Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems

Hien Quoc Ngo, Erik G. Larsson and Thomas L. Marzetta

Hien Quoc Ngo, Erik G. Larsson and Thomas L. Marzetta, Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems, 2013, IEEE Transactions on Communications, (61), 4, 1436-1449.
http://dx.doi.org/10.1109/TCOMM.2013.020413.110848

Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems

Hien Quoc Ngo, Erik G. Larsson, and Thomas L. Marzetta

***Abstract*—A multiplicity of autonomous terminals simultaneously transmits data streams to a compact array of antennas. The array uses imperfect channel-state information derived from transmitted pilots to extract the individual data streams. The power radiated by the terminals can be made inversely proportional to the square-root of the number of base station antennas with no reduction in performance. In contrast if perfect channel-state information were available the power could be made inversely proportional to the number of antennas. Lower capacity bounds for maximum-ratio combining (MRC), zero-forcing (ZF) and minimum mean-square error (MMSE) detection are derived. An MRC receiver normally performs worse than ZF and MMSE. However as power levels are reduced, the cross-talk introduced by the inferior maximum-ratio receiver eventually falls below the noise level and this simple receiver becomes a viable option. The tradeoff between the energy efficiency (as measured in bits/J) and spectral efficiency (as measured in bits/channel use/terminal) is quantified for a channel model that includes small-scale fading but not large-scale fading. It is shown that the use of moderately large antenna arrays can improve the spectral and energy efficiency with orders of magnitude compared to a single-antenna system.**

***Index Terms*—Energy efficiency, spectral efficiency, multiuser MIMO, very large MIMO systems**

I. Introduction

In multiuser multiple-input multiple-output (MU-MIMO) systems, a base station (BS) equipped with multiple antennas serves a number of users. Such systems have attracted much attention for some time now [2]. Conventionally, the communication between the BS and the users is performed by orthogonalizing the channel so that the BS communicates with each user in separate time-frequency resources. This is not optimal from an information-theoretic point of view, and higher rates can be achieved if the BS communicates with several users in the same time-frequency resource [3], [4]. However, complex techniques to mitigate interuser interference must then be used, such as maximum-likelihood multiuser detection on the uplink [5], or "dirty-paper coding" on the downlink [6], [7].

Manuscript received Dec. 15, 2011; revised May 2, 2012 and Aug. 20, 2012; accepted Nov. 1, 2012. The associate editor coordinating the review of this paper and approving it for publication was B. Clerckx. This work was supported in part by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), and ELLIIT. E. Larsson was a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation. Parts of this work were presented at the 2011 Allerton Conf. Commun., Control and Comput. [1].

H. Q. Ngo and E. G. Larsson are with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden (Email: nqhien@isy.liu.se; egl@isy.liu.se).

T. L. Marzetta is with Bell Laboratories, Alcatel-Lucent, 600 Moutain Avenue, Murray Hill, NJ 07974, USA (Email: tom.marzetta@alcatel-lucent.com).

Digital Object Identifier xxx/xxx

Recently, there has been a great deal of interest in MU-MIMO with *very large antenna arrays* at the BS. Very large arrays can substantially reduce intracell interference with simple signal processing [8]. We refer to such systems as "very large MU-MIMO systems" here, and with very large we mean arrays comprising say a hundred, or a few hundreds, of antennas, simultaneously serving tens of users. The design and analysis of very large MU-MIMO systems is a fairly new subject that is attracting substantial interest [8]–[11]. The vision is that each individual antenna can have a small physical size, and be built from inexpensive hardware. With a very large antenna array, things that were random before start to look deterministic. As a result, the effect of small-scale fading can be averaged out. Furthermore, when the number of BS antennas grows large, the random channel vectors between the users and the BS become pairwisely orthogonal [10]. In the limit of an infinite number of antennas, with simple matched filter processing at the BS, uncorrelated noise and intracell interference disappear completely [8]. Another important advantage of large MIMO systems is that they enable us to reduce the transmitted power. On the uplink, reducing the transmit power of the terminals will drain their batteries slower. On the downlink, much of the electrical power consumed by a BS is spent by power amplifiers and associated circuits and cooling systems [12]. Hence reducing the emitted RF power would help in cutting the electricity consumption of the BS.

This paper analyzes the potential for power savings on the uplink of very large MU-MIMO systems. We derive new capacity bounds of the uplink for finite number of BS antennas. While it is well known that MIMO technology can offer improved power efficiency, owing to both array gains and diversity effects [13], we are not aware of any work that analyzes power efficiency of MU-MIMO systems with receiver structures that are realistic for very large MIMO.[1] We consider both single-cell and multicell systems, but focus on the analysis of single-cell MU-MIMO systems since: i) the results are easily comprehensible; ii) it bounds the performance of a multicell system; and iii) the single-cell performance can be actually attained if one uses successively less-aggressive frequency-reuse (e.g., with reuse factor 3, or 7). Our results are different from recent results in [14] and [15]. In [14] and [15], the authors derived a deterministic equivalent of the SINR assuming that the number of transmit antennas and the number

[1]After submitting this work, other papers have also addressed the tradeoff between spectral and energy efficiency in MU-MIMO. An analysis related to the one presented here but for the downlink was given in [16]. However, the analysis of the downlink is quantitatively and qualitatively different both in what concerns systems aspects and the corresponding the capacity bounds.

of users go to infinity but their ratio remains bounded for the downlink of network MIMO systems using a sophisticated scheduling scheme and MISO broadcast channels using zero-forcing (ZF) precoding, respectively. The paper makes the following specific contributions:

- We show that, when the number of BS antennas M grows without bound, we can reduce the transmitted power of each user proportionally to $1/M$ if the BS has perfect channel state information (CSI), and proportionally to $1/\sqrt{M}$ if CSI is estimated from uplink pilots. This holds true even when using simple, linear receivers. We also derive closed-form lower bounds on the uplink achievable rates for finite M, for the cases of perfect and imperfect CSI, assuming MRC, ZF, and minimum mean-squared error (MMSE) receivers, respectively. See Section III.
- We study the tradeoff between spectral efficiency and energy efficiency. For imperfect CSI, in the low transmit power regime, we can simultaneously increase the spectral-efficiency and energy-efficiency. We further show that in large-scale MIMO, very high spectral efficiency can be obtained even with simple MRC processing at the same time as the transmit power can be cut back by orders of magnitude and that this holds true even when taking into account the losses associated with acquiring CSI from uplink pilots. MRC also has the advantage that it can be implemented in a distributed manner, i.e., each antenna performs multiplication of the received signals with the conjugate of the channel, without sending the entire baseband signal to the BS for processing. Quantitatively, our energy-spectral efficiency tradeoff analysis incorporates the effects of small-scale fading but neglects those of large-scale fading, leaving an analysis of the effect of large-scale fading for future work. See Section IV.

II. System Model and Preliminaries

A. MU-MIMO System Model

We consider the uplink of a MU-MIMO system. The system includes one BS equipped with an array of M antennas that receive data from K single-antenna users. The nice thing about single-antenna users is that they are inexpensive, simple, and power-efficient, and each user still gets typically high throughput. Furthermore, the assumption that users have single antennas can be considered as a special case of users having multiple antennas when we treat the extra antennas as if they were additional autonomous users.[2] The users transmit their data in the same time-frequency resource. The $M \times 1$ received vector at the BS is

$$\boldsymbol{y} = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

[2]Note that under the assumptions on favorable propagation (see Section II-C), having n autonomous single-antenna users or having one n-antenna user (where the antennas cooperate in the encoding), represent two cases with equal energy and spectral efficiency. To see why, consider two cases: the case of 2 autonomous single-antenna users of which each spends power P, and the case of one dual-antenna user with a total power constraint of $2P$. Then, the sum rates for the two cases are the same and equal to $\log_2\left(1+\frac{P\|\boldsymbol{h}_1\|^2}{N_0}\right)+\log_2\left(1+\frac{P\|\boldsymbol{h}_2\|^2}{N_0}\right)=\log_2\left|\boldsymbol{I}+\frac{1}{N_0}[\boldsymbol{h}_1\ \boldsymbol{h}_2]\begin{bmatrix}P & 0\\ 0 & P\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_1^H\\ \boldsymbol{h}_2^H\end{bmatrix}\right|$, where $\boldsymbol{h}_i$ is the channel vector between the ith user (or ith antenna) to the BS, and N_0 is the variance of noise.

where $\boldsymbol{G}$ represents the $M \times K$ channel matrix between the BS and the K users, i.e., $g_{mk} \triangleq [\boldsymbol{G}]_{mk}$ is the channel coefficient between the mth antenna of the BS and the kth user; $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}$ is the vector of symbols simultaneously transmitted by the K users (the average transmitted power of each user is p_{u}); and $\boldsymbol{n}$ is a vector of additive white, zero-mean Gaussian noise. We take the noise variance to be 1, to minimize notation, but without loss of generality. With this convention, p_{u} has the interpretation of normalized "transmit" SNR and is therefore dimensionless. The model (1) also applies to wideband channels handled by OFDM over restricted intervals of frequency.

The channel matrix $\boldsymbol{G}$ models independent fast fading, geometric attenuation, and log-normal shadow fading. The coefficient g_{mk} can be written as

$$g_{mk} = h_{mk}\sqrt{\beta_k}, \quad m = 1, 2, ..., M \tag{2}$$

where h_{mk} is the fast fading coefficient from the kth user to the mth antenna of the BS. $\sqrt{\beta_k}$ models the geometric attenuation and shadow fading which is assumed to be independent over m and to be constant over many coherence time intervals and known a priori. This assumption is reasonable since the distances between the users and the BS are much larger than the distance between the antennas, and the value of β_k changes very slowly with time. Then, we have

$$\boldsymbol{G} = \boldsymbol{H}\boldsymbol{D}^{1/2} \tag{3}$$

where $\boldsymbol{H}$ is the $M \times K$ matrix of fast fading coefficients between the K users and the BS, i.e., $[\boldsymbol{H}]_{mk} = h_{mk}$, and $\boldsymbol{D}$ is a $K \times K$ diagonal matrix, where $[\boldsymbol{D}]_{kk} = \beta_k$.

B. Review of Some Results on Very Long Random Vectors

We review some limit results for random vectors [17] that will be useful later on. Let $\boldsymbol{p} \triangleq [p_1\ ...\ p_n]^T$ and $\boldsymbol{q} \triangleq [q_1\ ...\ q_n]^T$ be mutually independent $n \times 1$ vectors whose elements are i.i.d. zero-mean random variables (RVs) with $\mathbb{E}\left\{|p_i|^2\right\} = \sigma_p^2$, and $\mathbb{E}\left\{|q_i|^2\right\} = \sigma_q^2$, $i = 1, ..., n$. Then from the law of large numbers, we have

$$\frac{1}{n}\boldsymbol{p}^H\boldsymbol{p} \overset{a.s.}{\to} \sigma_p^2, \text{ and } \frac{1}{n}\boldsymbol{p}^H\boldsymbol{q} \overset{a.s.}{\to} 0, \text{ as } n \to \infty. \tag{4}$$

where $\overset{a.s.}{\to}$ denotes the almost sure convergence. Also, from the Lindeberg-Lévy central limit theorem, we have

$$\frac{1}{\sqrt{n}}\boldsymbol{p}^H\boldsymbol{q} \overset{d}{\to} \mathcal{CN}\left(0, \sigma_p^2\sigma_q^2\right), \text{ as } n \to \infty \tag{5}$$

where $\overset{d}{\to}$ denotes convergence in distribution.

C. Favorable Propagation

Throughout the rest of the paper, we assume that the fast fading coefficients, i.e., the elements of $\boldsymbol{H}$ are i.i.d. RVs with zero mean and unit variance. Then the conditions in (4)–(5) are satisfied with $\boldsymbol{p}$ and $\boldsymbol{q}$ being any two distinct columns of $\boldsymbol{G}$. In this case we have

$$\frac{\boldsymbol{G}^H\boldsymbol{G}}{M} = \boldsymbol{D}^{1/2}\frac{\boldsymbol{H}^H\boldsymbol{H}}{M}\boldsymbol{D}^{1/2} \approx \boldsymbol{D}, \quad M \gg K$$

and we say that we have *favorable propagation*. Clearly, if all fading coefficients are i.i.d. and zero mean, we have favorable

propagation. Recent channel measurements campaigns have shown that multiuser MIMO systems with large antenna arrays have characteristics that approximate the favorable-propagation assumption fairly well [10], and therefore provide experimental justification for this assumption.

To understand why favorable propagation is desirable, consider an $M \times K$ uplink (multiple-access) MIMO channel $\boldsymbol{H}$, where $M \geq K$, neglecting for now path loss and shadowing factors in $\boldsymbol{D}$. This channel can offer a sum-rate of

$$R = \sum_{k=1}^{K} \log_2 \left(1 + p_{\mathrm{u}} \lambda_k^2\right) \tag{6}$$

where p_{u} is the power spent per terminal and $\{\lambda_k\}_{k=1}^{K}$ are the singular values of $\boldsymbol{H}$, see [13]. If the channel matrix is normalized such that $|H_{ij}| \sim 1$ (where $\sim$ means equality of the order of magnitude), then $\sum_{k=1}^{K} \lambda_k^2 = \|\boldsymbol{H}\|^2 \approx MK$. Under this constraint the rate R is bounded as

$$\log_2 \left(1 + MKp_{\mathrm{u}}\right) \leq R \leq K \log_2 \left(1 + Mp_{\mathrm{u}}\right). \tag{7}$$

The lower bound (left inequality) is satisfied with equality if $\lambda_1^2 = MK$ and $\lambda_2^2 = \cdots = \lambda_K^2 = 0$ and corresponds to a rank-one (line-of-sight) channel. The upper bound (right inequality) is achieved if $\lambda_1^2 = \cdots = \lambda_K^2 = M$. This occurs if the columns of $\boldsymbol{H}$ are mutually orthogonal and have the same norm, which is the case when we have favorable propagation.

III. Achievable Rate and Asymptotic ($M \to \infty$) Power Efficiency

By using a large antenna array, we can reduce the transmitted power of the users as M grows large, while maintaining a given, desired quality-of-service. In this section, we quantify this potential for power decrease, and derive achievable rates of the uplink. Theoretically, the BS can use the maximum-likelihood detector to obtain optimal performance. However, the complexity of this detector grows exponentially with K. The interesting operating regime is when both M and K are large, but M is still (much) larger than K, i.e., $1 \ll K \ll M$. It is known that in this case, linear detectors (MRC, ZF and MMSE) perform fairly well [8] and therefore we will restrict consideration to those detectors in this paper. We treat the cases of perfect CSI (Section III-A) and estimated CSI (Section III-B) separately.

A. Perfect Channel State Information

We first consider the case when the BS has perfect CSI, i.e. it knows $\boldsymbol{G}$. Let $\boldsymbol{A}$ be an $M \times K$ linear detector matrix which depends on the channel $\boldsymbol{G}$. By using the linear detector, the received signal is separated into streams by multiplying it with $\boldsymbol{A}^H$ as follows

$$\boldsymbol{r} = \boldsymbol{A}^H \boldsymbol{y}. \tag{8}$$

We consider three conventional linear detectors MRC, ZF, and MMSE, i.e.,

$$\boldsymbol{A} = \begin{cases} \boldsymbol{G} & \text{for MRC} \\ \boldsymbol{G}\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1} & \text{for ZF} \\ \boldsymbol{G}\left(\boldsymbol{G}^H\boldsymbol{G} + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K\right)^{-1} & \text{for MMSE} \end{cases} \tag{9}$$

From (1) and (8), the received vector after using the linear detector is given by

$$\boldsymbol{r} = \sqrt{p_{\mathrm{u}}}\boldsymbol{A}^H\boldsymbol{G}\boldsymbol{x} + \boldsymbol{A}^H\boldsymbol{n}. \tag{10}$$

Let r_k and x_k be the kth elements of the $K \times 1$ vectors $\boldsymbol{r}$ and $\boldsymbol{x}$, respectively. Then,

$$r_k = \sqrt{p_{\mathrm{u}}}\boldsymbol{a}_k^H\boldsymbol{g}_k x_k + \sqrt{p_{\mathrm{u}}} \sum_{i=1, i\neq k}^{K} \boldsymbol{a}_k^H\boldsymbol{g}_i x_i + \boldsymbol{a}_k^H\boldsymbol{n} \tag{11}$$

where $\boldsymbol{a}_k$ and $\boldsymbol{g}_k$ are the kth columns of the matrices $\boldsymbol{A}$ and $\boldsymbol{G}$, respectively. For a fixed channel realization $\boldsymbol{G}$, the noise-plus-interference term is a random variable with zero mean and variance $p_{\mathrm{u}} \sum_{i=1,i\neq k}^{K} |\boldsymbol{a}_k^H\boldsymbol{g}_i|^2 + \|\boldsymbol{a}_k\|^2$. By modeling this term as additive Gaussian noise independent of x_k we can obtain a lower bound on the achievable rate. Assuming further that the channel is ergodic so that each codeword spans over a large (infinite) number of realizations of the fast-fading factor of $\boldsymbol{G}$, the ergodic achievable uplink rate of the kth user is

$$R_{\mathrm{P},k} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}|\boldsymbol{a}_k^H\boldsymbol{g}_k|^2}{p_{\mathrm{u}} \sum\limits_{i=1,i\neq k}^{K} |\boldsymbol{a}_k^H\boldsymbol{g}_i|^2 + \|\boldsymbol{a}_k\|^2}\right)\right\}. \tag{12}$$

To approach this capacity lower bound, the message has to be encoded over many realizations of all sources of randomness that enter the model (noise and channel). In practice, assuming wideband operation, this can be achieved by coding over the frequency domain, using, for example coded OFDM.

Proposition 1: Assume that the BS has perfect CSI and that the transmit power of each user is scaled with M according to $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$, where E_{u} is fixed. Then,[3]

$$R_{\mathrm{P},k} \to \log_2\left(1 + \beta_k E_{\mathrm{u}}\right), M \to \infty. \tag{13}$$

Proof: We give the proof for the case of an MRC receiver. With MRC, $\boldsymbol{A} = \boldsymbol{G}$ so $\boldsymbol{a}_k = \boldsymbol{g}_k$. From (12), the achievable uplink rate of the kth user is

$$R_{\mathrm{P},k}^{\mathtt{mrc}} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^4}{p_{\mathrm{u}} \sum_{i=1,i\neq k}^{K} |\boldsymbol{g}_k^H\boldsymbol{g}_i|^2 + \|\boldsymbol{g}_k\|^2}\right)\right\}. \tag{14}$$

Substituting $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$ into (14), and using (4), we obtain (13). By using the law of large numbers, we can arrive at the same result for the ZF and MMSE receivers. Note from (3) and (4) that when M grows large, $\frac{1}{M}\boldsymbol{G}^H\boldsymbol{G}$ tends to $\boldsymbol{D}$, and hence the ZF and MMSE filters tend to that of the MRC. ■

Proposition 1 shows that with perfect CSI at the BS and a large M, the performance of a MU-MIMO system with M antennas at the BS and a transmit power per user of E_{u}/M is equal to the performance of a SISO system with transmit power E_{u}, without any intra-cell interference and without any fast fading. In other words, by using a large number of BS antennas, we can scale down the transmit power proportionally to $1/M$. At the same time we increase the spectral efficiency K times by simultaneously serving K users in the same time-frequency resource.

[3]As mentioned after (1), p_{u} has the interpretation of normalized transmit SNR, and it is dimensionless. Therefore E_{u} is dimensionless too.

1) Maximum-Ratio Combining: For MRC, from (14), by the convexity of $\log_2\left(1+\frac{1}{x}\right)$ and using Jensen's inequality, we obtain the following lower bound on the achievable rate:

$$R_{\mathrm{P},k}^{\mathtt{mrc}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} \triangleq \log_2\left(1+\left(\mathbb{E}\left\{\frac{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\boldsymbol{g}_k^H\boldsymbol{g}_i|^2+\|\boldsymbol{g}_k\|^2}{p_{\mathrm{u}}\|\boldsymbol{g}_k\|^4}\right\}\right)^{-1}\right). \quad (15)$$

Proposition 2: With perfect CSI, Rayleigh fading, and $M \geq 2$, the uplink achievable rate from the kth user for MRC can be lower bounded as follows:

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} = \log_2\left(1+\frac{p_{\mathrm{u}}\left(M-1\right)\beta_k}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}\beta_i+1}\right). \quad (16)$$

Proof: See Appendix A. ■

If $p_{\mathrm{u}} = E_{\mathrm{u}}/M$, and M grows without bound, then

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mrc}} = \log_2\left(1+\frac{\frac{E_{\mathrm{u}}}{M}\left(M-1\right)\beta_k}{\frac{E_{\mathrm{u}}}{M}\sum_{i=1,i\neq k}^{K}\beta_i+1}\right) \rightarrow \log_2(1+\beta_k E_{\mathrm{u}}). \quad (17)$$

Equation (17) shows that the lower bound in (16) becomes equal to the exact limit in Proposition 1 as $M \rightarrow \infty$.

2) Zero-Forcing Receiver: With ZF, $\boldsymbol{A}^H = \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H$, or $\boldsymbol{A}^H\boldsymbol{G} = \boldsymbol{I}_K$. Therefore, $\boldsymbol{a}_k^H\boldsymbol{g}_i = \delta_{ki}$, where $\delta_{ki} = 1$ when $k = i$ and 0 otherwise. From (12), the uplink rate for the kth user is

$$R_{\mathrm{P},k}^{\mathtt{zf}} = \mathbb{E}\left\{\log_2\left(1+\frac{p_{\mathrm{u}}}{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}}\right)\right\}. \quad (18)$$

By using Jensen's inequality, we obtain the following lower bound on the achievable rate:

$$R_{\mathrm{P},k}^{\mathtt{zf}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1+\frac{p_{\mathrm{u}}}{\mathbb{E}\left\{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}\right\}}\right). \quad (19)$$

Proposition 3: When using ZF, in Rayleigh fading, and provided that $M \geq K+1$, the achievable uplink rate for the kth user is lower bounded by

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1+p_{\mathrm{u}}\left(M-K\right)\beta_k\right). \quad (20)$$

Proof: See Appendix B. ■

If $p_{\mathrm{u}} = E_{\mathrm{u}}/M$, and M grows large, we have

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{zf}} = \log_2\left(1+\frac{\beta_k E_{\mathrm{u}}}{M}\left(M-K\right)\right) \rightarrow \log_2(1+\beta_k E_{\mathrm{u}}). \quad (21)$$

We can see again from (21) that the lower bound becomes exact for large M.

3) Minimum Mean-Squared Error Receiver: For MMSE, the detector matrix $\boldsymbol{A}$ is

$$\boldsymbol{A}^H = \left(\boldsymbol{G}^H\boldsymbol{G}+\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K\right)^{-1}\boldsymbol{G}^H = \boldsymbol{G}^H\left(\boldsymbol{G}\boldsymbol{G}^H+\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M\right)^{-1}. \quad (22)$$

Therefore, the kth column of $\boldsymbol{A}$ is given by [18]

$$\boldsymbol{a}_k = \left(\boldsymbol{G}\boldsymbol{G}^H+\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M\right)^{-1}\boldsymbol{g}_k = \frac{\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k}{\boldsymbol{g}_k^H\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k+1} \quad (23)$$

where $\boldsymbol{\Lambda}_k \triangleq \sum_{i=1,i\neq k}^{K}\boldsymbol{g}_i\boldsymbol{g}_i^H + \frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M$. Substituting (23) into (12), we obtain the uplink rate for user k:

$$\begin{aligned} R_{\mathrm{P},k}^{\mathtt{mmse}} &= \mathbb{E}\left\{\log_2\left(1+\boldsymbol{g}_k^H\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k\right)\right\} \\ &\overset{(a)}{=} \mathbb{E}\left\{\log_2\left(\frac{1}{1-\boldsymbol{g}_k^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M+\boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{g}_k}\right)\right\} \\ &= \mathbb{E}\left\{\log_2\left(\frac{1}{1-\left[\boldsymbol{G}^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M+\boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{G}\right]_{kk}}\right)\right\} \\ &\overset{(b)}{=} \mathbb{E}\left\{\log_2\left(\frac{1}{\left[\left(\boldsymbol{I}_K+p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}}\right)\right\} \end{aligned} \quad (24)$$

where (a) is obtained directly from (23), and (b) is obtained by using the identity

$$\begin{aligned} \boldsymbol{G}^H\left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_M+\boldsymbol{G}\boldsymbol{G}^H\right)^{-1}\boldsymbol{G} &= \left(\frac{1}{p_{\mathrm{u}}}\boldsymbol{I}_K+\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{G} \\ &= \boldsymbol{I}_K-\left(\boldsymbol{I}_K+p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}. \end{aligned}$$

By using Jensen's inequality, we obtain the following lower bound on the achievable uplink rate:

$$R_{\mathrm{P},k}^{\mathtt{mmse}} \geq \tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1+\frac{1}{\mathbb{E}\left\{1/\gamma_k\right\}}\right) \quad (25)$$

where $\gamma_k \triangleq \frac{1}{\left[\left(\boldsymbol{I}_K+p_{\mathrm{u}}\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}}-1$. For Rayleigh fading, the exact distribution of γ_k can be found in [19]. This distribution is analytically intractable. To proceed, we approximate it with a distribution which has an analytically tractable form. More specifically, the PDF of γ_k can be approximated by a Gamma distribution as follows [20]:

$$p_{\gamma_k}\left(\gamma\right) = \frac{\gamma^{\alpha_k-1}e^{-\gamma/\theta_k}}{\Gamma\left(\alpha_k\right)\theta_k^{\alpha_k}} \quad (26)$$

where

$$\begin{aligned} \alpha_k &= \frac{\left(M-K+1+\left(K-1\right)\mu\right)^2}{M-K+1+\left(K-1\right)\kappa}, \\ \theta_k &= \frac{M-K+1+\left(K-1\right)\kappa}{M-K+1+\left(K-1\right)\mu}p_{\mathrm{u}}\beta_k \end{aligned} \quad (27)$$

where μ and κ are determined by solving following equations:

$$\begin{aligned} \mu &= \frac{1}{K-1}\sum_{i=1,i\neq k}^{K}\frac{1}{Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1} \\ \kappa&\left(1+\sum_{i=1,i\neq k}^{K}\frac{p_{\mathrm{u}}\beta_i}{\left(Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1\right)^2}\right) \\ &= \sum_{i=1,i\neq k}^{K}\frac{p_{\mathrm{u}}\beta_i\mu+1/(K-1)}{\left(Mp_{\mathrm{u}}\beta_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\mu\right)+1\right)^2}. \end{aligned} \quad (28)$$

Using the approximate PDF of γ_k given by (26), we have the following proposition.

Proposition 4: With perfect CSI, Rayleigh fading, and MMSE, the lower bound on the achievable rate for the kth user can be approximated as

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1 + \left(\alpha_k - 1\right)\theta_k\right). \tag{29}$$

Proof: Substituting (26) into (25), and using the identity [21, eq. (3.326.2)], we obtain

$$\tilde{R}_{\mathrm{P},k}^{\mathtt{mmse}} = \log_2\left(1 + \frac{\Gamma\left(\alpha_k\right)}{\Gamma\left(\alpha_k - 1\right)}\theta_k\right) \tag{30}$$

where $\Gamma\left(\cdot\right)$ is the Gamma function. Then, using $\Gamma\left(x+1\right) = x\Gamma\left(x\right)$, we obtain the desired result (29). ■

Remark 1: From (12), the achievable rate $R_{\mathrm{P},k}$ can be rewritten as

$$\begin{aligned} R_{\mathrm{P},k} &= \mathbb{E}\left\{\log_2\left(1 + \frac{|\boldsymbol{a}_k^H \boldsymbol{g}_k|^2}{\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k \boldsymbol{a}_k}\right)\right\} \\ &\leq \mathbb{E}\left\{\log_2\left(1 + \frac{\|\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k^{1/2}\|^2 \|\boldsymbol{\Lambda}_k^{-1/2}\boldsymbol{g}_k\|^2}{\boldsymbol{a}_k^H \boldsymbol{\Lambda}_k \boldsymbol{a}_k}\right)\right\} \\ &= \mathbb{E}\left\{\log_2\left(1 + \boldsymbol{g}_k^H \boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k\right)\right\}. \end{aligned} \tag{31}$$

The inequality is obtained by using Cauchy-Schwarz' inequality, which holds with equality when $\boldsymbol{a}_k = c\boldsymbol{\Lambda}_k^{-1}\boldsymbol{g}_k$, for any $c \in \mathbb{C}$. This corresponds to the MMSE detector (see (23)). This implies that the MMSE detector is optimal in the sense that it maximizes the achievable rate given by (12).

B. Imperfect Channel State Information

In practice, the channel matrix $\boldsymbol{G}$ has to be estimated at the BS. The standard way of doing this is to use uplink pilots. A part of the coherence interval of the channel is then used for the uplink training. Let T be the length (time-bandwidth product) of the coherence interval and let τ be the number of symbols used for pilots. During the training part of the coherence interval, all users simultaneously transmit mutually orthogonal pilot sequences of length τ symbols. The pilot sequences used by the K users can be represented by a $\tau \times K$ matrix $\sqrt{p_{\mathrm{p}}}\boldsymbol{\Phi}$ ($\tau \geq K$), which satisfies $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$, where $p_{\mathrm{p}} \triangleq \tau p_{\mathrm{u}}$. Then, the $M \times \tau$ received pilot matrix at the BS is given by

$$\boldsymbol{Y}_{\mathrm{p}} = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}\boldsymbol{\Phi}^T + \boldsymbol{N} \tag{32}$$

where $\boldsymbol{N}$ is an $M \times \tau$ matrix with i.i.d. $\mathcal{CN}\left(0,1\right)$ elements. The MMSE estimate of $\boldsymbol{G}$ given $\boldsymbol{Y}$ is

$$\hat{\boldsymbol{G}} = \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{Y}_{\mathrm{p}}\boldsymbol{\Phi}^*\tilde{\boldsymbol{D}} = \left(\boldsymbol{G} + \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{W}\right)\tilde{\boldsymbol{D}} \tag{33}$$

where $\boldsymbol{W} \triangleq \boldsymbol{N}\boldsymbol{\Phi}^*$, and $\tilde{\boldsymbol{D}} \triangleq \left(\frac{1}{p_{\mathrm{p}}}\boldsymbol{D}^{-1} + \boldsymbol{I}_K\right)^{-1}$. Since $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$, $\boldsymbol{W}$ has i.i.d. $\mathcal{CN}\left(0,1\right)$ elements. Note that our analysis takes into account the fact that pilot signals cannot take advantage of the large number of receive antennas since channel estimation has to be done on a per-receive antenna basis. All results that we present take this fact into account. Denote by $\boldsymbol{\mathcal{E}} \triangleq \hat{\boldsymbol{G}} - \boldsymbol{G}$. Then, from (33), the elements of the ith column of $\boldsymbol{\mathcal{E}}$ are RVs with zero means and variances $\frac{\beta_i}{p_{\mathrm{p}}\beta_i + 1}$. Furthermore, owing to the properties of MMSE estimation, $\boldsymbol{\mathcal{E}}$ is independent of $\hat{\boldsymbol{G}}$. The received vector at the BS can be rewritten as

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{A}}^H\left(\sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{G}}\boldsymbol{x} - \sqrt{p_{\mathrm{u}}}\boldsymbol{\mathcal{E}}\boldsymbol{x} + \boldsymbol{n}\right). \tag{34}$$

Therefore, after using the linear detector, the received signal associated with the kth user is

$$\begin{aligned} \hat{r}_k = \sqrt{p_{\mathrm{u}}}\hat{\boldsymbol{a}}_k^H\hat{\boldsymbol{g}}_k x_k + \sqrt{p_{\mathrm{u}}}\sum_{i=1, i\neq k}^{K}\hat{\boldsymbol{a}}_k^H\hat{\boldsymbol{g}}_i x_i \\ - \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{K}\hat{\boldsymbol{a}}_k^H\boldsymbol{\varepsilon}_i x_i + \hat{\boldsymbol{a}}_k^H\boldsymbol{n} \end{aligned} \tag{35}$$

where $\hat{\boldsymbol{a}}_k$, $\hat{\boldsymbol{g}}_i$, and $\boldsymbol{\varepsilon}_i$ are the ith columns of $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{G}}$, and $\boldsymbol{\mathcal{E}}$, respectively.

Since $\hat{\boldsymbol{G}}$ and $\boldsymbol{\mathcal{E}}$ are independent, $\hat{\boldsymbol{A}}$ and $\boldsymbol{\mathcal{E}}$ are independent too. The BS treats the channel estimate as the true channel, and the part including the last three terms of (35) is considered as interference and noise. Therefore, an achievable rate of the uplink transmission from the kth user is given by (36) shown at the bottom of the page.

Intuitively, if we cut the transmitted power of each user, both the data signal and the pilot signal suffer from the reduction in power. Since these signals are multiplied together at the receiver, we expect that there will be a "squaring effect". As a consequence, we cannot reduce power proportionally to $1/M$ as in the case of perfect CSI. The following proposition shows that it is possible to reduce the power (only) proportionally to $1/\sqrt{M}$.

Proposition 5: Assume that the BS has imperfect CSI, obtained by MMSE estimation from uplink pilots, and that the transmit power of each user is $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{\sqrt{M}}$, where E_{u} is fixed. Then,

$$R_{\mathrm{IP},k} \to \log_2\left(1 + \tau\beta_k^2 E_{\mathrm{u}}^2\right), M \to \infty. \tag{37}$$

Proof: For MRC, substituting $\hat{\boldsymbol{a}}_k = \hat{\boldsymbol{g}}_k$ into (36), we obtain the achievable uplink rate as

$$R_{\mathrm{IP},k}^{\mathtt{mrc}} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}\|\hat{\boldsymbol{g}}_k\|^4}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\hat{\boldsymbol{g}}_k^H\hat{\boldsymbol{g}}_i|^2 + p_{\mathrm{u}}\|\hat{\boldsymbol{g}}_k\|^2\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i + 1} + \|\hat{\boldsymbol{g}}_k\|^2}\right)\right\}. \tag{38}$$

Substituting $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$ into (38), and again using (4) along with the fact that each element of $\hat{\boldsymbol{g}}_k$ is a RV with zero

$$R_{\mathrm{IP},k} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}|\hat{\boldsymbol{a}}_k^H\hat{\boldsymbol{g}}_k|^2}{p_{\mathrm{u}}\sum_{i=1,i\neq k}^{K}|\hat{\boldsymbol{a}}_k^H\hat{\boldsymbol{g}}_i|^2 + p_{\mathrm{u}}\|\hat{\boldsymbol{a}}_k\|^2\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i + 1} + \|\hat{\boldsymbol{a}}_k\|^2}\right)\right\} \tag{36}$$

mean and variance $\frac{p_{\rm p}\beta_k^2}{p_{\rm p}\beta_k+1}$, we obtain (37). We can obtain the limit in (37) for ZF and MMSE in a similar way. ■

Proposition 5 implies that with imperfect CSI and a large M, the performance of a MU-MIMO system with an M-antenna array at the BS and with the transmit power per user set to $E_{\rm u}/\sqrt{M}$ is equal to the performance of an interference-free SISO link with transmit power $\tau\beta_k E_{\rm u}^2$, without fast fading.

Remark 2: From the proof of Proposition 5, we see that if we cut the transmit power proportionally to $1/M^{\alpha}$, where $\alpha > 1/2$, then the SINR of the uplink transmission from the kth user will go to zero as $M \to \infty$. This means that $1/\sqrt{M}$ is the fastest rate at which we can cut the transmit power of each user and still maintain a fixed rate.

Remark 3: In general, each user can use different transmit powers which depend on the geometric attenuation and the shadow fading. This can be done by assuming that the kth user knows β_k and performs power control. In this case, the reasoning leading to Proposition 5 can be extended to show that to achieve the same rate as in a SISO system using transmit power $E_{\rm u}$, we must choose the transmit power of the kth user to be $\sqrt{\frac{E_{\rm u}}{M\tau\beta_k}}$.

Remark 4: It can be seen directly from (14) and (38) that the power-scaling laws still hold even for the most unfavorable propagation case (where $\boldsymbol{H}$ has rank one). However, for this case, the multiplexing gains do not materialize since the intracell interference cannot be cancelled when M grows without bound.

1) Maximum-Ratio Combining: By following a similar line of reasoning as in the case of perfect CSI, we can obtain lower bounds on the achievable rate.

Proposition 6: With imperfect CSI, Rayleigh fading, MRC processing, and for $M \geq 2$, the achievable uplink rate for the kth user is lower bounded by

$$\tilde{R}_{{\rm IP},k}^{\tt mrc} = \log_2\left(1+\frac{\tau p_{\rm u}\left(M-1\right)\beta_k^2}{\left(\tau p_{\rm u}\beta_k+1\right)\sum\limits_{i=1,i\neq k}^{K}\beta_i+\left(\tau+1\right)\beta_k+\frac{1}{p_{\rm u}}}\right). \tag{39}$$

By choosing $p_{\rm u} = E_{\rm u}/\sqrt{M}$, we obtain

$$\tilde{R}_{{\rm IP},k}^{\tt mrc} \to \log_2\left(1+\tau\beta_k^2 E_{\rm u}^2\right),\ M\to\infty. \tag{40}$$

Again, when $M \to \infty$, the asymptotic bound on the rate equals the exact limit obtained from Proposition 5.

2) ZF Receiver: For the ZF receiver, we have $\hat{\boldsymbol{a}}_k^H\hat{\boldsymbol{g}}_i = \delta_{ki}$. From (36), we obtain the achievable uplink rate for the kth user as

$$R_{{\rm IP},k}^{\tt zf} = \mathbb{E}\left\{\log_2\left(1+\frac{p_{\rm u}}{\left(\sum\limits_{i=1}^{K}\frac{p_{\rm u}\beta_i}{\tau p_{\rm u}\beta_i+1}+1\right)\left[\left(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}\right)^{-1}\right]_{kk}}\right)\right\}. \tag{41}$$

Following the same derivations as in Section III-A2 for the case of perfect CSI, we obtain the following lower bound on the achievable uplink rate.

Proposition 7: With ZF processing using imperfect CSI, Rayleigh fading, and for $M \geq K+1$, the achievable uplink rate for the kth user is bounded as

$$\tilde{R}_{{\rm IP},k}^{\tt zf} = \log_2\left(1+\frac{\tau p_{\rm u}^2\left(M-K\right)\beta_k^2}{\left(\tau p_{\rm u}\beta_k+1\right)\sum\limits_{i=1}^{K}\frac{p_{\rm u}\beta_i}{\tau p_{\rm u}\beta_i+1}+\tau p_{\rm u}\beta_k+1}\right). \tag{42}$$

Similarly, with $p_{\rm u} = E_{\rm u}/\sqrt{M}$, when $M \to \infty$, the achievable uplink rate and its lower bound tend to the ones for MRC (see (40)), i.e.,

$$\tilde{R}_{{\rm IP},k}^{\tt zf} \to \log_2\left(1+\tau\beta_k^2 E_{\rm u}^2\right),\ M\to\infty \tag{43}$$

which equals the rate value obtained from Proposition 5.

3) MMSE Receiver: With imperfect CSI, the received vector at the BS can be rewritten as

$$\boldsymbol{y} = \sqrt{p_{\rm u}}\hat{\boldsymbol{G}}\boldsymbol{x} - \sqrt{p_{\rm u}}\boldsymbol{\mathcal{E}}\boldsymbol{x} + \boldsymbol{n}. \tag{44}$$

Therefore, for the MMSE receiver, the kth column of $\hat{\boldsymbol{A}}$ is given by

$$\begin{aligned}\hat{\boldsymbol{a}}_k &= \left(\hat{\boldsymbol{G}}\hat{\boldsymbol{G}}^H + \frac{1}{p_{\rm u}}{\tt Cov}\left(-\sqrt{p_{\rm u}}\boldsymbol{\mathcal{E}}\boldsymbol{x}+\boldsymbol{n}\right)\right)^{-1}\hat{\boldsymbol{g}}_k \\ &= \frac{\hat{\boldsymbol{\Lambda}}_k^{-1}\hat{\boldsymbol{g}}_k}{\hat{\boldsymbol{g}}_k^H\hat{\boldsymbol{\Lambda}}_k^{-1}\hat{\boldsymbol{g}}_k+1}\end{aligned} \tag{45}$$

where ${\tt Cov}\left(\boldsymbol{a}\right)$ denotes the covariance matrix of a random vector $\boldsymbol{a}$, and

$$\hat{\boldsymbol{\Lambda}}_k \triangleq \sum_{i=1,i\neq k}^{K}\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^H + \left(\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\rm u}\beta_i+1}+\frac{1}{p_{\rm u}}\right)\boldsymbol{I}_M. \tag{46}$$

Similarly to in Remark 1, by using Cauchy-Schwarz' inequality, we can show that the MMSE receiver given by (45) is the optimal detector in the sense that it maximizes the rate given by (36).

Substituting (45) into (36), we get the achievable uplink rate for the kth user with MMSE receivers as

$$\begin{aligned}R_{{\rm P},k}^{\tt mmse} &= \mathbb{E}\left\{\log_2\left(1+\hat{\boldsymbol{g}}_k^H\hat{\boldsymbol{\Lambda}}_k^{-1}\hat{\boldsymbol{g}}_k\right)\right\} \\ &= -\mathbb{E}\left\{\log_2\left(\left[\left(\boldsymbol{I}_K+\left(\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\rm u}\beta_i+1}+\frac{1}{p_{\rm u}}\right)\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}}\right)^{-1}\right]_{kk}\right)\right\}.\end{aligned} \tag{47}$$

Again, using an approximate distribution for the SINR, we can obtain a lower bound on the achievable uplink rate in closed form.

Proposition 8: With imperfect CSI and Rayleigh fading, the achievable rate for the kth user with MMSE processing is approximately lower bounded as follows:

$$\tilde{R}_{{\rm IP},k}^{\tt mmse} = \log_2\left(1+\left(\hat{\alpha}_k-1\right)\hat{\theta}_k\right) \tag{48}$$

where

$$\hat{\alpha}_k = \frac{\left(M-K+1+(K-1)\,\hat{\mu}\right)^2}{M-K+1+(K-1)\,\hat{\kappa}},$$
$$\hat{\theta}_k = \frac{M-K+1+(K-1)\,\hat{\kappa}}{M-K+1+(K-1)\,\hat{\mu}}\omega\hat{\beta}_k \tag{49}$$

where $\omega \triangleq \left(\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}+\frac{1}{p_{\mathrm{u}}}\right)^{-1}$, $\hat{\beta}_k \triangleq \frac{\tau p_{\mathrm{u}}\beta_k^2}{\tau p_{\mathrm{u}}\beta_k+1}$, $\hat{\mu}$ and $\hat{\kappa}$ are obtained by using following equations:

$$\hat{\mu} = \frac{1}{K-1}\sum_{i=1,i\neq k}^{K}\frac{1}{M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1}$$
$$\hat{\kappa}\left(1+\sum_{i=1,i\neq k}^{K}\frac{\omega\hat{\beta}_i}{\left(M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1\right)^2}\right)$$
$$=\sum_{i=1,i\neq k}^{K}\frac{\omega\hat{\beta}_i\hat{\mu}+1/(K-1)}{\left(M\omega\hat{\beta}_i\left(1-\frac{K-1}{M}+\frac{K-1}{M}\hat{\mu}\right)+1\right)^2}. \tag{50}$$

Table I summarizes the lower bounds on the achievable rates for linear receivers derived in this section, distinguishing between the cases of perfect and imperfect CSI, respectively. Here $C(x) \triangleq \log_2(1+x)$.

We have considered a *single-cell* MU-MIMO system. This simplifies the analysis, and it gives us important insights into how power can be scaled with the number of antennas in very large MIMO systems. A natural question is to what extent this power-scaling law still holds for *multicell* MU-MIMO systems. Intuitively, when we reduce the transmit power of each user, the effect of interference from other cells also reduces and hence, the SINR will stay unchanged. Therefore we will have the same power-scaling law as in the single-cell scenario. The next section explains this argument in more detail.

C. Power-Scaling Law for Multicell MU-MIMO Systems

We will use the MRC for our analysis. A similar analysis can be performed for the ZF and MMSE detectors. Consider the uplink of a multicell MU-MIMO system with L cells sharing the same frequency band. Each cell includes one BS equipped with M antennas and K single-antenna users. The $M\times 1$ received vector at the lth BS is given by

$$\boldsymbol{y}_l = \sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{G}_{li}\boldsymbol{x}_i + \boldsymbol{n}_l \tag{51}$$

TABLE I
LOWER BOUNDS ON THE ACHIEVABLE RATES OF THE UPLINK TRANSMISSION FOR THE kTH USER.

	Perfect CSI	Imperfect CSI
MRC	$C\left(\frac{p_{\mathrm{u}}(M-1)\beta_k}{p_{\mathrm{u}}\sum_{i\neq k}^{K}\beta_i+1}\right)$	$C\left(\frac{\tau p_{\mathrm{u}}(M-1)\beta_k^2}{(\tau p_{\mathrm{u}}\beta_k+1)\sum_{i\neq k}^{K}\beta_i+(\tau+1)\beta_k+\frac{1}{p_{\mathrm{u}}}}\right)$
ZF	$C\left(p_{\mathrm{u}}(M-K)\beta_k\right)$	$C\left(\frac{\tau p_{\mathrm{u}}(M-K)\beta_k^2}{(\tau p_{\mathrm{u}}\beta_k+1)\sum_{i=1}^{K}\frac{\beta_i}{\tau p_{\mathrm{u}}\beta_i+1}+\tau\beta_k+\frac{1}{p_{\mathrm{u}}}}\right)$
MMSE	$C\left((\alpha_k-1)\theta_k\right)$	$C\left((\hat{\alpha}_k-1)\hat{\theta}_k\right)$

where $\sqrt{p_{\mathrm{u}}}\boldsymbol{x}_i$ is the $K\times 1$ transmitted vector of K users in the ith cell; $\boldsymbol{n}_l$ is an AWGN vector, $\boldsymbol{n}_l \sim \mathcal{CN}(0,\boldsymbol{I}_M)$; and $\boldsymbol{G}_{li}$ is the $M\times K$ channel matrix between the lth BS and the K users in the ith cell. The channel matrix $\boldsymbol{G}_{li}$ can be represented as

$$\boldsymbol{G}_{li} = \boldsymbol{H}_{li}\boldsymbol{D}_{li}^{1/2} \tag{52}$$

where $\boldsymbol{H}_{li}$ is the fast fading matrix between the lth BS and the K users in the ith cell whose elements have zero mean and unit variance; and $\boldsymbol{D}_{li}$ is a $K\times K$ diagonal matrix, where $[\boldsymbol{D}_{li}]_{kk} = \beta_{lik}$, with β_{lik} represents the large-scale fading between the kth user in the i cell and the lth BS.

1) Perfect CSI: With perfect CSI, the received signal at the lth BS after using MRC is given by

$$\boldsymbol{r}_l = \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{ll}^H\boldsymbol{G}_{ll}\boldsymbol{x}_l + \sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq l}^{L}\boldsymbol{G}_{ll}^H\boldsymbol{G}_{li}\boldsymbol{x}_i + \boldsymbol{G}_{ll}^H\boldsymbol{n}_l. \tag{53}$$

With $p_{\mathrm{u}} = \frac{E_{\mathrm{u}}}{M}$, (53) can be rewritten as

$$\frac{1}{\sqrt{M}}\boldsymbol{r}_l = \sqrt{E_{\mathrm{u}}}\frac{\boldsymbol{G}_{ll}^H\boldsymbol{G}_{ll}}{M}\boldsymbol{x}_l + \sqrt{p_{\mathrm{u}}}\sum_{i=1,i\neq l}^{L}\frac{\boldsymbol{G}_{ll}^H\boldsymbol{G}_{li}}{M}\boldsymbol{x}_i + \frac{1}{\sqrt{M}}\boldsymbol{G}_{ll}^H\boldsymbol{n}_l. \tag{54}$$

From (4)–(5), when M grows large, the interference from other cells disappears. More precisely,

$$\frac{1}{\sqrt{M}}\boldsymbol{r}_l \to \sqrt{E_{\mathrm{u}}}\boldsymbol{D}_{ll}\boldsymbol{x}_l + \boldsymbol{D}_{ll}^{1/2}\tilde{\boldsymbol{n}}_l \tag{55}$$

where $\tilde{\boldsymbol{n}}_l \sim \mathcal{CN}(0,\boldsymbol{I})$. Therefore, the SINR of the uplink transmission from the kth user in the lth cell converges to a constant value when M grows large, more precisely

$$\mathsf{SINR}_{l,k}^{\mathrm{P}} \to \beta_{llk}E_{\mathrm{u}}, \text{ as } M\to\infty. \tag{56}$$

This means that the power scaling law derived for single-cell systems is valid in multicell systems too.

2) Imperfect CSI: In this case, the channel estimate from the uplink pilots is contaminated by interference from other cells. The MMSE channel estimate of the channel matrix $\boldsymbol{G}_{ll}$ is given by [11]

$$\hat{\boldsymbol{G}}_{ll} = \left(\sum_{i=1}^{L}\boldsymbol{G}_{li} + \frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{W}_l\right)\tilde{\boldsymbol{D}}_{ll} \tag{57}$$

where $\tilde{\boldsymbol{D}}_{ll}$ is a diagonal matrix where the kth diagonal element $\left[\tilde{\boldsymbol{D}}_{ll}\right]_{kk} = \beta_{llk}\left(\sum_{i=1}^{L}\beta_{lik}+\frac{1}{p_{\mathrm{p}}}\right)^{-1}$. The received signal at the lth BS after using MRC is given by

$$\hat{\boldsymbol{r}}_l = \hat{\boldsymbol{G}}_{ll}^H\boldsymbol{y}_l$$
$$= \tilde{\boldsymbol{D}}_{ll}\left(\sum_{i=1}^{L}\boldsymbol{G}_{li}+\frac{1}{\sqrt{p_{\mathrm{p}}}}\boldsymbol{W}_l\right)^H\left(\sqrt{p_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{G}_{li}\boldsymbol{x}_i+\boldsymbol{n}_l\right). \tag{58}$$

With $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$, we have

$$\frac{1}{M^{3/4}}\tilde{\boldsymbol{D}}_{ll}^{-1}\hat{\boldsymbol{r}}_l = \sqrt{E_{\mathrm{u}}}\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\boldsymbol{G}_{li}^H\boldsymbol{G}_{lj}}{M}\boldsymbol{x}_j + \sum_{i=1}^{L}\frac{\boldsymbol{G}_{li}^H\boldsymbol{n}_l}{M^{3/4}}$$
$$+\frac{1}{\sqrt{\tau}}\sum_{i=1}^{L}\frac{\boldsymbol{W}_l^H\boldsymbol{G}_{li}}{M^{3/4}}\boldsymbol{x}_i + \frac{1}{\sqrt{\tau E_{\mathrm{u}}}}\frac{\boldsymbol{W}_l^H\boldsymbol{n}_l}{M^{1/2}}. \tag{59}$$

By using (4) and (5), as M grows large, we obtain

$$\frac{1}{M^{3/4}}\tilde{\boldsymbol{D}}_{ll}^{-1}\hat{\boldsymbol{r}}_l \to \sqrt{E_{\mathrm{u}}}\sum_{i=1}^{L}\boldsymbol{D}_{li}\boldsymbol{x}_i + \frac{1}{\sqrt{\tau E_{\mathrm{u}}}}\tilde{\boldsymbol{w}}_l \qquad (60)$$

where $\tilde{\boldsymbol{w}}_l \sim \mathcal{CN}(0, \boldsymbol{I}_M)$. Therefore, the asymptotic SINR of the uplink from the kth user in the lth cell is

$$\mathsf{SINR}_{l,k}^{\mathrm{IP}} \to \frac{\tau\beta_{llk}^2 E_{\mathrm{u}}^2}{\tau\sum_{i\neq l}^{L}\beta_{lik}^2 E_{\mathrm{u}}^2 + 1}, \text{ as } M \to \infty. \qquad (61)$$

We can see that the $1/\sqrt{M}$ power-scaling law still holds. Furthermore, transmission from users in other cells constitutes residual interference. The reason is that the pilot reuse gives pilot-contamination-induced inter-cell interference which grows with M at the same rate as the desired signal.

Remark 5: The MMSE channel estimate (57) is obtained by the assumption that, for uplink training, all cells simultaneously transmit pilot sequences, and that the same set of pilot sequences is used in all cells. This assumption makes no fundamental difference compared with using different pilot sequences in different cells, as explained [8, Section VII-F]. Nor does this assumption make any fundamental difference to the case when users in other cells transmit *data* when the users in the cell of interest send their pilots. The reason is that whatever data is transmitted in other cells, it can always be expanded in terms of the orthogonal pilot sequences that are transmitted in the cell of interest, so pilot contamination ensues. For example, consider the uplink training in cell 1 of a MU-MIMO system with $L = 2$ cells. Assume that, during an interval of length τ symbols ($\tau \geq K$), K users in cell 1 are transmitting uplink pilots $\boldsymbol{\Phi}^T$ at the same time as K users in cell 2 are transmitting uplink data $\boldsymbol{X}_2$. Here $\boldsymbol{\Phi}$ is a $\tau \times K$ matrix which satisfies $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$. The received signal at base station 1 is

$$\boldsymbol{Y}_1 = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}_{11}\boldsymbol{\Phi}^T + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\boldsymbol{X}_2 + \boldsymbol{N}_1$$

where $\boldsymbol{N}_1 \in \mathbb{C}^{M\times\tau}$ is AWGN at base station 1. By projecting the received signal $\boldsymbol{Y}_1$ onto $\boldsymbol{\Phi}^*$, we obtain

$$\tilde{\boldsymbol{Y}}_1 \triangleq \boldsymbol{Y}_1\boldsymbol{\Phi}^* = \sqrt{p_{\mathrm{p}}}\boldsymbol{G}_{11} + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{X}}_2 + \tilde{\boldsymbol{N}}_1$$

where $\tilde{\boldsymbol{X}}_2 \triangleq \boldsymbol{X}_2\boldsymbol{\Phi}^*$, and $\tilde{\boldsymbol{N}}_1 \triangleq \boldsymbol{N}_1\boldsymbol{\Phi}^*$. The kth column of $\tilde{\boldsymbol{Y}}_1$ is given by

$$\tilde{\boldsymbol{y}}_{1k} = \sqrt{p_{\mathrm{p}}}\boldsymbol{g}_{11k} + \sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{x}}_{2k} + \tilde{\boldsymbol{n}}_{1k}$$

where $\boldsymbol{g}_{11k}$, $\tilde{\boldsymbol{x}}_{2k}$, and $\tilde{\boldsymbol{n}}_{1k}$ are the kth columns of $\boldsymbol{G}_{11}$, $\tilde{\boldsymbol{X}}_2$, and $\tilde{\boldsymbol{N}}_1$, respectively. By using the Lindeberg-Lévy central limit theorem, we find that each element of the vector $\sqrt{p_{\mathrm{u}}}\boldsymbol{G}_{12}\tilde{\boldsymbol{x}}_{2,k}$ (ignoring the large-scale fading in this argument) is approximately Gaussian distributed with zero mean and variance Kp_{u}. If $K = \tau$, then $Kp_{\mathrm{u}} = p_{\mathrm{p}}$ and this result means that the effect of payload interference is just as bad as if users in cell 2 transmitted pilot sequences.

IV. Energy-Efficiency versus Spectral-Efficiency Tradeoff

The energy-efficiency (in bits/Joule) of a system is defined as the spectral-efficiency (sum-rate in bits/channel use) divided by the transmit power expended (in Joules/channel use). Typically, increasing the spectral efficiency is associated with increasing the power and hence, with decreasing the energy-efficiency. Therefore, there is a fundamental tradeoff between the energy efficiency and the spectral efficiency. However, in one operating regime it is possible to jointly increase the energy and spectral efficiencies, and in this regime there is no tradeoff. This may appear a bit counterintuitive at first, but it falls out from the analysis in Section IV-A. Note, however, that this effect occurs in an operating regime that is probably of less interest in practice.

In this section, we study the energy-spectral efficiency tradeoff for the uplink of MU-MIMO systems using linear receivers at the BS. Certain activities (multiplexing to many users rather than beamforming to a single user and increasing the number of service antennas) can simultaneously benefit both the spectral-efficiency and the radiated energy-efficiency. Once the number of service antennas is set, one can adjust other system parameters (radiated power, numbers of users, duration of pilot sequences) to obtain increased spectral-efficiency at the cost of reduced energy-efficiency, and vice-versa. This should be a desirable feature for service providers: they can set the operating point according to the current traffic demand (high energy-efficiency and low spectral-efficiency, for example, during periods of low demand).

A. Single-Cell MU-MIMO Systems

We define the spectral efficiency for perfect and imperfect CSI, respectively, as follows

$$R_{\mathrm{P}}^A = \sum_{k=1}^{K}\tilde{R}_{\mathrm{P},k}^A, \text{ and } R_{\mathrm{IP}}^A = \frac{T-\tau}{T}\sum_{k=1}^{K}\tilde{R}_{\mathrm{IP},k}^A \qquad (62)$$

where $A \in \{\mathtt{mrc}, \mathtt{zf}, \mathtt{mmse}\}$ corresponds to MRC, ZF and MMSE, and T is the coherence interval in symbols. The energy-efficiency for perfect and imperfect CSI is defined as

$$\eta_{\mathrm{P}}^A = \frac{1}{p_{\mathrm{u}}}R_{\mathrm{P}}^A, \text{ and } \eta_{\mathrm{IP}}^A = \frac{1}{p_{\mathrm{u}}}R_{\mathrm{IP}}^A. \qquad (63)$$

The large-scale fading can be incorporated by substituting (39) and (42) into (62). However, this yields energy and spectral efficiency formulas of an intractable form and which are very difficult (if not impossible) to use for obtaining further insights. Note that the large number of antennas effectively removes the small-scale fading, but the effects of path loss and large-scale fading will remain. This may give different users vastly different SNRs. As a result, power control may be desired. In principle, a power control factor could be included by letting p_{u} in (39) and (42) depend on k. The optimal transmit power for each user would depend only on the large-scale fading, not on the small-scale fading and effective power-control rules could be developed straightforwardly from the resulting expressions. However, the introduction of such power control may bring new trade-offs, for example that of fairness between users near and far from the BS. In addition, the spectral versus energy efficiency tradeoff relies on optimization of the number of active users. If the users have grossly different large-scale fading coefficients, then the

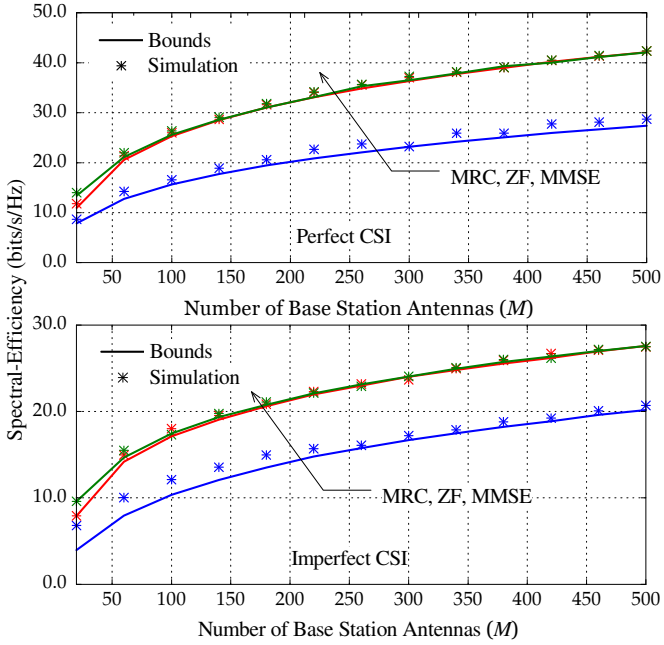

Fig. 1. Lower bounds and numerically evaluated values of the spectral efficiency for different numbers of BS antennas for MRC, ZF, and MMSE with perfect and imperfect CSI. In this example there are $K = 10$ users, the coherence interval $T = 196$, the transmit power per terminal is $p_{\mathrm{u}} = 10$ dB, and the propagation channel parameters were $\sigma_{\mathrm{shadow}} = 8$ dB, and $\nu = 3.8$.

issue will arise as to whether these coefficients should be fixed before the optimization or whether for a given number of users K, these coefficients should be drawn randomly. Both ways can be justified, but have different operational meaning in terms of scheduling. This leads, among others, to issues with fairness versus total throughput, which we would like to avoid here as this matter could easily obscure the main points of our analysis. Therefore, for analytical tractability, we ignore the effect of the large-scale fading here, i.e., we set $\boldsymbol{D} = \boldsymbol{I}_K$. Also, we only consider MRC and ZF receivers.[4]

For perfect CSI, it is straightforward to show from (16), (20), and (63) that when the spectral efficiency increases, the energy efficiency decreases. For imperfect CSI, this is not always so, as we shall see next. In what follows, we focus on the case of imperfect CSI since this is the case of interest in practice.

1) Maximum-Ratio Combining: From (39), the spectral efficiency and energy efficiency with MRC processing are given by

$$R_{\mathrm{IP}}^{\mathtt{mrc}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-1)\,p_{\mathrm{u}}^2}{\tau\,(K-1)\,p_{\mathrm{u}}^2 + (K+\tau)\,p_{\mathrm{u}} + 1}\right),$$
$$\eta_{\mathrm{IP}}^{\mathtt{mrc}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{mrc}}. \tag{64}$$

We have

$$\lim_{p_{\mathrm{u}}\to 0} \eta_{\mathrm{IP}}^{\mathtt{mrc}} = \lim_{p_{\mathrm{u}}\to 0} \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{mrc}}$$
$$= \lim_{p_{\mathrm{u}}\to 0} \frac{T-\tau}{T} K \frac{(\log_2 e)\,\tau\,(M-1)\,p_{\mathrm{u}}}{\tau\,(K-1)\,p_{\mathrm{u}}^2 + (K+\tau)\,p_{\mathrm{u}} + 1} = 0 \tag{65}$$

and

$$\lim_{p_{\mathrm{u}}\to \infty} \eta_{\mathrm{IP}}^{\mathtt{mrc}} = \lim_{p_{\mathrm{u}}\to \infty} \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{mrc}} = 0. \tag{66}$$

Equations (65) and (66) imply that for low p_{u}, the energy efficiency increases when p_{u} increases, and for high p_{u} the energy efficiency decreases when p_{u} increases. Since $\frac{\partial R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}} > 0$, $\forall p_{\mathrm{u}} > 0$, $R_{\mathrm{IP}}^{\mathtt{mrc}}$ is a monotonically increasing function of p_{u}. Therefore, at low p_{u} (and hence at low spectral efficiency), the energy efficiency increases as the spectral efficiency increases and vice versa at high p_{u}. The reason is that, the spectral efficiency suffers from a "squaring effect" when the received data signal is multiplied with the received pilots. Hence, at $p_{\mathrm{u}} \ll 1$, the spectral-efficiency behaves as $\sim p_{\mathrm{u}}^2$. As a consequence, the energy efficiency (which is defined as the spectral efficiency divided by p_{u}) increases linearly with p_{u}. In more detail, expanding the rate in a Taylor series for $p_{\mathrm{u}} \ll 1$, we obtain

$$R_{\mathrm{IP}}^{\mathtt{mrc}} \approx R_{\mathrm{IP}}^{\mathtt{mrc}}|_{p_{\mathrm{u}}=0} + \left.\frac{\partial R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}}\right|_{p_{\mathrm{u}}=0} p_{\mathrm{u}} + \frac{1}{2}\left.\frac{\partial^2 R_{\mathrm{IP}}^{\mathtt{mrc}}}{\partial p_{\mathrm{u}}^2}\right|_{p_{\mathrm{u}}=0} p_{\mathrm{u}}^2$$
$$= \frac{T-\tau}{T} K \log_2(e)\,\tau\,(M-1)\,p_{\mathrm{u}}^2. \tag{67}$$

This gives the following relation between the spectral efficiency and energy efficiency at $p_{\mathrm{u}} \ll 1$:

$$\eta_{\mathrm{IP}}^{\mathtt{mrc}} = \sqrt{\frac{T-\tau}{T} K \log_2(e)\,\tau\,(M-1)\,R_{\mathrm{IP}}^{\mathtt{mrc}}}. \tag{68}$$

We can see that when $p_{\mathrm{u}} \ll 1$, by doubling the spectral efficiency, or by doubling M, we can increase the energy efficiency by 1.5 dB.

2) Zero-Forcing Receiver: From (42), the spectral efficiency and energy efficiency for ZF are given by

$$R_{\mathrm{IP}}^{\mathtt{zf}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-K)\,p_{\mathrm{u}}^2}{(K+\tau)\,p_{\mathrm{u}} + 1}\right), \text{ and}$$
$$\eta_{\mathrm{IP}}^{\mathtt{zf}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{zf}}. \tag{69}$$

[4]When M is large, the performance of the MMSE receiver is very close to that of the ZF receiver (see Section V). Therefore, the insights on energy versus spectral efficiency obtained from studying the performance of ZF can be used to draw conclusions about MMSE as well.

$$R_{\mathrm{mul}}^{\mathtt{mrc}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-1)\,p_{\mathrm{u}}^2}{\tau\left(K\bar{L}^2 - 1 + \beta\left(\bar{L}-1\right)(M-2)\right)p_{\mathrm{u}}^2 + \bar{L}\,(K+\tau)\,p_{\mathrm{u}} + 1}\right), \text{and } \eta_{\mathrm{mul}}^{\mathtt{mrc}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{IP}}^{\mathtt{mrc}} \tag{73}$$

$$R_{\mathrm{mul}}^{\mathtt{zf}} = \frac{T-\tau}{T} K \log_2\left(1 + \frac{\tau\,(M-K)\,p_{\mathrm{u}}^2}{\tau K\left(\bar{L}^2 - \bar{L}\beta + \beta - 1\right)p_{\mathrm{u}}^2 + \bar{L}\,(K+\tau)\,p_{\mathrm{u}} + 1}\right), \text{ and } \eta_{\mathrm{IP}}^{\mathtt{zf}} = \frac{1}{p_{\mathrm{u}}} R_{\mathrm{ml}}^{\mathtt{zf}} \tag{74}$$

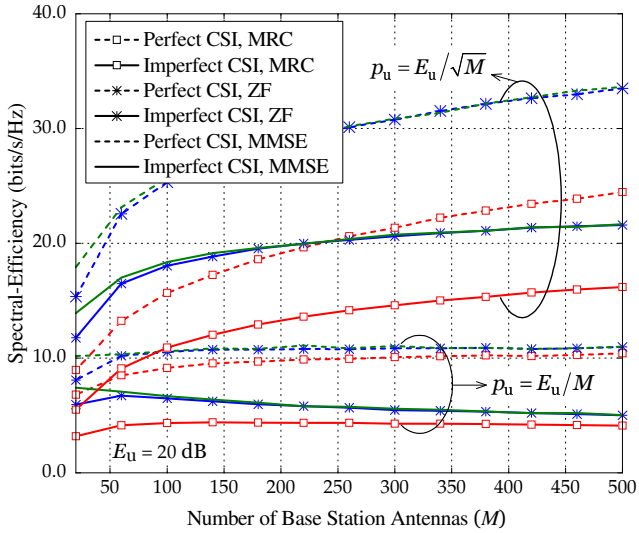

Fig. 2. Spectral efficiency versus the number of BS antennas M for MRC, ZF, and MMSE processing at the receiver, with perfect CSI and with imperfect CSI (obtained from uplink pilots). In this example $K = 10$ users are served simultaneously, the reference transmit power is $E_u = 20$ dB, and the propagation parameters were $\sigma_{\text{shadow}} = 8$ dB and $\nu = 3.8$.

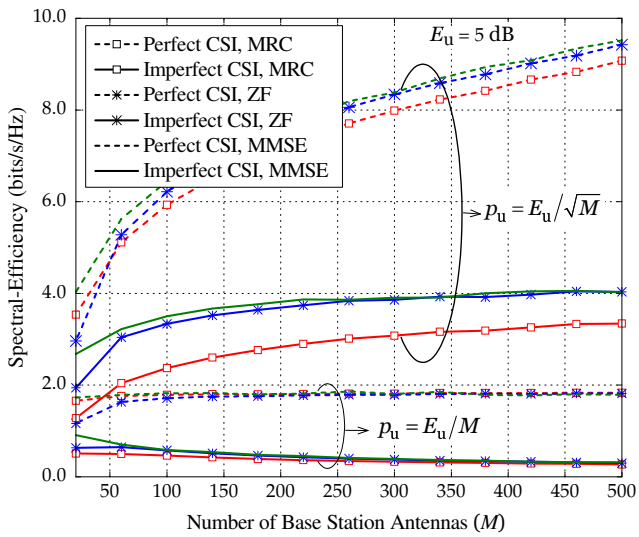

Fig. 3. Same as Figure 2, but with $E_u = 5$ dB.

Similarly to in the analysis of MRC, we can show that at low transmit power p_u, the energy efficiency increases when the spectral efficiency increases. In the low-p_u regime, we obtain the following Taylor series expansion

$$R_{\text{IP}}^{\text{zf}} \approx \frac{T-\tau}{T} K \log_2(e)\, \tau\, (M-K)\, p_u^2, \text{ for } p_u \ll 1. \quad (70)$$

Therefore,

$$\eta_{\text{IP}}^{\text{zf}} = \sqrt{\frac{T-\tau}{T} K \log_2(e)\, \tau\, (M-K)\, R_{\text{IP}}^{\text{zf}}}. \quad (71)$$

Again, at $p_u \ll 1$, by doubling M or $R_{\text{IP}}^{\text{zf}}$, we can increase the energy efficiency by 1.5 dB.

B. Multicell MU-MIMO Systems

In this section, we derive expressions for the energy-efficiency and spectral-efficiency for a multicell system. These are used for the simulation in the Section V. Here, we consider a simplified channel model, i.e., $\boldsymbol{D}_{ll} = \boldsymbol{I}_K$, and $\boldsymbol{D}_{li} = \beta \boldsymbol{I}_K$, where $\beta \in [0,1]$ is an intercell interference factor. Note that from (57), the estimate of the channel between the kth user in the lth cell and the lth BS is given by

$$\hat{\boldsymbol{g}}_{llk} = \left(\bar{L} + \frac{1}{p_p}\right)^{-1} \left(\boldsymbol{h}_{llk} + \sum_{i \neq k}^{L} \sqrt{\beta} \boldsymbol{h}_{lik} + \frac{1}{\sqrt{p_p}} \boldsymbol{w}_{lk}\right). \quad (72)$$

where $\bar{L} \triangleq (L-1)\beta + 1$. The term $\sum_{i\neq k}^{L} \sqrt{\beta} \boldsymbol{h}_{lik}$ represents the pilot contamination, therefore

$$\frac{\sum_{i\neq k}^{L} \mathbb{E}\left\{\|\sqrt{\beta}\boldsymbol{h}_{lik}\|^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{h}_{llk}\|^2\right\}} = \beta(L-1)$$

can be considered as the effect of pilot contamination.

Following a similar derivation as in the case of single-cell MU-MIMO systems, we obtain the spectral efficiency and energy efficiency for imperfect CSI with MRC and ZF receivers, respectively, as (73) and (74) shown at the bottom of the previous page. The principal complexity in the derivation is the correlation between pilot-contaminated channel estimates.

We can see that the spectral efficiency is a decreasing function of β and L. Furthermore, when $L = 1$, or $\beta = 0$, the results (73) and (74) coincide with (64) and (69) for single-cell MU-MIMO systems.

V. Numerical Results

A. Single-Cell MU-MIMO Systems

We consider a hexagonal cell with a radius (from center to vertex) of 1000 meters. The users are located uniformly at random in the cell and we assume that no user is closer to the BS than $r_h = 100$ meters. The large-scale fading is modelled via $\beta_k = z_k/(r_k/r_h)^\nu$, where z_k is a log-normal random variable with standard deviation σ_{shadow}, r_k is the distance between the kth user and the BS, and ν is the path loss exponent. For all examples, we choose $\sigma_{\text{shadow}} = 8$ dB, and $\nu = 3.8$.

We assume that the transmitted data are modulated with OFDM. Here, we choose parameters that resemble those of LTE standard: an OFDM symbol duration of $T_s = 71.4\mu$s, and a useful symbol duration of $T_u = 66.7\mu$s. Therefore, the guard interval length is $T_g = T_s - T_u = 4.7\mu$s. We choose the channel coherence time to be $T_c = 1$ ms. Then, $T = \frac{T_c}{T_s}\frac{T_u}{T_g} = 196$, where $\frac{T_c}{T_s} = 14$ is the number of OFDM symbols in a 1 ms coherence interval, and $\frac{T_u}{T_g} = 14$ corresponds to the "frequency smoothness interval" [8].

1) Power-Scaling Law: We first conduct an experiment to validate the tightness of our proposed capacity bounds. Fig. 1 shows the simulated spectral efficiency and the proposed analytical bounds for MRC, ZF, and MMSE receivers with perfect and imperfect CSI at $p_u = 10$ dB. In this example there are $K = 10$ users. For CSI estimation from uplink pilots, we choose pilot sequences of length $\tau = K$. (This is the smallest amount of training that can be used.) Clearly, all bounds are very tight, especially at large M. Therefore, in the following, we will use these bounds for all numerical work.

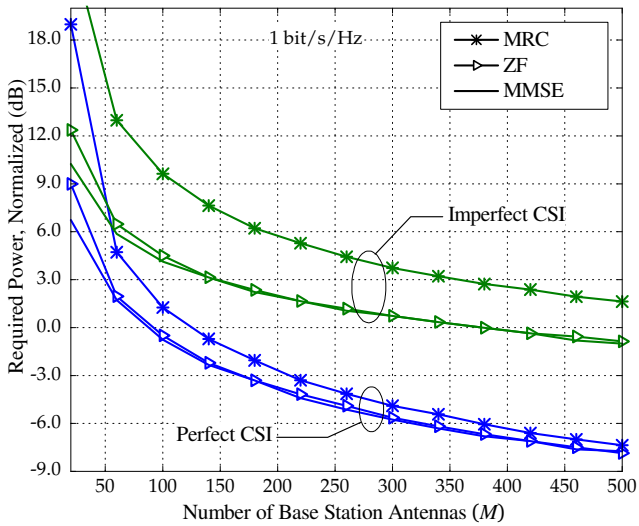

Fig. 4. Transmit power required to achieve 1 bit/channel use per user for MRC, ZF, and MMSE processing, with perfect and imperfect CSI, as a function of the number M of BS antennas. The number of users is fixed to $K = 10$, and the propagation parameters are $\sigma_{\text{shadow}} = 8$ dB and $\nu = 3.8$.

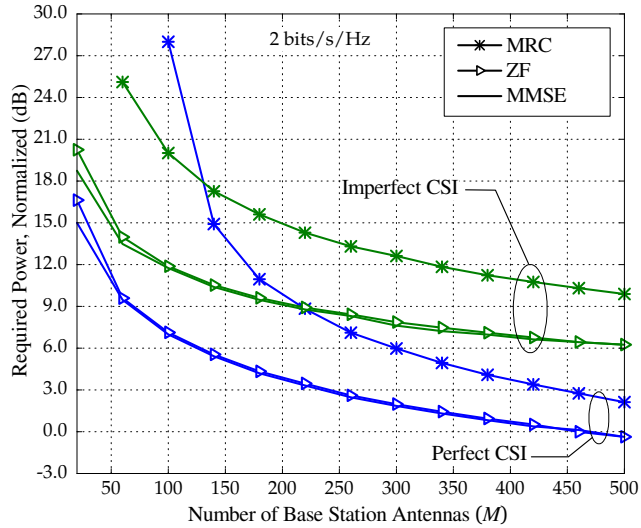

Fig. 5. Same as Figure 4 but for a target spectral efficiency of 2 bits/channel use per user.

We next illustrate the power scaling laws. Fig. 2 shows the spectral efficiency on the uplink versus the number of BS antennas for $p_{\text{u}} = E_{\text{u}}/M$ and $p_{\text{u}} = E_{\text{u}}/\sqrt{M}$ with perfect and imperfect receiver CSI, and with MRC, ZF, and MMSE processing, respectively. Here, we choose $E_{\text{u}} = 20$ dB. At this SNR, the spectral efficiency is in the order of 10–30 bits/s/Hz, corresponding to a spectral efficiency per user of 1–3 bits/s/Hz. These operating points are reasonable from a practical point of view. For example, 64-QAM with a rate-1/2 channel code would correspond to 3 bits/s/Hz. (Figure 3, see below, shows results at lower SNR.) As expected, with $p_{\text{u}} = E_{\text{u}}/M$, when M increases, the spectral efficiency approaches a constant value for the case of perfect CSI, but decreases to 0 for the case of imperfect CSI. However, with $p_{\text{u}} = E_{\text{u}}/\sqrt{M}$, for the case of perfect CSI the spectral efficiency grows without bound (logarithmically fast with M) when $M \to \infty$ and with imperfect CSI, the spectral efficiency converges to a nonzero limit as $M \to \infty$. These results confirm that we can scale down the transmitted power of each user as E_{u}/M for the perfect CSI case, and as $E_{\text{u}}/\sqrt{M}$ for the imperfect CSI case when M is large.

Typically ZF is better than MRC at high SNR, and vice versa at low SNR [13]. MMSE always performs the best across the entire SNR range (see Remark 1). When comparing MRC and ZF in Fig. 2, we see that here, when the transmitted power is proportional to $1/\sqrt{M}$, the power is not low enough to make MRC perform as well as ZF. But when the transmitted power is proportional to $1/M$, MRC performs almost as well as ZF for large M. Furthermore, as we can see from the figure, MMSE is always better than MRC or ZF, and its performance is very close to ZF.

In Fig. 3, we consider the same setting as in Fig. 2, but we choose $E_{\text{u}} = 5$ dB. This figure provides the same insights as Fig. 2. The gap between the performance of MRC and that of ZF (or MMSE) is reduced compared with Fig. 2. This is so because the relative effect of crosstalk interference (the interference from other users) as compared to the thermal noise is smaller here than in Fig. 2.

We next show the transmit power per user that is needed to reach a fixed spectral efficiency. Fig. 4 shows the normalized power (p_{u}) required to achieve 1 bit/s/Hz per user as a function of M. As predicted by the analysis, by doubling M, we can cut back the power by approximately 3 dB and 1.5 dB for the cases of perfect and imperfect CSI, respectively. When M is large ($M/K \gtrapprox 6$), the difference in performance between MRC and ZF (or MMSE) is less than 1 dB and 3 dB for the cases of perfect and imperfect CSI, respectively. This difference increases when we increase the target spectral efficiency. Fig. 5 shows the normalized power required for 2 bit/s/Hz per user. Here, the crosstalk interference is more significant (relative to the thermal noise) and hence the ZF and MMSE receivers perform relatively better.

2) Energy Efficiency versus Spectral Efficiency Tradeoff : We next examine the tradeoff between energy efficiency and spectral efficiency in more detail. Here, we ignore the effect of large-scale fading, i.e., we set $\boldsymbol{D} = \boldsymbol{I}_K$. We normalize the energy efficiency against a reference mode corresponding to a single-antenna BS serving one single-antenna user with $p_{\text{u}} = 10$ dB. For this reference mode, the spectral efficiencies and energy efficiencies for MRC, ZF, and MMSE are equal, and given by (from (38) and (62))

$$R_{\text{IP}}^{0} = \frac{T-\tau}{T}\,\mathbb{E}\left\{\log_2\left(1+\frac{\tau p_{\text{u}}^2|z|^2}{1+p_{\text{u}}\left(1+\tau\right)}\right)\right\}$$
$$\eta_{\text{IP}}^{0} = R_{\text{IP}}^{0}/p_{\text{u}}$$

where z is a Gaussian RV with zero mean and unit variance. For the reference mode, the spectral-efficiency is obtained by choosing the duration of the uplink pilot sequence τ to maximize R_{IP}^{0}. Numerically we find that $R_{\text{IP}}^{0} = 2.65$ bits/s/Hz and $\eta_{\text{IP}}^{0} = 0.265$ bits/J.

Fig. 6 shows the relative energy efficiency versus the the spectral efficiency for MRC and ZF. The relative energy efficiency is obtained by normalizing the energy efficiency by

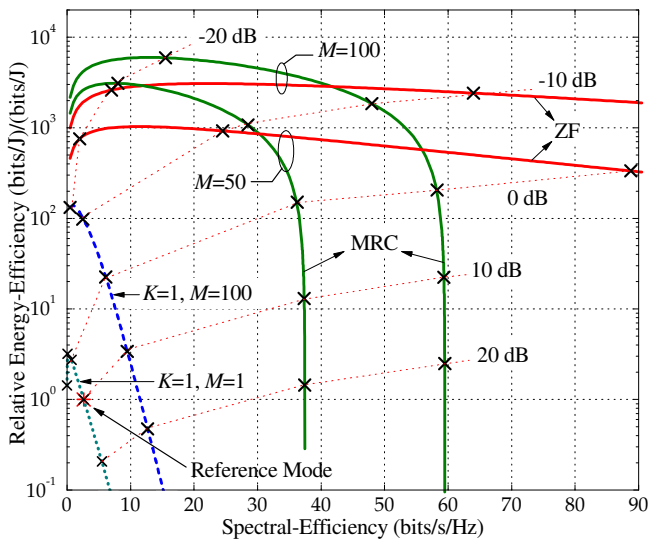

Fig. 6. Energy efficiency (normalized with respect to the reference mode) versus spectral efficiency for MRC and ZF with imperfect CSI. The reference mode corresponds to $K = 1, M = 1$ (single antenna, single user), and a transmit power of $p_{\mathrm{u}} = 10$ dB. The coherence interval is $T = 196$ symbols. For the dashed curves (marked with $K = 1$), the transmit power p_{u} and the fraction of the coherence interval τ/T spent on training was optimized in order to maximize the energy efficiency for a fixed spectral efficiency. For the green and red curves (marked MRC and ZF; shown for $M = 50$ and $M = 100$ antennas, respectively), the number of users K was optimized jointly with p_{u} and τ/T to maximize the energy efficiency for given spectral efficiency. Any operating point on the curves can be obtained by appropriately selecting p_{u} and optimizing with respect to K and τ/T. The number marked next to the $\times$ marks on each curve is the power p_{u} spent by the transmitter.

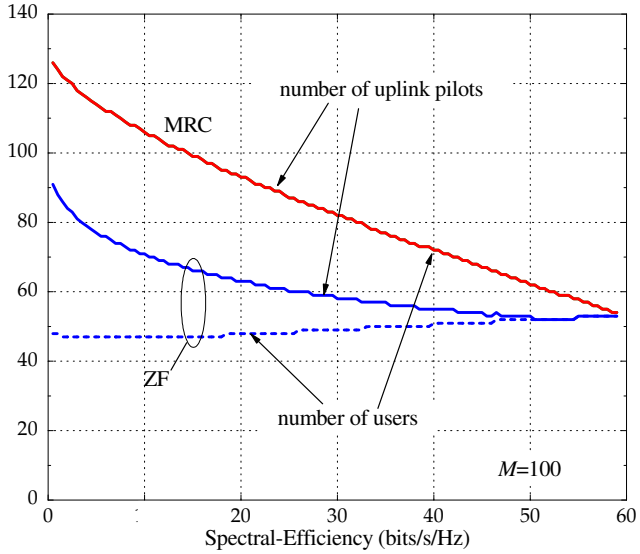

Fig. 7. Optimal number of users K and number of symbols τ spent on training, out of a total of $T = 196$ symbols per coherence interval, for the curves in Fig. 6 corresponding to $M = 100$ antennas.

η_{IP}^{0} and it is therefore dimensionless. The dotted and dashed lines show the performances for the cases of $M = 1, K = 1$ and $M = 100, K = 1$, respectively. Each point on the curves is obtained by choosing the transmit power p_{u} and pilot sequence length τ to maximize the energy efficiency for a given spectral efficiency. The solid lines show the performance for the cases of $M = 50$, and 100. Each point on these curves is computed by jointly choosing K, τ, and p_{u} to maximize the energy-efficiency subject a fixed spectral-efficiency, i.e.,

$$\arg \max_{p_{\mathrm{u}}, K, \tau} \ \eta_{\mathrm{IP}}^{\mathtt{A}}, \qquad \text{s.t. } R_{\mathrm{IP}}^{\mathtt{A}} = \text{const.}, K \leq \tau \leq T$$

We first consider a single-user system with $K = 1$. We compare the performance of the cases $M = 1$ and $M = 100$. Since $K = 1$ the performances of MRC and ZF are equal. With the same power used as in the reference mode, i.e., $p_{\mathrm{u}} = 10$ dB, using 100 antennas can increase the spectral efficiency and the energy efficiency by factors of 4 and 3, respectively. Reducing the transmit power by a factor of 100, from 10 dB to -10 dB yields a 100-fold improvement in energy efficiency compared with that of the reference mode with no reduction in spectral-efficiency.

We next consider a multiuser system ($K > 1$). Here the transmit power p_{u}, the number of users K, and the duration of pilot sequences τ are chosen optimally for fixed M. We consider $M = 50$ and 100. The system performance improves very significantly compared to the single-user case. For example, with MRC, at $p_{\mathrm{u}} = 0$ dB, compared with the case of $M = 1, K = 1$, the spectral-efficiency increases by factors of 50 and 80, while the energy-efficiency increases by factors of 55 and 75 for $M = 50$ and $M = 100$, respectively. As discussed in Section IV, at low spectral efficiency, the energy efficiency increases when the spectral efficiency increases. Furthermore, we can see that at high spectral efficiency, ZF outperforms MRC. This is due to the fact that MRC is limited by the intracell interference, which is significant at high spectral efficiency. As a consequence, when p_{u} is increased, the spectral efficiency of MRC approaches a constant value, while the energy efficiency goes to zero (see (66)).

The corresponding optimum values of K and τ as functions of the spectral efficiency for $M = 100$ are shown in Fig. 7. For MRC, the optimal number of users and uplink pilots are the same (this means that the minimal possible length of training sequences are used). For ZF, more of the coherence interval is used for training. Generally, at low transmit power and therefore at low spectral efficiency, we spend more time on training than on payload data transmission. At high power (high spectral efficiency and low energy efficiency), we can serve around 55 users, and $K = \tau$ for both MRC and ZF.

B. Multicell MU-MIMO Systems

Next, we examine the effect of pilot contamination on the energy and spectral efficiency for multicell systems. We consider a system with $L = 7$ cells. Each cell has the same size as in the single-cell system. When shrinking the cell size, one typically also cuts back on the power. Hence, the relation between signal and interference power would not be substantially different in systems with smaller cells and in that sense, the analysis is largely independent of the actual physical size of the cell [23]. Note that, setting $L = 7$ means that we consider the performance of a given cell with the interference from 6 nearest-neighbor cells. We assume $\boldsymbol{D}_{ll} = \boldsymbol{I}_K$, and $\boldsymbol{D}_{li} = \beta \boldsymbol{I}_K$, for $i \neq l$. To examine the performance in a practical scenario, the intercell interference factor, β, is chosen as follows. We consider two users, the 1st user is located

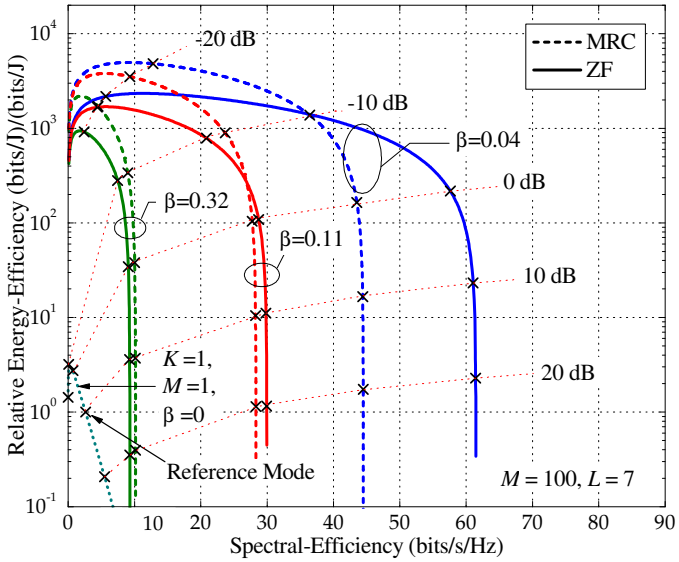

Fig. 8. Same as Figure 6, but for a multicell scenario, with $L = 7$ cells, and coherence interval $T = 196$.

uniformly at random in the first cell, and the 2nd user is located uniformly at random in one of the 6 nearest-neighbor cells of the 1st cell. Let $\bar{\beta}_1$ and $\bar{\beta}_2$ be the large scale fading from the 1st user and the 2nd user to the 1st BS, respectively. (The large scale fading is modelled as in Section V-A1.) Then we compute β as $\mathbb{E}\left\{\bar{\beta}_2/\bar{\beta}_1\right\}$. By simulation, we obtain $\beta = 0.32, 0.11$, and 0.04 for the cases of ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3.8, f_{\text{reuse}} = 1$), ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3, f_{\text{reuse}} = 1$), and ($\sigma_{\text{shadow}} = 8$ dB, $\nu = 3.8, f_{\text{reuse}} = 3$), respectively, where f_{reuse} is the frequency reuse factor.

Fig. 8 shows the relative energy efficiency versus the spectral efficiency for MRC and ZF of the multicell system. The reference mode is the same as the one in Fig. 6 for a single-cell system. The dotted line shows the performance for the case of $M = 1, K = 1$, and $\beta = 0$. The solid and dashed lines show the performance for the cases of $M = 100$, and $L = 7$, with different intercell interference factors β of $0.32, 0.11$, and 0.04. Each point on these curves is computed by jointly choosing τ, K, and p_{u} to maximize the energy efficiency for a given spectral efficiency. We can see that the pilot contamination significantly degrades the system performance. For example, when β increases from 0.11 to 0.32 (and hence, the pilot contamination increases), with the same power, $p_{\text{u}} = 10$ dB, the spectral efficiency and the energy efficiency reduce by factors of 3 and 2.7, respectively. However, with low transmit power where the spectral efficiency is smaller than 10 bits/s/Hz, the system performance is not affected much by the pilot contamination. Furthermore, we can see that in a multicell scenario with high pilot contamination, MRC achieves a better performance than ZF.

VI. Conclusion

Very large MIMO systems offer the opportunity of increasing the spectral efficiency (in terms of bits/s/Hz sum-rate) by one or two orders of magnitude, and simultaneously improving the energy efficiency (in terms of bits/J) by three orders of magnitude. This is possible with simple linear processing such as MRC or ZF at the BS, and using channel estimates obtained from uplink pilots even in a high mobility environment where half of the channel coherence interval is used for training. Generally, ZF outperforms MRC owing to its ability to cancel intracell interference. However, in multicell environments with strong pilot contamination, this advantage tends to diminish. MRC has the additional benefit of facilitating a distributed per-antenna implementation of the detector. Quantitatively, with MRC, 100 antennas can serve about 50 terminals in the same time-frequency resource, each terminal having a fading-free throughput of about 1 bpcu, and hence the system offering a sum-throughput of about 50 bpcu. These conclusions are valid under a channel model that includes the effects of small-scale Rayleigh fading, but neglects the effects of large-scale fading (see the discussion after (63)).

Appendix

A. Proof of Proposition 2

From (15), we have

$$\tilde{R}^{\mathtt{mrc}}_{\text{P},k} = \log_2\left(1 + \left(\mathbb{E}\left\{\frac{p_{\text{u}}\sum_{i=1,i\neq k}^{K}|\tilde{g}_i|^2 + 1}{p_{\text{u}}\|\boldsymbol{g}_k\|^2}\right\}\right)^{-1}\right) \quad (75)$$

where $\tilde{g}_i \triangleq \frac{\boldsymbol{g}_k^H \boldsymbol{g}_i}{\|\boldsymbol{g}_k\|}$. Conditioned on $\boldsymbol{g}_k$, $\tilde{g}_i$ is a Gaussian RV with zero mean and variance β_i which does not depend on $\boldsymbol{g}_k$. Therefore, $\tilde{g}_i$ is Gaussian distributed and independent of $\boldsymbol{g}_k$, $\tilde{g}_i \sim \mathcal{CN}(0, \beta_i)$. Then,

$$\begin{aligned}\mathbb{E}&\left\{\frac{p_{\text{u}}\sum_{i=1,i\neq k}^{K}|\tilde{g}_i|^2 + 1}{p_{\text{u}}\|\boldsymbol{g}_k\|^2}\right\} \\ &= \left(p_{\text{u}}\sum_{i=1,i\neq k}^{K}\mathbb{E}\left\{|\tilde{g}_i|^2\right\} + 1\right)\mathbb{E}\left\{\frac{1}{p_{\text{u}}\|\boldsymbol{g}_k\|^2}\right\} \\ &= \left(p_{\text{u}}\sum_{i=1,i\neq k}^{K}\beta_i + 1\right)\mathbb{E}\left\{\frac{1}{p_{\text{u}}\|\boldsymbol{g}_k\|^2}\right\}. \quad (76)\end{aligned}$$

Using the identity [22]

$$\mathbb{E}\left\{\mathtt{tr}\left(\boldsymbol{W}^{-1}\right)\right\} = m/(n-m) \quad (77)$$

where $\boldsymbol{W} \sim \mathcal{W}_m(n, \boldsymbol{I}_n)$ is an $m \times m$ central complex Wishart matrix with n ($n > m$) degrees of freedom, we obtain

$$\mathbb{E}\left\{\frac{1}{p_{\text{u}}\|\boldsymbol{g}_k\|^2}\right\} = \frac{1}{p_{\text{u}}(M-1)\beta_k}, \text{ for } M \geq 2. \quad (78)$$

Substituting (78) into (76), we arrive at the desired result (16).

B. Proof of Proposition 3

From (3), we have

$$\begin{aligned}\mathbb{E}\left\{\left[\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\right]_{kk}\right\} &= \frac{1}{\beta_k}\mathbb{E}\left\{\left[\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\right]_{kk}\right\} \\ &= \frac{1}{K\beta_k}\mathbb{E}\left\{\mathtt{tr}\left[\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\right]\right\} \\ &\overset{(a)}{=} \frac{1}{(M-K)\beta_k}, \text{ for } M \geq K+1 \quad (79)\end{aligned}$$

where (a) is obtained by using (77). Using (79), we get (20).

REFERENCES

[1] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. Allerton Conf. Commun., Control, Comput.*, Urbana-Champaign, IL., Sept. 2011.

[2] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Sig. Proc. Mag.*, vol. 24, no. 5, pp. 36–46, 2007.

[3] 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, 2010.

[4] 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.

[5] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[6] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality" *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[7] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

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

[9] ——, "How much training is required for multiuser MIMO," in *Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC '06)*, Pacific Grove, CA, USA, Oct. 2006, pp. 359–363.

[10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Mag.*, accepted. [Online]. Available: arxiv.org/abs/1201.3210.

[11] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," in *IEEE J. Sel. Areas Commun.*, 2012, accepted.

[12] A. Fehske, G. Fettweis, J. Malmodin and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Communications Magazine*, pp. 55-62, August 2011.

[13] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[14] H. Huh, G. Caire, H. C. Papadopoulos, S. A. Rampshad, "Achieving large spectral efficiency with TDD and not-so-many base station antennas," in *Proc. IEEE Antennas and Propagation in Wireless Communications (APWC)*, 2011.

[15] S. Wagner, R. Couillet, D. T. M. Slock, and M. Debbah, "Large system analysis of zero-forcing precoding in MISO broadcast channels with limited feedback," in *Proc. IEEE Int. Works. Signal Process. Adv. Wireless Commun. (SPAWC)*, 2010.

[16] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beanforming in large-scale antenna systems", *IEEE J. Select. Areas Commun.*, 2012, accepted.

[17] H. Cramér, *Random Variables and Probability Distributions*. Cambridge, UK: Cambridge University Press, 1970.

[18] N. Kim and H. Park, "Performance analysis of MIMO system with linear MMSE receiver," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4474–4478, Nov. 2008.

[19] H. Gao, P. J. Smith, and M. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.

[20] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA: Academic, 2007.

[22] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

[23] A. Lozano, R. W. Heath Jr., and J. G. Andrews, "Fundamentral limits of cooperation," Mar. 2012. [Online]. Available: arxiv.org/abs/1204.0011.

Hien Quoc Ngo received the B.S. degree in Electrical Engineering from Ho Chi Minh City University of Technology, Vietnam, in 2007, and the M.S. degree in Electronics and Radio Engineering from Kyung Hee University, Korea, in 2010. From 2008 to 2010, he was with the Communication and Coding Theory Laboratory, Kyung Hee University, where he did research on wireless communication and information theories, in particular are cooperative communications, game theory, and network connectivity. Since April 2010, he is a Ph.D. student of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. His current research interests include large-scale (massive) MIMO systems, cooperative communications, and interference networks.

Erik G. Larsson received his Ph.D. degree from Uppsala University, Sweden, in 2002. Since 2007, he is Professor and Head of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. He has previously been Associate Professor (Docent) at the Royal Institute of Technology (KTH) in Stockholm, Sweden, and Assistant Professor at the University of Florida and the George Washington University, USA.

His main professional interests are within the areas of wireless communications and signal processing. He has published some 80 journal papers on these topics, he is co-author of the textbook *Space-Time Block Coding for Wireless Communications* (Cambridge Univ. Press, 2003) and he holds 10 patents on wireless technology.

He is Associate Editor for the *IEEE Transactions on Communications* and he has previously been Associate Editor for several other IEEE journals. He is a member of the IEEE Signal Processing Society SAM and SPCOM technical committees. He is active in conference organization, most recently as the Technical Chair of the Asilomar Conference on Signals, Systems and Computers 2012 and Technical Program co-chair of the International Symposium on Turbo Codes and Iterative Information Processing 2012.

Thomas L. Marzetta was born in Washington, D.C. He received the PhD in electrical engineering from the Massachusetts Institute of Technology in 1978. His dissertation extended, to two dimensions, the three-way equivalence of autocorrelation sequences, minimum-phase prediction error filters, and reflection coefficient sequences. He worked for Schlumberger-Doll Research (1978-1987) to modernize geophysical signal processing for petroleum exploration. He headed a group at Nichols Research Corporation (1987-1995) which improved automatic target recognition, radar signal processing, and video motion detection. He joined Bell Laboratories in 1995 (formerly part of AT&T, then Lucent Technologies, now Alcatel-Lucent). He has had research supervisory responsibilities in communication theory, statistics, and signal processing. He specializes in multiple-antenna wireless, with a particular emphasis on the acquisition and exploitation of channel-state information. He is the originator of Large-Scale Antenna Systems which can provide huge improvements in wireless spectral-efficiency and energy-efficiency over 4G technology.

Dr. Marzetta was a member of the IEEE Signal Processing Society Technical Committee on Multidimensional Signal Processing, a member of the Sensor Array and Multichannel Technical Committee, an associate editor for the IEEE Transactions on Signal Processing, an associate editor for the IEEE Transactions on Image Processing, and a guest associate editor for the IEEE Transactions on Information Theory Special Issue on Signal Processing Techniques for Space-Time Coded Transmissions (Oct. 2002), for the IEEE Transactions on Information Theory Special Issue on Space-Time Transmission, Reception, Coding, and Signal Design (Oct. 2003), and for the IEEE JSAC Special Issue on Large-Scale Multiple Antenna Wireless Systems (Feb. 2013).

Dr. Marzetta was the recipient of the 1981 ASSP Paper Award from the IEEE Signal Processing Society. He was elected a Fellow of the IEEE in Jan. 2003.