Power Normalization in Massive MIMO Systems: How to Scale Down the Number of Antennas

Meysam Sadeghi[§†], Luca Sanguinetti[†*], Romain Couillet[‡], Chau Yuen[§]
[§]Singapore University of Technology and Design (SUTD), Singapore
[†]Large Networks and System Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France
[*]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
[‡]Telecommunication Department of CentraleSupélec, Gif-sur-Yvette, France

***Abstract*—This work considers the downlink of a massive MIMO system in which L base stations (BSs) of N antennas each communicate with K single-antenna user equipments randomly positioned in the coverage area. Within this setting, we are interested in studying the effect of power normalization on the sum rate of the system when maximum ratio transmission (MRT) or zero forcing (ZF) are employed as precoding schemes. In particular, we consider the most common two power normalization methods known as vector and matrix normalizations. The analysis is conducted assuming that N and K grow large with a non-trivial ratio K/N under the assumption that the data transmission in each cell is affected by channel estimation errors, pilot contamination, an arbitrary large scale attenuation, and antenna correlation at the BSs. The asymptotic results are instrumental to get insights and make comparisons. For medium to high signal-to-noise ratios, simulations and theory concur to show that vector normalization largely outperforms matrix normalization for both MRT and ZF and thus allows to scale down the number of antennas required to achieve a target sum rate.**

I. Introduction

Massive MIMO (also known as large scale MIMO) is considered as one of the most promising technology for next generation cellular networks [1]–[4]. The massive MIMO technology aims at evolving the conventional base stations (BSs) by using arrays with a hundred or more small dipole antennas. This allows for coherent multi-user MIMO transmission where tens of users can be multiplexed in both the uplink (UL) and downlink (DL) of each cell. It is worth observing that, contrary to what the name "massive" suggests, massive MIMO arrays are rather compact; 160 dual-polarized antennas at 3.7 GHz fit into the form factor of a flat-screen television [5].

In this work, we consider the downlink of a massive MIMO system in which L base stations (BSs) of N antennas each communicate with K single-antenna user equipments randomly positioned in the coverage area. We assume that maximum ratio transmission (MRT) or zero forcing (ZF) are employed as linear precoding schemes. Both techniques have been extensively studied and compared in the massive MIMO literature under different operating conditions and using different power normalization methods. The two most common normalization techniques are known as vector and matrix normalizations. For example, in [6] the authors consider a simple single cell network and compare the performance of MRT and ZF in terms of spectral efficiency and energy efficiency under the assumption that matrix normalization is employed. The same analysis is carried out in [3] for a more realistic multi-cellular network in which the data transmission in each cell is affected by channel estimation errors, pilot contamination, an arbitrary large scale attenuation, and antenna correlation at the BSs. The vector normalization method is employed in [7] wherein a new multi-cell minimum-mean-square-error (MMSE) scheme is proposed for massive MIMO networks to exploit all available pilots for interference suppression. The vector normalization is also used in [8] to deploy dense networks for maximal energy efficiency. To the best of our knowledge, it is not clear yet which one of the two normalization methods should be used. A first attempt to evaluate the impact of power normalization is carried out in [9] wherein the authors consider the DL sum rate of a massive MIMO system composed by three BSs connected via one digital unit. It turns out that vector normalization is better for ZF while matrix normalization is better for MRT. However, the analysis is conducted under the assumption of perfect channel state information, full cooperation among the three BSs. Also, the large scale attenuation is neglected though this has a fundamental effect on the choice of normalization method.

The aim of this work is to provide a detailed treatment on the impact of the two above power normalization methods on the DL sum rate of a massive MIMO system affected by channel estimation errors, pilot contamination, an arbitrary large scale attenuation, and antenna correlation at the BSs. The analysis is conducted assuming that N and K grow large with a non trivial ratio K/N. The asymptotic results are necessary to provide insights and enable us to compare the effect of normalization without the need to carry out extensive Monte-Carlo simulations. For a practical network setting and different values of N and K, it turns out vector normalization largely outperforms matrix normalization for both precoding schemes. In turn, this means that the use of vector normalization allows to scale down the number of antennas N required to meet a given target sum rate.

The remainder of this paper is organized as follows. Next section introduces the system model and precoder design. The

The research of Luca Sanguinetti and Romain Couillet has been supported by the ERC Starting Grant 305123 MORE and by the HUAWEI project RMTin5G.

asymptotic analysis is performed in Section IV and validated by means of numerical results in Section V. Conclusions are drawn in Section VI.

II. System Model and Precoder Design

Consider[1] the downlink of a multi-cell multi-user MIMO system composed of L cells, the BS of each cell comprising N antennas to communicate with $K < N$ single-antenna UEs. A double index notation is used to refer to each UE as e.g., "user k in cell j". Under this convention, let $\mathbf{h}_{jlk} \in \mathbb{C}^N$ be the channel from BS j to UE k in cell l within a block and assume that

$$\mathbf{h}_{jlk} = \mathbf{\Theta}_{jlk}^{1/2} \mathbf{z}_{jlk} \tag{1}$$

where $\mathbf{z}_{jlk} \in \mathbb{C}^N$ is the small-scale fading channel assumed to be Gaussian with zero mean and unit covariance and $\mathbf{\Theta}_{jlk} \in \mathbb{C}^{N \times N}$ accounts for the corresponding channel correlation matrix (from BS l to UE k in cell j). Note that (1) is very general and includes many channel models in the literature as special cases [10].

Denoting by $\mathbf{g}_{jk} \in \mathbb{C}^N$ the precoding vector of UE k in cell j, its received signal can be written as

$$y_{jk} = \mathbf{h}_{jjk}^H \mathbf{g}_{jk} s_{jk} + \sum_{i=1, i\neq k}^{K} \mathbf{h}_{jjk}^H \mathbf{g}_{ji} s_{ji} + \sum_{l=1, l\neq j}^{L} \sum_{i=1}^{K} \mathbf{h}_{ljk}^H \mathbf{g}_{li} s_{li} + n_{jk} \tag{2}$$

with $s_{li} \in \mathbb{C}$ being the signal intended to UE i in cell l, assumed independent across (l, i) pairs, of zero mean and unit variance, and $n_{jk} \sim \mathcal{CN}(0, 1/\rho_{\mathrm{dl}})$ where ρ_{dl} accounts for the signal-to-noise ratio (SNR) in the DL. We assume that the BS and UEs are perfectly synchronized and operate according to a time-division duplex (TDD) protocol wherein the downlink data transmission phase is preceded in the uplink by a training phase for channel estimation.

If an MMSE estimator is employed [3], then the estimate $\hat{\mathbf{h}}_{jjk}$ of $\mathbf{h}_{jjk}$ is distributed as $\hat{\mathbf{h}}_{jjk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_{jjk})$ with

$$\mathbf{\Phi}_{ljk} = \mathbf{\Theta}_{llk} \left(\sum_{n=1}^{L} \mathbf{\Theta}_{lnk} + \frac{1}{\rho_{\mathrm{tr}}} \mathbf{I}_N \right)^{-1} \mathbf{\Theta}_{ljk} \tag{3}$$

where ρ_{tr} accounts for the SNR during the uplink training phase. Therefore the estimated uplink channel of cell j becomes $\hat{\mathbf{H}}_{jj} = [\hat{\mathbf{h}}_{jj1}, \ldots, \hat{\mathbf{h}}_{jjK}]$.

As in [1], [3], [11], [12] (among many others), we assume that there are no downlink pilots such that the UEs do not have knowledge of the current channels but can only learn the average channel gain $\mathbb{E}\{\mathbf{h}_{jjk}^H \mathbf{g}_{jk}\}$ and the total interference power. Using the same technique from [13], an ergodic achievable information rate for UE k in cell j is obtained as

$$r_{jk} = \log_2(1 + \gamma_{jk}) \tag{4}$$

where γ_{jk} is given by

$$\gamma_{jk} = \frac{|\mathbb{E}[\mathbf{h}_{jjk}^H \mathbf{g}_{jk}]|^2}{\frac{1}{\rho_{\mathrm{dl}}} + \sum_{l=1}^{L} \sum_{i=1}^{K} \mathbb{E}[|\mathbf{h}_{ljk}^H \mathbf{g}_{li}|^2] - |\mathbb{E}[\mathbf{h}_{jjk}^H \mathbf{g}_{jk}]|^2} \tag{5}$$

where the expectation is taken with respect to the channel realizations. The above result holds true for any precoding scheme and is obtained by treating the inter-user interference (from the same and other cells) and channel uncertainty as worst-case Gaussian noise. As mentioned earlier, we consider MRT and ZF as precoding schemes [1], [3], [4], [11]. Denoting by $\mathbf{G}_j = [\mathbf{g}_{j1}, \ldots, \mathbf{g}_{jK}] \in \mathbb{C}^{N \times K}$ the precoding matrix of BS j, we have that

$$\mathbf{G}_j = \mathbf{F}_j \mathbf{D}_j^{-1/2} \tag{6}$$

where $\mathbf{F}_j = [\mathbf{f}_{j1}, \ldots, \mathbf{f}_{jK}] \in \mathbb{C}^{N \times K}$ takes the form

$$\mathbf{F}_j = \begin{cases} \hat{\mathbf{H}}_{jj} & \text{for MRT} \\ \hat{\mathbf{H}}_{jj} \left(\hat{\mathbf{H}}_{jj}^H \hat{\mathbf{H}}_{jj} \right)^{-1} & \text{for ZF} \end{cases} \tag{7}$$

whereas $\mathbf{D}_j \in \mathbb{C}^{K \times K}$ is a diagonal matrix whose entries are chosen so as to satisfy the following average power constraint $\mathbb{E}[\mathtt{tr}\mathbf{G}_j \mathbf{G}_j^H] = K \ \forall j$. The two common approaches for satisfying power constraints are known as vector and matrix normalizations [9]. If vector normalization (VN) is used, then the kth element of $\mathbf{D}_j$ is computed as

$$[\mathbf{D}_j]_{k,k} = \mathbb{E}[\mathbf{f}_{jk}^H \mathbf{f}_{jk}]. \tag{8}$$

On the other hand, if matrix normalization (MN) is employed, then $\mathbf{D}_j = \eta_j \mathbf{I}_K$ with

$$\eta_j = \mathbb{E}[\mathtt{tr}\mathbf{F}_j \mathbf{F}_j^H]. \tag{9}$$

The objective of this work is to study the effect of VN and MN on the ergodic sum rate of the system defined as:

$$r_E = \sum_{j=1}^{L} \sum_{k=1}^{K} \log_2(1 + \gamma_{jk}). \tag{10}$$

Plugging $\mathbf{G}_j$ into (5), it follows that for both precoding schemes γ_{jk} depends on the statistical distribution of $\{\mathbf{h}_{jlk}\}$ and $\{\hat{\mathbf{h}}_{jlk}\}$. This makes it hard to compute γ_{jk} and, consequently, r_{jk} in closed-form. Therefore, a large system analysis is provided in the next section to find tight approximations for $\{\gamma_{jk}\}$.

[1]The following notation is used throughout the paper. Scalars are denoted by lower case letters whereas boldface lower (upper) case letters are used for vectors (matrices). We denote by $\mathbf{I}_N$ the identity matrix of order N and call $[\mathbf{A}]_{i,k}$ the (i,k)th element of the enclosed matrix. A random vector $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{C})$ is complex Gaussian distributed with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. The trace, transpose, conjugate transpose, real part, and expectation operators are denoted by $\mathtt{tr}(\cdot)$, $(\cdot)^T$, $(\cdot)^H$, $\mathtt{Re}(\cdot)$, and $\mathbb{E}[\cdot]$.

III. Large System Analysis

In what follows, asymptotic approximations (also called deterministic equivalents) of γ_{jk} for MRT and ZF with MN and VN are derived. These results can then be used to find tight approximations for the individual rates $\{r_{jk}\}$. By applying the continuous mapping theorem, the almost sure convergence of the results illustrated below implies that $r_{jk} - \overline{r}_{jk} \to 0$ almost surely with [3]

$$\overline{r}_{jk} = \log_2\left(1 + \overline{\gamma}_{jk}\right) \tag{11}$$

where $\overline{\gamma}_{jk}$ denotes one of the asymptotic approximations computed below.

A. Preliminary assumptions and results

We begin by assuming the following grow rate of system dimensions:

Assumption 1. *The dimensions N and K grow to infinity at the same pace, that is:*

$$1 < \liminf N/K \leq \limsup N/K < \infty. \tag{12}$$

For technical reasons, the following reasonable assumptions are also imposed on the system settings [3].

Assumption 2. *As $N, K \to \infty$, the correlation matrix $\forall j, l, k$ $\mathbf{\Theta}_{jlk}$ has uniformly bounded spectral norm on N, i.e.,*

$$\limsup ||\mathbf{\Theta}_{jlk}^{1/2}|| < \infty \quad \liminf \frac{1}{N}\mathrm{tr}\left(\mathbf{\Theta}_{jlk}\right) > 0. \tag{13}$$

Assumption 3. *There exists $\epsilon > 0$ such that, for all large N, we have $\lambda_{\min}(\frac{1}{N}\mathbf{H}_{jj}^H\mathbf{H}_{jj}) > \epsilon$ with probability 1.*

Let us also introduce the fundamental equations that will be needed to express asymptotic approximations of $\{\gamma_{jk}\}$ for ZF with MN and VN. The following set of equations:

$$u_{lk} = \frac{1}{N}\mathsf{tr}\left(\mathbf{\Phi}_{llk}\mathbf{T}_l\right) \tag{14}$$

with

$$\mathbf{T}_l = \left(\frac{1}{N}\sum_{i=1}^{K}\frac{\mathbf{\Phi}_{lli}}{u_{li}} + \mathbf{I}_N\right)^{-1} \tag{15}$$

admits a unique positive solution [10]. Moreover, we call

$$\mathbf{T}'_{lk} = \mathbf{T}_l\left(\frac{1}{N}\sum_{i=1}^{K}\frac{\mathbf{\Phi}_{lli}u'_{k,li}}{u_{li}^2} + \mathbf{\Phi}_{llk}\right)\mathbf{T}_l \tag{16}$$

where $\mathbf{u}'_{k,l} = [u'_{k,l1}, \ldots, u'_{k,lK}]^T \in \mathbb{C}^K$ is computed as

$$\mathbf{u}'_{k,l} = \left(\mathbf{I}_K - \mathbf{J}_l\right)^{-1}\mathbf{v}_{k,l} \tag{17}$$

with the entries of $\mathbf{J}_l \in \mathbb{C}^{K\times K}$ and $\mathbf{v}_{k,l} \in \mathbb{C}^K$ given by:

$$[\mathbf{J}_l]_{n,i} = \frac{1}{N^2}\frac{\mathsf{tr}\left(\mathbf{\Phi}_{lln}\mathbf{T}_l\mathbf{\Phi}_{lli}\mathbf{T}_l\right)}{u_{li}^2} \tag{18}$$

$$[\mathbf{v}_{k,l}]_i = \frac{1}{N}\mathsf{tr}\left(\mathbf{\Phi}_{lli}\mathbf{T}_l\mathbf{\Phi}_{llk}\mathbf{T}_l\right). \tag{19}$$

B. Matrix Normalization

An asymptotic expression of γ_{jk} for MRT with MN is given in [3] and is reported below for completeness.

Theorem 1. *[3, Theorem 4] Let Assumptions 1 – 3 hold true. If MRT with MN is used, then $\gamma_{jk} - \overline{\gamma}_{jk}^{(\mathrm{MRT-MN})} \to 0$ almost surely with*

$$\overline{\gamma}_{jk}^{(\mathrm{MRT-MN})} = \frac{\lambda_j\left(\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{jjk}\right)^2}{\frac{1}{N\rho_{\mathrm{dl}}} + \frac{1}{N}\sum_{l=1}^{L}\sum_{i=1}^{K}\lambda_l z_{jk,li} + \sum_{l=1,l\neq j}^{L}\lambda_l|\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{ljk}|^2} \tag{20}$$

where

$$\lambda_j = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{jjk}\right)^{-1} \tag{21}$$

$$z_{jk,li} = \frac{1}{N}\mathtt{tr}\mathbf{\Theta}_{ljk}\mathbf{\Phi}_{lli}. \tag{22}$$

Although ZF is not considered in [3] (only the regularized ZF is investigated), an asymptotic expression of γ_{jk} for ZF with MN can be retrieved from [3, Theorem 3] assuming $\mathbf{Z}_j^{\mathrm{dl}} = \mathbf{0}_N$ and letting $\varphi_l^{\mathrm{dl}} \to 0$. In doing so, we obtain:

Theorem 2. *Let Assumptions 1 – 3 hold true. If ZF with MN is used, then $\gamma_{jk} - \overline{\gamma}_{jk}^{(\mathrm{ZF-MN})} \to 0$ almost surely with*

$$\overline{\gamma}_{jk}^{(\mathrm{ZF-MN})} = \frac{\lambda_j}{\frac{1}{N\rho_{\mathrm{dl}}} + \sum_{l=1}^{L}\sum_{i=1}^{K}\lambda_l\frac{\epsilon_{jk,li}}{u_{li}^2} + \sum_{l=1,l\neq j}^{L}\lambda_l\frac{u_{ljk}^2}{u_{lk}^2}} \tag{23}$$

where

$$\lambda_j = \left(\frac{1}{K}\sum_{i=1}^{K}\frac{1}{u_{ji}}\right)^{-1} \tag{24}$$

and

$$u_{ljk} = \frac{1}{N}\mathsf{tr}\left(\mathbf{\Phi}_{ljk}\mathbf{T}_l\right) \tag{25}$$

$$\epsilon_{jk,li} = \frac{1}{N}\mathsf{tr}\left(\mathbf{\Theta}_{ljk}\mathbf{T}'_{li}\right) - 2\frac{u_{ljk}}{u_{lk}}\frac{1}{N}\mathsf{tr}\left(\mathbf{\Phi}_{ljk}\mathbf{T}'_{li}\right) + + \frac{u_{ljk}^2}{u_{lk}^2}\frac{1}{N}\mathsf{tr}\left(\mathbf{\Phi}_{llk}\mathbf{T}'_{li}\right). \tag{26}$$

C. Vector Normalization

We now provide asymptotic approximations for γ_{jk} with VN. The proofs follow from random matrix theory results as those used in [3], [10] and are omitted for space limitations.[2] We begin by considering MRT, for which we have the following result:

Theorem 3. *Let Assumptions 1 – 3 hold true. If MRT with VN is used, then $\gamma_{jk} - \overline{\gamma}_{jk}^{(\mathrm{MRT-VN})} \to 0$ almost surely with*

$$\overline{\gamma}_{jk}^{(\mathrm{MRT-VN})} = \frac{\vartheta_{jk}\left(\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{jjk}\right)^2}{\frac{1}{N\rho_{\mathrm{dl}}} + \frac{1}{N}\sum_{l=1}^{L}\sum_{i=1}^{K}\vartheta_{li}z_{jk,li} + \sum_{l=1,l\neq j}^{L}\vartheta_{lk}|\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{ljk}|^2} \tag{27}$$

[2]They will be provided in the extended version [14].

where

$$\vartheta_{li} = \Big(\frac{1}{N}\mathtt{tr}\mathbf{\Phi}_{lli}\Big)^{-1} \tag{28}$$

and $z_{jk,li}$ *is given in* (22)*.*

On the other hand, if ZF is employed we have that:

Theorem 4. *Let Assumptions 1 – 3 hold true. If ZF with VN is employed, then* $\gamma_{jk} - \overline{\gamma}_{jk}^{(\mathrm{ZF-VN})} \to 0$ *almost surely with*

$$\overline{\gamma}_{jk}^{(\mathrm{ZF-VN})} = \frac{u_{jk}}{\frac{1}{N\rho_{\mathrm{dl}}} + \frac{1}{N}\sum\limits_{l=1}^{L}\sum\limits_{i=1}^{K}\frac{\epsilon_{jk,li}}{u_{li}} + \sum\limits_{l=1,l\neq j}^{L}\frac{u_{ljk}^2}{u_{lk}}} \tag{29}$$

where u_{ljk} *and* $\epsilon_{jk,li}$ *are given by* (25) *and* (26)*, respectively.*

Note the above results can be achieved by considering vector normalized precoding in (5) and following the same procedure as in [3, Theorem 4] and [10, Theorem 3].

D. A case study

The above asymptotic results will be validated in Section IV and used to make comparisons among the different precoding schemes. As we shall see, VN will achieve better performance for both MRT and ZF in medium to high SNR regimes. In addition to this, they can be used to get some insights. To this end, let us focus on ZF and assume that $\mathbf{\Theta}_{jk} = d_{jlk}\mathbf{I}_N$ $\forall j,k$ where d_{jlk} accounts for the large scale fading channel attenuation. Therefore, (1) reduces to

$$\mathbf{h}_{jlk} = \sqrt{d_{jlk}}\mathbf{z}_{jk}. \tag{30}$$

Let

$$\nu_{jk} = \frac{1}{N\rho_{\mathrm{dl}}} + \frac{K}{N}\sum_{l=1}^{L} d_{ljk}\left(1 - \frac{d_{ljk}}{\alpha_{lk}}\right) \tag{31}$$

with $\alpha_{lk} = \sum_{n=1}^{L} d_{lnk} + \frac{1}{\rho_{\mathrm{tr}}}$ accounting for channel estimation errors and pilot contamination. Then, we have that:

Corollary 1. *If the channel is modeled as in* (30) *and ZF is used, then*

$$\overline{\gamma}_{jk}^{(\mathrm{ZF-VN})} = \frac{\frac{d_{jjk}^2}{\alpha_{jk}}\bar{u}}{\nu_{jk} + \sum\limits_{l=1,l\neq j}^{L}\frac{d_{ljk}^2}{\alpha_{lk}}\bar{u}} \tag{32}$$

$$\overline{\gamma}_{jk}^{(\mathrm{ZF-MN})} = \frac{\lambda_j}{\nu_{jk} + \sum\limits_{l=1,l\neq j}^{L}\lambda_l\frac{d_{ljk}^2}{d_{llk}^2}} \tag{33}$$

with $\lambda_j = \frac{\bar{u}}{\frac{1}{K}\sum\limits_{i=1}^{K}\frac{\alpha_{ji}}{d_{jji}^2}}$ *and* $\bar{u} = 1 - \frac{K}{N}$.

Proof: See Appendix A. ∎

As seen, that the residual interference term is the same for both normalization methods and given by ν_{jk}. On the other hand, the term due to pilot contamination is different. Applied to practical networks, the above results may lead to important insights on how the two normalization methods affect the network performance (more details on these aspects will be given in the extended version [14]). Consider for example a single-cell network. Then, under the assumption of perfect channel knowledge (i.e., $\rho_{\mathrm{tr}} \to \infty$) we have that:

Lemma 1. *If* $L = 1$ *and perfect channel knowledge is assumed, then ZF with VN outperforms ZF with MN and the rate gap* $\Delta r \geq 0$ *is given by*

$$\Delta r = \sum_k \log\left(1 + \frac{1}{\frac{1}{N\rho_{\mathrm{dl}}\bar{u}}\frac{1}{d_k}}\right) - K\log\left(1 + \frac{1}{\frac{1}{N\rho_{\mathrm{dl}}\bar{u}}\frac{1}{K}\sum\limits_{i=1}^{K}\frac{1}{d_i}}\right). \tag{34}$$

Proof: From Corollary 1, setting $L = 1$ and assuming $\rho_{\mathrm{tr}} \to \infty$ yields $\nu_k = \frac{1}{N\rho_{\mathrm{dl}}}$ and $\frac{d_k^2}{\alpha_k} = d_k$. Then, applying the Jensen' inequality (observe that $\log(1 + 1/x)$ is a convex function) the result easily follows. ∎

A similar result holds true in the high SINR regime:

Lemma 2. *If* $L = 1$ *and the high SINR regime is considered, then ZF with VN outperforms ZF with MN and the rate gap* $\Delta r \geq 0$ *is given by*

$$\Delta r = K\log\left(\frac{1}{K}\sum_{i=1}^{K}\frac{1 + \frac{1}{d_i\rho_{\mathrm{tr}}}}{d_i}\right) - \sum_{k=1}^{K}\log\left(\frac{1 + \frac{1}{d_k\rho_{\mathrm{tr}}}}{d_k}\right). \tag{35}$$

Proof: The proof easily follows from Corollary 1 when $L = 1$ and the rate is approximated as $\log(\overline{\gamma}_k)$. Applying the Jensen's inequality, it follows that $\Delta r \geq 0$. ∎

Observe that if $\forall k$ $d_k = d$ in (34) or (35) then $\Delta r = 0$. This means that the benefits of VN vanish as UEs have similar large scale channel statics. This confirms the intuition that large scale channel attenuation have a fundamental effect on the normalization methods.

IV. Numerical results

Monte-Carlo (MC) simulations are now used to validate the asymptotic analysis for different values of N and K. We consider a multicell network composed of $L = 7$ cells, one in the center and six around. Each cell radius is normalized to one. The UEs are randomly distributed within each cell. The channel is modeled as in [15]. In particular, we assume that the matrices $\mathbf{\Theta}_{ljk}^{1/2}$ are given by

$$\mathbf{\Theta}_{ljk}^{1/2} = \sqrt{d_{ljk}}\mathbf{A} \tag{36}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_N)] \in \mathbb{C}^N$ with $\mathbf{a}(\theta_i)$ given by

$$\mathbf{a}(\theta_i) = \frac{1}{\sqrt{N}}[1, e^{-i2\pi\omega\sin(\theta_i)}, \ldots, e^{-i2\pi\omega(N-1)\sin(\theta_i)}]^T \tag{37}$$

where $\omega = 0.3$ is the antenna spacing and $\theta_i = -\pi/2 + (i-1)\pi/N$. Also, d_{ljk} is the large scale attenuation, which is modelled as $d_{ljk} = d_0/x_{ljk}^{\beta}$ where x_{ljk} denotes the distance of UE k in cell j from BS l, $\beta = 3.7$ is the path loss exponent and $d_0 = 10^{-3.53}$. We let $\rho_{\mathrm{tr}} = 6$ dB and $\rho_{\mathrm{dl}} = 10$ dB,

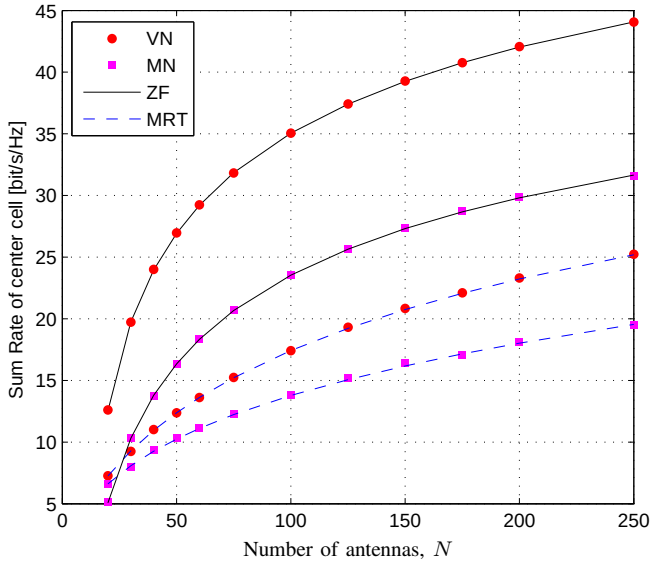

Fig. 1. Ergodic achievable sum rate of center cell for MRT and ZF with VN and MN when $K = 10$.

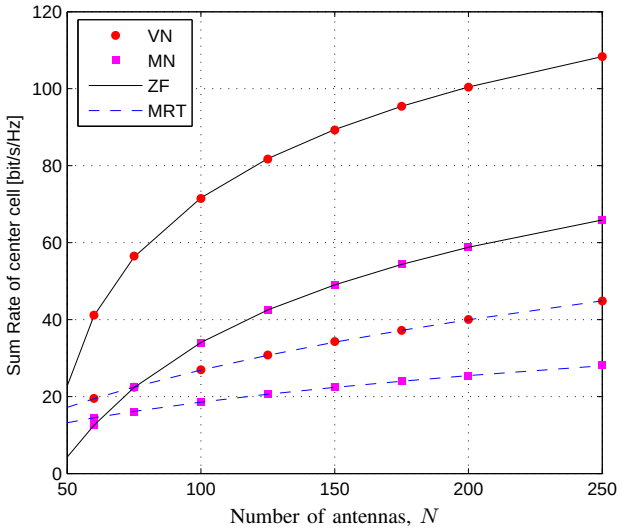

Fig. 2. Ergodic achievable sum rate of center cell for MRT and ZF with VN and MN when $K = 10$.

which corresponds to a practical setting [3], [8]. The results are obtained for 1000 different channel realizations and 10 different UE distributions.

Figs. 1 and 2 illustrate the sum rate of the investigated cell vs. N when $K = 10$ and 30, respectively. Bold lines are obtained using the asymptotic results of Theorems 1 – 4 and markers are obtained from MC simulation. As seen, the approximation achieved from the deterministic equivalents (DEs) matches very well with MC simulations even for small values of K and N, e.g., $K = 10$ and $N = 20$. The results

TABLE I
ASYMPTOTIC CALUES OF THE SUM RATE OF THE CENTER CELL FOR DIFFERENT PRECODING SCHEMES.

No. of ant.	Precoding Scheme	$K = 10$	$K = 20$	$K = 30$
$N = 75$	MRT-VN	15.24	19.59	22.44
	MRT-MN	12.25	11.50	16.10
	ZF-VN	31.83	49.89	56.50
	ZF-MN	20.68	24.07	22.43
	RZF-MN	22.38	29.08	33.17
$N = 100$	MRT-VN	17.41	23.18	26.96
	MRT-MN	13.79	12.95	18.61
	ZF-VN	35.06	57.58	71.48
	ZF-MN	23.57	30.59	33.97
	RZF-MN	25.16	34.90	42.20
$N = 150$	MRT-VN	20.84	28.70	34.28
	MRT-MN	16.37	15.28	22.55
	ZF-VN	39.28	67.64	89.30
	ZF-MN	27.33	39.22	48.93
	RZF-MN	28.84	42.94	55.38
$N = 200$	MRT-VN	23.3	32.94	40.04
	MRT-MN	18.11	17.43	25.39
	ZF-VN	42.06	74.03	100.42
	ZF-MN	29.89	44.86	58.80
	RZF-MN	31.35	48.34	65.57
$N = 250$	MRT-VN	25.23	36.74	44.84
	MRT-MN	19.47	19.10	28.26
	ZF-VN	44.06	78.75	108.33
	ZF-MN	31.61	49.00	65.95
	RZF-MN	33.06	52.37	71.33

show that VN largely outperforms MN for both MRT and ZF.

Since the asymptotic results of Theorems 1 – 4 match very well with MC simulations, they are now used to make comparisons among the different precoding schemes. For completeness, we also consider the RZF with MN whose asymptotic analysis is provided in [3, Theorem 3]. Table I reports the DEs of the investigated schemes for different values of K and N when $\rho_{\text{tr}} = 6$ dB and $\rho_{\text{dl}} = 10$ dB. As seen, if $K = 10$ ZF-VN requires only $N = 75$ antennas to nearly achieve the same sum rate of ZF-MN or RZF-MN with $N = 250$. A similar result holds true for $K = 20$ and 30. This means that VN allows to reduce the number of required antennas by a factor of 3 while providing the same sum rate. Similar conclusions can be drawn for MRT. From Table I, it also follows that for a given N MRT-VN improves the sum rate by 90% (or even more) compared to MRT-MN. For ZF and RZF, the improvement is around 200% and 70%, respectively.

V. CONCLUSIONS

We investigated the effect of vector and matrix normalizations on the DL sum rate of MRT and ZF precoding schemes in massive MIMO systems with a very general channel model under the assumption that data transmission was affected by channel estimation errors and pilot contamination. Recent results from random matrix theory were used to find asymptotic approximations for the sum rate of the investigated precoding schemes. These results were used to make comparisons among the two normalization methods. For a practical network setting operating from a medium to high SNR regime, it turned out that vector normalization provides better performance than

matrix normalization for both MRT and ZF. In particular, we showed that the number of antennas N required to achieve a target sum rate with vector normalization is smaller than the one required by matrix normalization by a factor of 3 or 4. The analysis can be easily extended to other precoding techniques such the RZF precoding and the MMSE scheme proposed in [8]. Furthermore, it can be used to get important insights into the way the two normalization methods perform, especially with respect to the pilot reuse factor, large-scale attenuations, operating SNR as well as the number of antennas and UEs. More details on all the aspects will be given in the extended version.

APPENDIX A

For simplicity, we only consider ZF with VN. The same steps can be used for ZF with MN. If the channel is modelled as in (30), then $\mathbf{\Theta}_{ljk} = d_{ljk}$ and

$$\mathbf{\Phi}_{ljk} = \frac{d_{llk} d_{ljk}}{\alpha_{lk}} \mathbf{I}_N \tag{38}$$

with $\alpha_{lk} = \sum_{n=1}^{L} d_{lnk} + \frac{1}{\rho_{\text{tr}}}$. Plugging (38) into (14) and (15) yields

$$u_{lk} = \frac{d_{llk}^2}{\alpha_{lk}} \frac{1}{N} \mathsf{tr}\left(\mathbf{T}_l\right) \tag{39}$$

with

$$\mathbf{T}_l = \left(\frac{1}{N} \sum_{i=1}^{K} \frac{1}{\frac{1}{N}\mathsf{tr}\left(\mathbf{T}_l\right)} + 1 \right)^{-1} \mathbf{I}_N. \tag{40}$$

Call $\bar{u} = \frac{1}{N}\mathsf{tr}\left(\mathbf{T}_l\right)$. Therefore, we have that

$$\bar{u} = \frac{1}{N}\mathsf{tr}\left(\mathbf{T}_l\right) = \left(\frac{K}{N} \frac{1}{\bar{u}} + 1 \right)^{-1}. \tag{41}$$

Solving with respect to $\bar{u}$ yields $\bar{u} = 1 - \frac{K}{N}$. Then, we eventually have that

$$u_{lk} = \frac{d_{llk}^2}{\alpha_{lk}} \bar{u} \tag{42}$$

and also

$$u_{ljk} = \frac{d_{llk} d_{ljk}}{\alpha_{lk}} \bar{u}. \tag{43}$$

Therefore, the pilot contamination term in γ_{jk} reduces to

$$\sum_{l=1, l\neq j}^{L} \frac{u_{ljk}^2}{u_{lk}} = \sum_{l=1, l\neq j}^{L} \frac{d_{ljk}^2}{\alpha_{lk}} \bar{u}. \tag{44}$$

Let's now compute $[\mathbf{J}_l]_{n,i}$ defined as in (18). Using the above results yields

$$[\mathbf{J}_l]_{n,i} = \frac{1}{N^2} \frac{d_{lln}^2}{\alpha_{ln}} \frac{d_{lli}^2}{\alpha_{li}} \frac{1}{u_{li}^2} \mathsf{tr}\left(\mathbf{T}^2\right) = \frac{1}{N} \frac{d_{lln}^2}{\alpha_{ln}} \frac{\alpha_{li}}{d_{lli}^2}. \tag{45}$$

Similarly, we have that

$$[\mathbf{v}_{k,l}]_i = \frac{d_{lli}^2}{\alpha_{li}} \frac{d_{llk}^2}{\alpha_{lk}} \bar{u}^2. \tag{46}$$

In compact form, we may write $\mathbf{J}_l$ and $\mathbf{v}_{l,k}$ as

$$\mathbf{J}_l = \frac{1}{N} \mathbf{a}_l \mathbf{b}_l^T \qquad \mathbf{v}_{l,k} = \frac{d_{llk}^2}{\alpha_{lk}} \bar{u}^2 \mathbf{a}_l \tag{47}$$

with $[\mathbf{a}_l]_i = d_{lli}^2/\alpha_{li}$ and $[\mathbf{b}_l]_i = 1/[\mathbf{a}_l]_i$. Then, we have that (applying the matrix inversion lemma)

$$\mathbf{u}'_{k,l} = \frac{d_{llk}^2}{\alpha_{lk}} \bar{u}^2 \left(\mathbf{I}_K - \frac{1}{N} \mathbf{a}_l \mathbf{b}_l^T \right)^{-1} \mathbf{a}_l = \frac{d_{llk}^2}{\alpha_{lk}} \bar{u} \mathbf{a}_l = u_{lk} \mathbf{a}_l. \tag{48}$$

Plugging the above result into (16) produces

$$\mathbf{T}'_{li} = \frac{d_{lli}^2}{\alpha_{li}} \mathbf{T} \left(\frac{K}{N} \frac{1}{\bar{u}} + 1 \right) \mathbf{T} = \frac{d_{lli}^2}{\alpha_{li}} \bar{u} \mathbf{I}_N = u_{li} \mathbf{I}_N. \tag{49}$$

We are thus left with evaluating (26). Using the above results yields

$$\begin{aligned} \epsilon_{jk,li} = {} & \frac{d_{ljk}}{N^2} \mathsf{tr}\left(\mathbf{T}'_{li}\right) - 2 \frac{d_{ljk}}{d_{llk}} \frac{1}{N^2} \mathsf{tr}\left(\mathbf{\Phi}_{ljk} \mathbf{T}'_{li}\right) + \\ & + \frac{d_{ljk}^2}{d_{llk}^2} \frac{1}{N^2} \mathsf{tr}\left(\mathbf{\Phi}_{llk} \mathbf{T}'_{li}\right) \end{aligned} \tag{50}$$

from which, using (38) and (49), we obtain

$$\epsilon_{jk,li} = \frac{d_{ljk}}{N} u_{li} - \frac{d_{ljk}^2}{\alpha_{lk}} \frac{1}{N} u_{li}.$$

Therefore, we have that

$$\frac{\epsilon_{jk,li}}{u_{li}} = \frac{1}{N} d_{ljk} \left(1 - \frac{d_{ljk}}{\alpha_{lk}} \right). \tag{51}$$

Plugging (42), (44) and (51) into (29) produces

$$\sum_{l=1}^{L} \sum_{i=1}^{K} \frac{\epsilon_{jk,li}}{u_{li}} = \frac{K}{N} \sum_{l=1}^{L} d_{ljk} \left(1 - \frac{d_{ljk}}{\alpha_{lk}} \right). \tag{52}$$

Collecting all the above results together completes the proof.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590–3600, 2010.

[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 Processing Magazine*, vol. 30, no. 1, pp. 40 – 60, Jan 2013.

[3] 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, 2013.

[4] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186 – 195, February 2014.

[5] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. C. Wong, V. Öwall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *Proc. of IEEE Globecom Workshop - Massive MIMO: From Theory to Practice*, 2014.

[6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 172–179, 2013.

[7] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *CoRR*, vol. abs/1505.03682, 2015. [Online]. Available: http://arxiv.org/abs/1505.03682

[8] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE J. Sel. Areas Commun.*, 2015, submitted. [Online]. Available: http://arxiv.org/abs/1505.01181

[9] Y. Lim, C. Chae, and G. Caire, "Performance analysis of massive MIMO for cell-boundary users," *IEEE Trans. Wireless Commun.*, 2015, to appear.

[10] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback," *Information Theory, IEEE Transactions on*, vol. 58, no. 7, pp. 4509–4537, 2012.

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

[12] 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, 2011.

[13] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[14] M. Sadeghi, L. Sanguinetti, R. Couillet, and C. Yuen, "Large system analysis of massive MIMO systems with different power normalization techniques," *IEEE Trans. Wireless Commun.*, 2015, to be submitted.

[15] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser mimo systems for physical channel models," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3464–3467.