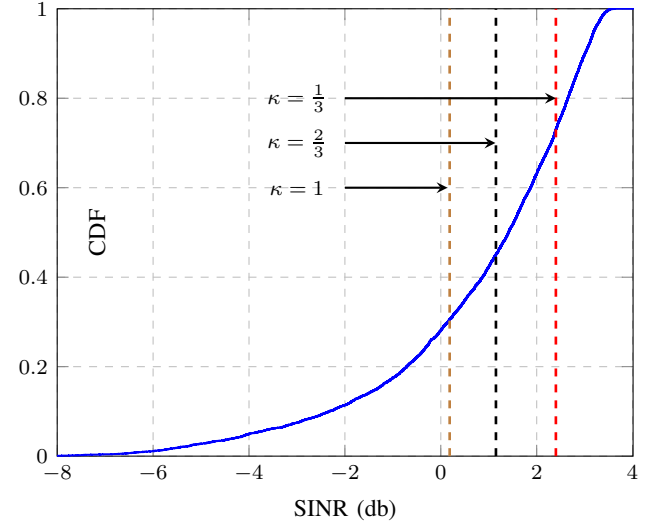

Fig. 2. CDF of $\bar{\gamma}_{ki}^{s,p}$ (solid) and $\bar{\gamma}_i^p$ (dashed) for $\alpha = 0.1$ and $\kappa = \{\frac{1}{3}, \frac{2}{3}, 1\}$

gives a better performance (SINR). The figure also reveals that there is some probability that $\bar{\gamma}_i^p$ is larger than $\bar{\gamma}_{ki}^{s,p}$ and vice-versa. For example, when $\kappa = 1$, there is a 30% chance that $\bar{\gamma}_i^p$ is larger than $\bar{\gamma}_{ki}^{s,p}$. When smaller $\kappa = \frac{1}{3}$ is used, the probability becomes higher i.e., around 75%. This probability also changes when α varies. Thus, it is interesting to see how the user capacity (which is related to α) is affected by both schemes.

The user capacity for the different orthogonal training sequences (DOP scheme) has been presented in Section IV. For the FPR scheme, the formulation of user capacity is slightly different since $\bar{\gamma}_{ki}^{s,p}$ is a random quantity contributed by the large-scale fading terms of the pilot contamination. Hence, the user capacity region, for the simplest case ($K_j = K, \forall j$)
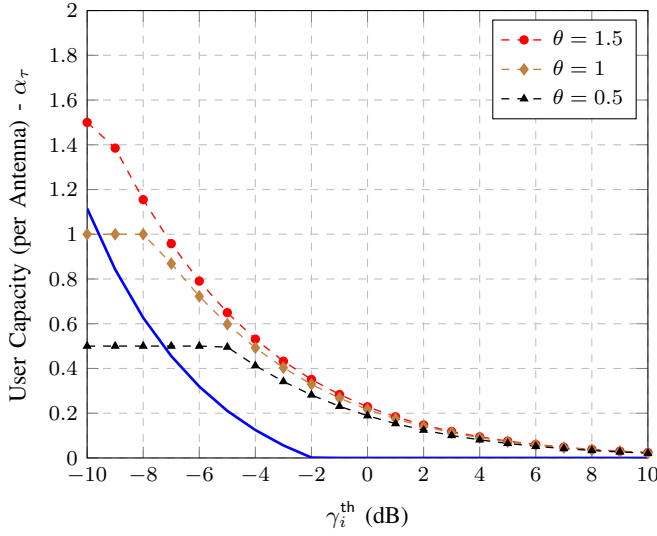
Fig. 3. User Capacity of the FPR (solid) and the DOP (mark/dashed) schemes for different values of $\theta = \frac{\tau}{N}$.
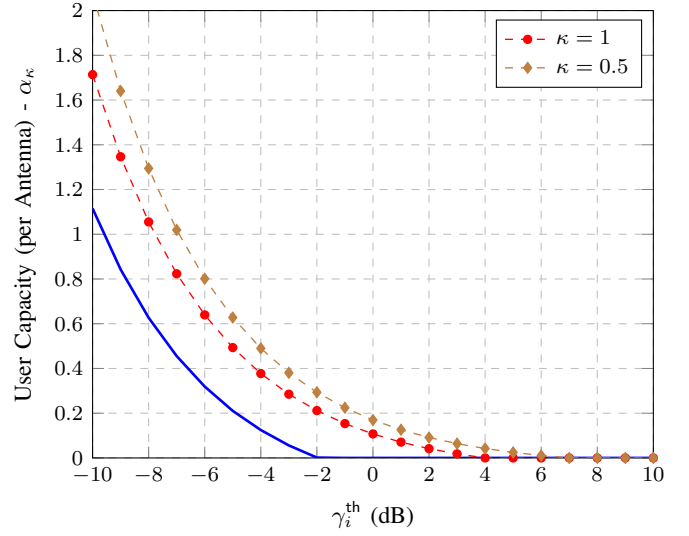

Fig. 4. User Capacity of the FPR (solid) and the DOP (mark/dashed) schemes for different values of κ.

is defined as the value of α that satisfies

$$\mathbb{P}\left(\bar{\gamma}_{ki}^{s,p} \geq \gamma_i^{\mathsf{th}}\right) \geq 1 - \beta \tag{23}$$

where β is the outage probability. The above is equivalent to $\mathbb{P}\left(X_{ki} \leq 0\right) \geq 1 - \beta$ or $F_{X_{ki}}(0) \geq 1 - \beta$ where

$$X_{ki} = \alpha \sum_{j=1}^{L} \mathbb{E}\left[\bar{\mathsf{L}}_{ji}\right]\left(1 + \frac{\sigma^2}{P_u} + \sum_{j \neq i}^{L} \bar{\ell}_{kji}\right) + \sum_{j \neq i}^{L} \bar{\ell}_{kji}^2 - \frac{1}{\gamma_i^{\mathsf{th}}}.$$

Figure 3 presents the comparisons of the user capacities of the FPR obtained from (23) with $\beta = 0.05$ and of the DOP represented by α_τ in (21) with $\theta = \{0.5, 1, 1.5\}$. As expected, higher outages lead to higher user capacities for the FPR while higher θ (higher τ with N fixed) gives higher user capacities for DOP. In general, we can see that the FPR has a higher user capacity compared to the DOP for very low γ_i^{th} (less than -6 dB) and the reverse occurs for low to high γ_i^{th}s. The user capacity of the FPR scheme is zero starting at $\gamma_i^{\mathsf{th}} = -2$ dB for outage probability 5% and at $\gamma_i^{\mathsf{th}} = 0$ dB. On the other hand, the user capacities for the DOP scheme are about $0.28 - 0.35$ and 0.2 when $\gamma_i^{\mathsf{th}} = -2$ dB and $\gamma_i^{\mathsf{th}} = 0$ dB, respectively. This shows the advantage of the DOP scheme over the FPR scheme in terms of user capacity and is consistent with the findings in [6]. A similar conclusion can be drawn by comparing the user capacities of FPR and α_κ of DOP with $\kappa = \{0.5, 1\}$ as shown in Figure 4. Even for the worst case, $\kappa = 1$, the user capacities of the DOP are higher than those of the FPR at all considered values of γ_i^{th}. For example, at $\gamma_i^{\mathsf{th}} = -10$ dB, the user capacities in the FPR scheme and the DOP scheme are about 1.1 and 1.7, respectively. Hence, for example, if the BS has 500 antennas, the DOP scheme can admit 300 more users than the FPR scheme can.

VI. Conclusion

In this paper, we consider the performance analysis of up-link massive MIMO systems with different sets of orthogonal training pilots in different cells employed during the uplink training. We obtained a novel expression for the asymptotic SINR in the large system regime that approximates finite-size systems accurately. We show that under a simple power control, the asymptotic SINR is a deterministic quantity and its denominator contains the average interference and mean-squared interference (pilot contamination) of all users from the interfering cells. We also derived the user capacity of the system and compared to that of the FPR scheme. Our simulation results showed that the DOP generally gives a better user capacity even for low values of the required SINRs.

Appendix A
Proof of Theorem 1

We begin the proof by establishing the limiting result for the channel estimate variance υ_{ki}. Then, we proceed to obtain the limiting results for the numerator and the denominator of the uplink SINR γ_{ki} which are denoted by A_{ki} and B_{ki} respectively. In the proof we use the following definitions and results from large random matrices.

Definition 1 ([15, Definition 2.6]): Two infinite sequences a_n and b_n are defined to be asymptotically equivalent as $n \to \infty$, denoted by $a_n \asymp b_n$, if only if $|a_n - b_n| \xrightarrow{a.s.} 0$.

Definition 2 ([15, Definition 2.7]): Two infinite sequences $a_{k,n}$ and $b_{k,n}$, indexed by $k = 1, 2, \cdots, n$, are defined to be uniformly asymptotically equivalent as $n \to \infty$, denoted by $a_{k,n} \overset{k}{\asymp} b_{k,n}$, if only if $\max_{k \leq N}|a_{k,n} - b_{k,n}| \xrightarrow{a.s.} 0$.

Lemma 1 ([15, Lemma 2.5]): If $a_{k,n} \overset{k}{\asymp} b_{k,n}$, then $\frac{1}{n}\sum_{k=1}^{n} a_{k,n} \asymp \frac{1}{n}\sum_{k=1}^{n} b_{k,n}$.

Lemma 2 ([16, Lemma 1]): Let $\mathbf{A} \in \mathbb{C}^{N \times N}$ be a deterministic matrix with uniformly bounded spectral radius for

all N. Let $\mathbf{x} = \frac{1}{\sqrt{N}}[x_1, x_2, \cdots, x_N]^\mathsf{T}$ where the x_i's are i.i.d with zero mean, unit variance and finite eighth moment. Let $\mathbf{y}$ be a similar vector independent of $\mathbf{x}$. Then, we have $\mathbf{x}\mathbf{A}\mathbf{x}^\mathsf{H} - \frac{1}{N}\mathrm{Tr}(\mathbf{A}) \xrightarrow{a.s.} 0$, and $\mathbf{x}\mathbf{A}\mathbf{y}^\mathsf{H} \xrightarrow{a.s.} 0$.

The following lemma which involves the columns of an isometric matrix is analogues to Lemma 2.

Lemma 3: Let $\mathbf{w} \in \mathbb{C}^N$ be a column of an isometric matrix[1] $\mathbf{W}$ and $\mathbf{z}$ be a column of an isometric $\mathbf{Z}$ matrix independent of $\mathbf{W}$. Let $\mathbf{A}$ be as defined in Lemma 2 and independent of $\mathbf{W}$ and $\mathbf{Z}$. Then, $\mathbf{w}^\mathsf{H}\mathbf{A}\mathbf{w} - \frac{1}{N}\mathrm{Tr}(\mathbf{A}) \xrightarrow{a.s.} 0$ and $\mathbf{w}^\mathsf{H}\mathbf{A}\mathbf{z} \xrightarrow{a.s.} 0$.

In the following subsections, we use the notation $\mathbf{X}_{[k]}$ for matrix $\mathbf{X}$ with k-th column removed. If $\mathbf{X}$ is a diagonal matrix, $\mathbf{X}_{[k]}$ denotes $\mathbf{X}$ with both k-th column and k-th row deleted.

A. The large system limit for υ_{ki}

Here, we only need to find the asymptotic limit for the second term in the denominator of υ_{ki} in (5). That term can be written as $\sum_{j\neq i}^{L}\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_j\boldsymbol{\Gamma}_{ji}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki}$. Since $\mathbf{q}_{ki}$ is independent of $\mathbf{Q}_j$ for $j \neq i$, $\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_j\boldsymbol{\Gamma}_{ji}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki} \asymp \frac{1}{\tau}\mathrm{Tr}(\mathbf{Q}_j\boldsymbol{\Gamma}_{ji}\mathbf{Q}_j^\mathsf{H})$ by Lemma 3. By employing the trace property, the right hand side is equal to $\frac{1}{\tau}\mathrm{Tr}(\boldsymbol{\Gamma}_{ji})$. Suppose that the e.s.d. of $\boldsymbol{\Gamma}_{ji}$ converges almost surely to the non-random limiting spectral distribution (l.s.d.) $F_{\boldsymbol{\Gamma}_{ji}}$ as $K_j, \tau \to \infty$ with $\frac{K_j}{\tau} \to \kappa_j$, $\frac{1}{\tau}\mathrm{Tr}(\boldsymbol{\Gamma}_{ji}) = \frac{K_j}{\tau}\frac{1}{K_j}\mathrm{Tr}(\boldsymbol{\Gamma}_{ji})$ converges almost surely to $\kappa_j\mathbb{E}[\Gamma_{ji}]$ where Γ_{ji} is a random variable with distribution $F_{\boldsymbol{\Gamma}_{ji}}$. Consequently, $\upsilon_{ki} \xrightarrow{a.s.} \bar{\upsilon}_{ki}$, where $\bar{\upsilon}_{ki}$ is defined in (10).

B. The large system limit for A_{ki}

From the channel model described in subsection II-A, $\widehat{\mathbf{h}}_{kii} \sim \mathcal{CN}(\mathbf{0}, \upsilon_{ki}\mathbf{I}_N)$. By using Lemma 2, $\frac{1}{N}\|\widehat{\mathbf{h}}_{kii}\|^2 \asymp \upsilon_{ki}$. Since $\upsilon_{ki} \xrightarrow{a.s.} \bar{\upsilon}_{ki}$, we have $\frac{1}{N^2}A_{ki} \xrightarrow{a.s.} \rho_{kii}(\bar{\upsilon}_{ki})^2$.

C. The large system limit for B_{ki}

First let us consider the self interference noise (ξ_{ki}). Note that while $\frac{1}{N}\widehat{\mathbf{h}}_{kii}^\mathsf{H}\widetilde{\mathbf{h}}_{kii} \asymp 0$ by Lemma 2. Consequently, $\frac{1}{N^2}\xi_{ki} \xrightarrow{a.s.} 0$. For the noise term, while $\frac{1}{N}\|\widehat{\mathbf{h}}_{kii}\|^2 \xrightarrow{a.s.} \bar{\upsilon}_{ki}$, we have $\frac{1}{N^2}\|\widehat{\mathbf{h}}_{kii}\|^2 \xrightarrow{a.s.} 0$. Thus, the noise term converges to zero almost surely.

For the intra-cell interference term, it can be written as $\frac{1}{N^2}\widehat{\mathbf{h}}_{kii}^\mathsf{H}\mathbf{H}_{ii[k]}\boldsymbol{\Gamma}_{ii[k]}\mathbf{H}_{ii[k]}^\mathsf{H}\widehat{\mathbf{h}}_{kii}$. Since $\widehat{\mathbf{h}}_{kii}$ is independent of $\mathbf{H}_{ii[k]}$ and $\boldsymbol{\Gamma}_{ii[k]}$ and by applying Lemma 2 and the rank-one perturbation lemma [14], the intra-cell interference is asymptotically equivalent to $\frac{\bar{\upsilon}_{ki}}{N}\mathrm{Tr}(\frac{1}{N}\mathbf{H}_{ii}\boldsymbol{\Gamma}_{ii}\mathbf{H}_{ii}^\mathsf{H})$. Note that $\frac{1}{N}\mathrm{Tr}(\frac{1}{N}\mathbf{H}_{ii}\boldsymbol{\Gamma}_{ii}\mathbf{H}_{ii}^\mathsf{H})$ is the first moment of the e.s.d. of $\frac{1}{N}\mathbf{H}_{ii}\boldsymbol{\Gamma}_{ii}\mathbf{H}_{ii}^\mathsf{H}$. It can be shown that (see also [13, eq. 2.118]), this moment converges almost surely to $\alpha_i\mathbb{E}[\Gamma_{ii}]$. Hence, the intra-cell interference power converges to $\alpha_i\bar{\upsilon}_{ki}\mathbb{E}[\Gamma_{ii}]$ almost surely.

Now, let us consider the inter-cell interference term. The (normalized) interference from cell j can be written as $\mathcal{I}_{kij} = \frac{1}{N^2}\widehat{\mathbf{h}}_{kii}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\widehat{\mathbf{h}}_{kii}$. Let us rewrite (6)

$$\widehat{\mathbf{h}}_{kii} = \frac{\upsilon_{ki}}{\sqrt{\rho_{kii}}}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}^{\frac{1}{2}}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki} + \mathbf{z}_{ki}$$

---

where $\mathbf{z}_{ki} = \upsilon_{ki}\mathbf{h}_{kii} + \frac{\upsilon_{ki}}{\sqrt{\rho_{kii}}}\left(\sum_{l\neq\{i,j\}}^{L}\mathbf{H}_{li}\boldsymbol{\Gamma}_{li}^{\frac{1}{2}}\mathbf{Q}_l^\mathsf{H}\mathbf{q}_{ki} + \hat{\mathbf{n}}_{ki}\right)$.

Let $\mathbf{a}_{c,j} = \frac{\upsilon_{ki}}{\sqrt{\rho_{kii}}}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}^{\frac{1}{2}}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki}$ that represents the (pilot) contamination from cell j. Thus,

$$\mathcal{I}_{kij} = \frac{1}{N^2}\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{z}_{ki} + \frac{1}{N^2}\mathbf{a}_{c,j}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{a}_{c,j}$$
$$+ \frac{2}{N^2}\Re\left\{\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{a}_{c,j}\right\}.$$

It is easy to check that $\mathbf{z}_{ki}$ and $\mathbf{a}_{c,j}$ are (mutually) independent. Furthermore, $\mathbf{z}_{ki}$ and $\mathbf{a}_{c,j}$ are zero mean Gaussian vectors with variances $\psi_{ki}\mathbf{I}_N$ with $\psi_{ki} = \upsilon_{ki}^2 + \frac{\upsilon_{ki}^2}{\rho_{kii}}\left(\sigma^2 + \sum_{l\neq\{i,j\}}^{L}\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_l\boldsymbol{\Gamma}_{li}\mathbf{Q}_l^\mathsf{H}\mathbf{q}_{ki}\right)$ and $\frac{\upsilon_{ki}^2}{\rho_{kii}}\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_j\boldsymbol{\Gamma}_{ji}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki}$, respectively. Let us denote $\mathcal{I}_{kij,1}$, $\mathcal{I}_{kij,2}$ and $\mathcal{I}_{kij,3}$ respectively for each term of $\mathcal{I}_{kij}$. Considering $\mathcal{I}_{kij,1}$, it is obvious that $\mathbf{z}_{ki}$ is independent of $\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}$. Hence, $\mathcal{I}_{kij,1} \asymp \frac{\psi_{ki}}{N}\mathrm{Tr}(\frac{1}{N}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji})$. As shown previously $\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_l\boldsymbol{\Gamma}_{li}\mathbf{Q}_l^\mathsf{H}\mathbf{q}_{ki} \xrightarrow{a.s.} \kappa_l\mathbb{E}[\Gamma_{li}]$. Thus, $\psi_{ki} \xrightarrow{a.s.} \bar{\psi}_{ki}$ where

$$\bar{\psi}_{ki} = \frac{\bar{\upsilon}_{ki}^2}{\rho_{kii}}\left(\rho_{kii} + \sigma^2 + \sum_{l\neq\{i,j\}}^{L}\kappa_l\mathbb{E}[\Gamma_{li}]\right).$$

Moreover, $\frac{1}{N}\mathrm{Tr}(\frac{1}{N}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji})$ converges to $\alpha_j\mathbb{E}[\Gamma_{ji}]$ almost surely. Thus,

$$\mathcal{I}_{kij,1} \xrightarrow{a.s.} \frac{\alpha_j\bar{\upsilon}_{ki}^2\mathbb{E}[\Gamma_{ji}]}{\rho_{kii}}\left(\rho_{kii} + \sigma^2 + \sum_{l\neq\{i,j\}}^{L}\kappa_l\mathbb{E}[\Gamma_{li}]\right).$$

For $\mathcal{I}_{kij,2}$, we have

$$\mathcal{I}_{kij,2} = \frac{\upsilon_{ki}^2}{\rho_{kii}}\left(\frac{1}{N^2}\mathbf{q}_{ki}^\mathsf{H}\mathbf{Q}_j\boldsymbol{\Gamma}_{ji}^{\frac{1}{2}}\mathbf{H}_{ji}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}^{\frac{1}{2}}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki}\right)$$
$$\asymp \frac{\upsilon_{ki}^2}{\rho_{kii}} \times \frac{N}{\tau} \times \frac{1}{N}\mathrm{Tr}\left(\left(\frac{1}{N}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\right)^2\right).$$

The last term in the last line is the second moment of the e.s.d. of $\frac{1}{N}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}$. It can be shown that (see also [13, eq. 2.118]) it converges almost surely to $\alpha_j\mathbb{E}[\Gamma_{ji}^2] + \alpha_j^2(\mathbb{E}[\Gamma_{ji}])^2$. By writing $\frac{N}{\tau} = \frac{K_j}{\tau}\frac{N}{K_j} \to \frac{\kappa_j}{\alpha_j}$, we obtain

$$\mathcal{I}_{kij,2} \xrightarrow{a.s.} \frac{\bar{\upsilon}_{ki}^2}{\rho_{kii}}\kappa_j\left(\mathbb{E}[\Gamma_{ji}^2] + \alpha_j(\mathbb{E}[\Gamma_{ji}])^2\right).$$

Let us consider the term $\frac{1}{N^2}\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{a}_{c,j}$ in $\mathcal{I}_{kij,3}$. It can be written as

$$\frac{\upsilon_{ki}}{N^2\sqrt{\rho_{kii}}}\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}^{\frac{1}{2}}\mathbf{Q}_j^\mathsf{H}\mathbf{q}_{ki}$$
$$= \frac{\upsilon_{ki}}{N^2\sqrt{\rho_{kii}}}\sum_{m=1}^{K_j}\sqrt{\rho_{mji}}(\mathbf{q}_{mj}^\mathsf{H}\mathbf{q}_{ki})\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji}\boldsymbol{\Gamma}_{ji}\mathbf{H}_{ji}^\mathsf{H}\mathbf{h}_{mji}$$
$$= \frac{\upsilon_{ki}}{N\sqrt{\rho_{kii}}}\sum_{m=1}^{K_j}\sqrt{\rho_{mji}}(\mathbf{q}_{mj}^\mathsf{H}\mathbf{q}_{ki})\frac{1}{N}\mathbf{z}_{ki}^\mathsf{H}\mathbf{H}_{ji[m]}\boldsymbol{\Gamma}_{ji[m]}\mathbf{H}_{ji[m]}^\mathsf{H}\mathbf{h}_{mji}$$
$$+ \frac{\upsilon_{ki}}{N\sqrt{\rho_{kii}}}\sum_{m=1}^{K_j}\rho_{mji}^{\frac{3}{2}}(\mathbf{q}_{mj}^\mathsf{H}\mathbf{q}_{ki})\frac{\mathbf{z}_{ki}^\mathsf{H}\mathbf{h}_{mji}\mathbf{h}_{mji}^\mathsf{H}\mathbf{h}_{mji}}{N}.$$

It can be checked that all the summands of the last equation are asymptotically equivalent to zero. Therefore, by using Lemma 1, $\mathcal{I}_{kij,3} \xrightarrow{a.s.} 0$.

Summing up over all interfering cells, the limiting result for the total inter-cell interference $\sum_{j \neq i} \mathcal{I}_{kij}$ is

$$\frac{\bar{v}_{ki}^2}{\rho_{kii}} \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\Gamma_{ji}^2\right]$$

$$+ \frac{\bar{v}_{ki}^2}{\rho_{kii}} \left(\rho_{kii} + \sum_{l \neq i}^{L} \kappa_l \mathbb{E}\left[\Gamma_{li}\right] + \sigma^2\right) \sum_{j \neq i}^{L} \alpha_j \mathbb{E}\left[\Gamma_{ji}\right].$$

Using the definition for $\bar{v}_{ki}$ in (10), the second term can be simplified to $\bar{v}_{ki} \sum_{j \neq i}^{L} \alpha_j \mathbb{E}\left[\Gamma_{ji}\right]$. Thus,

$$\sum_{j \neq i} \mathcal{I}_{kij} \xrightarrow{a.s.} \bar{v}_{ki} \left(\sum_{j \neq i}^{L} \alpha_j \mathbb{E}\left[\Gamma_{ji}\right] + \frac{\bar{v}_{ki}}{\rho_{kii}} \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\Gamma_{ji}^2\right]\right).$$

Bringing up the limiting results for each term of B_k, we obtain

$$\frac{1}{N^2} B_{ki} \xrightarrow{a.s.} \bar{v}_{ki} \left(\sum_{j=1}^{L} \alpha_j \mathbb{E}\left[\Gamma_{ji}\right] + \frac{\bar{v}_{ki}}{\rho_{kii}} \sum_{j \neq i}^{L} \kappa_j \mathbb{E}\left[\Gamma_{ji}^2\right]\right).$$

By putting the large system results for A_{ki} and B_{ki}, (9) follows immediately and this completes 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, February 2014.

[2] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal. Proc. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[3] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, 2010.

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

[5] N. Krishnan, R. Yates, and N. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, 2014.

[6] A. Sivamalai and J. Evans, "On uplink user capacity for massive MIMO cellular networks," in *Proc. IEEE Globecom*, 2016, to appear.

[7] P. J. Smith, C. Neil, M. Shafi, and P. A. Dmochowski, "On the convergence of massive MIMO systems," in *Proc. IEEE International Conference on Communications (ICC)*, June 2014, pp. 5191–5196.

[8] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Transactions on Information Theory*, vol. 49, no. 5, 2003.

[9] M. Peacock, I. Collings, and M. Honig, "Asymptotic spectral efficiency of multiuser multisignature CDMA in frequency-selective channels," *IEEE Transactions on Information Theory*, vol. 52, no. 3, 2006.

[10] J.-C. Shen, J. Zhang, and K. Letaief, "Downlink User Capacity of Massive MIMO Under Pilot Contamination," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3183–3193, 2015.

[11] N. Akbar, N. Yang, P. Sadeghi, and R. Kennedy, "Multi-cell multiuser massive MIMO networks: User capacity analysis and pilot design," *IEEE Transactions on Communications*, vol. PP, no. 99, p. 1, 2016.

[12] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, Feb. 2016.

[13] A. M. Tulino and S. Verdu, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[14] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[15] M. J. M. Peacock, "Random Matrix Theory Analysis of Fixed and Adaptive Linear Multiuser Receivers," Ph.D. dissertation, School of Electrical and Information Engineering, The University of Sydney, August 2005.

[16] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.