

Pilot Contamination and Precoding in Multi-Cell TDD Systems

Jubin Jose, *Student Member, IEEE*, Alexei Ashikhmin, *Senior Member, IEEE*,
Thomas L. Marzetta, *Fellow, IEEE*, and Sriram Vishwanath, *Senior Member, IEEE*

Abstract—This paper considers a multi-cell multiple antenna system with precoding used at the base stations for downlink transmission. Channel state information (CSI) is essential for precoding at the base stations. An effective technique for obtaining this CSI is time-division duplex (TDD) operation where uplink training in conjunction with reciprocity simultaneously provides the base stations with downlink as well as uplink channel estimates. This paper mathematically characterizes the impact that uplink training has on the performance of such multi-cell multiple antenna systems. When non-orthogonal training sequences are used for uplink training, the paper shows that the precoding matrix used by the base station in one cell becomes corrupted by the channel between that base station and the users in other cells in an undesirable manner. This paper analyzes this fundamental problem of pilot contamination in multi-cell systems. Furthermore, it develops a new multi-cell MMSE-based precoding method that mitigates this problem. In addition to being linear, this precoding method has a simple closed-form expression that results from an intuitive optimization. Numerical results show significant performance gains compared to certain popular single-cell precoding methods.

Index Terms—Time-division duplex systems, uplink training, pilot contamination, MMSE precoding.

I. INTRODUCTION

MULTIPLE antennas, especially at the base-station, have now become an accepted and a central feature of cellular networks. These networks have been studied extensively over the past one and a half decades (see [2] and references therein). It is now well understood that channel state information (CSI) at the base station is an essential component when trying to maximize network throughput. Systems with varying degrees of CSI have been studied in great detail in literature. The primary framework under which these have been studied is frequency division duplex (FDD) systems, where the CSI is typically obtained through (limited) feedback. There is a rich body of work in jointly designing this feedback mechanism with (pre)coding strategies to maximize

throughput in MIMO downlink [3]–[8]. Time division duplex (TDD) systems, however, have a fundamentally different architecture from the ones studied in FDD systems [9], [10]. The goal of this paper is to develop a clear understanding of mechanisms for acquiring CSI and subsequently designing precoding strategies for multi-cell MIMO TDD systems.

Although most current cellular systems are frequency-division duplex (FDD), there is a compelling case to be made for time-division duplex (TDD). Arguably the central problem in advanced wireless systems is the acquisition of CSI in a timely manner. A major point of this paper is the desirability of having a large excess of base station antennas compared with terminals. For fast-changing channels TDD offers the only way to acquire timely CSI since the training burden for reverse-link pilots in a TDD system is independent of the number of base station antennas, while conversely the training burden for forward-link pilots in an FDD system is proportional to the number of antennas. Use of FDD would forever impose a severe limit on the number of base station antennas.

An important distinguishing feature of TDD systems is the notion of *reciprocity*, where the reverse channel is used as an estimate of the forward channel. While utilizing reciprocity, the differences in the transfer characteristics of the amplifiers and the filters in the two directions must be accounted for. This can be handled by measuring the different scalar gains in the two directions. Arguably, this reverse channel estimation one of the best advantages of a TDD architecture, as it eliminates the need for feedback, and uplink training together with the reciprocity of the wireless medium [11], [12] is sufficient to provide us with the desired CSI. In [13], channel reciprocity has been validated through experiments. However, as we see next, this channel estimate is not without issues that must be addressed before it proves useful.

In this paper, we consider uplink training and transmit precoding in a multi-cell scenario with L cells, where each cell consists of a base station with M antennas and K users with single antenna each. The impact of uplink training on the resulting channel estimate (and thus system performance) in the multi-cell scenario is significantly different from that in a single-cell scenario. In the multi-cell scenario, non-orthogonal training sequences (pilots) must be utilized, as orthogonal pilots would need to be least $K \times L$ symbols long which is infeasible for large L . In particular, short channel coherence times due to mobility do not allow for such long training sequences.

This non-orthogonal nature causes *pilot contamination* [1],

Manuscript received June 29, 2010; revised January 27, 2011; accepted March 10, 2011. The associate editor coordinating the review of this paper and approving it for publication was A. Hjørungnes.

Results in this paper were presented in part at the IEEE International Symposium on Information Theory (ISIT) [1].

J. Jose and S. Vishwanath are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: {jubin, sriram}@austin.utexas.edu). The authors were supported in part by the DoD and the ARO Young Investigator Program (YIP).

A. Ashikhmin and T. L. Marzetta are with Bell Laboratories, Alcatel-Lucent Inc., Murray Hill, NJ 07974 USA (e-mail: aea@research.bell-labs.com; tom.marzetta@alcatel-lucent.com).

Digital Object Identifier 10.1109/TWC.2011.060711.101155

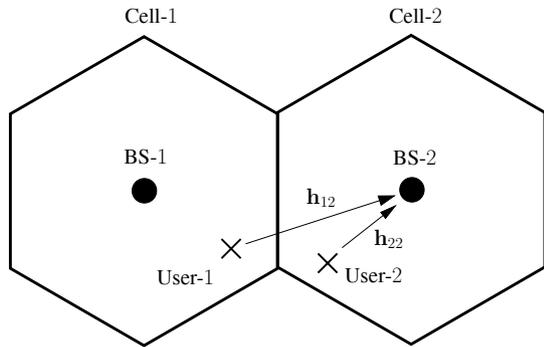


Fig. 1. A two-cell example with one user in each cell. Both users transmit non-orthogonal pilots during uplink training, which leads to pilot contamination at both the base stations.

[14], which is encountered only when analyzing a multi-cell MIMO system with training, and is lost when narrowing focus to a single-cell setting or to a multi-cell setting where channel information is assumed available at no cost. In this paper, we perform a more detailed study of this problem and consider precoding in its presence. First, we study the impact of pilot contamination (and thus achievable rates), and then, we develop methods that mitigate this contamination. In older generation cellular systems, multiple factors including large reuse distance of any training signal and randomization in the selection of pilots would have helped in keeping the impact of pilot contamination reasonable. However, with newer generation systems designed for more aggressive reuse of spectrum, this impact is very crucial to understand and mitigate. We note that pilot contamination must also figure in Cooperative MIMO (also called Network MIMO [15], [16]) where clusters of base stations are wired together to create distributed arrays, and where pilots must be re-used over multiple clusters.

The fundamental problem associated with pilot contamination is evident even in the simple multi-cell scenario shown in Figure 1. Consider two cells $i \in \{1, 2\}$, each consisting of one base station and one user. Let \mathbf{h}_{ij} denote the channel between the base station in the i -th cell and the user in the j -th cell. Let the training sequences used by both the users be same. In this case, the MMSE channel estimate of \mathbf{h}_{22} at the base station in the 2-nd cell is $\hat{\mathbf{h}}_{22} = c_1 \mathbf{h}_{12} + c_2 \mathbf{h}_{22} + c\mathbf{w}$. Here c_1 , c_2 and c are constants that depend on the propagation factors and the transmit powers of mobiles, and \mathbf{w} is $\mathcal{CN}(0, \mathbf{I})$ additive noise. The base station in the 2-nd cell uses this channel estimate to form a precoding vector $\mathbf{a}_2 = f(\hat{\mathbf{h}}_{22})$, which is usually aligned with the channel estimate, that is $\mathbf{a}_2 = \text{const} \cdot \hat{\mathbf{h}}_{22}$. However, by doing this, the base station (partially) aligns the transmitted signal with both \mathbf{h}_{22} (which is desirable) and \mathbf{h}_{12} (which is undesirable). Both signal ($\mathbf{h}_{22} \mathbf{a}_2^\dagger$) and interference ($\mathbf{h}_{12} \mathbf{a}_2^\dagger$) statistically behave similarly. Therefore, the general assumption that the precoding vector used by a base station in one cell is uncorrelated with the channel to users in other cells is not valid with uplink training using non-orthogonal training sequences. This fundamental problem is studied in further detail.

To perform pilot contamination analysis, we first develop analytical expressions using techniques similar to those used

in [9], [10]. For the setting with one user in every cell, we derive closed-form expressions for achievable rates. These closed-form expressions allow us to determine the extent to which pilot contamination impacts system performance. In particular, we show that the achievable rates can saturate with the number of antennas at the base station M . This analysis will allow system designers to determine the appropriate frequency/time/pilot reuse factor to maximize system throughput in the presence of pilot contamination.

In the multi-cell scenario, there has been significant work on utilizing coordination among base stations [15]–[18] when CSI is available. This existing body of work focuses on the gain that can be obtained through coordination of the base stations. Dirty paper coding based approaches and joint beamforming/precoding approaches are considered in [18]. Linear precoding methods for clustered networks with full intra-cluster coordination and limited inter-cluster coordination are proposed in [19]. These approaches generally require “good” channel estimates at the base stations. Due to non-orthogonal training sequences, the resulting channel estimate (of the channel between a base station and all users) can be shown to be rank deficient. We develop a multi-cell MMSE-based precoding method that depends on the set of training sequences assigned to the users. Note that this MMSE-based precoding is for the general setting with multiple users in every cell. Our approach does not need coordination between base stations required by the joint precoding techniques¹. When coordination is present, this approach can be applied at the inter-cluster level. The MMSE-based precoding derived in this paper has several advantages. In addition to being a linear precoding method, it has a simple closed-form expression that results from an intuitive optimization problem formulation. For many training sequence allocations, numerical results show that our approach gives significant gains over certain popular single-cell precoding methods including zero-forcing.

A. Related Work

Over the past decade, a variety of aspects of downlink and uplink transmission problems in a single cell setting have been studied. In information theoretic literature, these problems are studied as the broadcast channel (BC) and the multiple access channel (MAC) respectively. For Gaussian BC and general MAC, the problems have been studied for both single and multiple antenna cases. The sum capacity of the multi-antenna Gaussian BC has been shown to be achieved by dirty paper coding (DPC) in [20]–[23]. It was shown in [24] that DPC characterizes the full capacity region of the multi-antenna Gaussian BC. These results assume perfect CSI at the base station and the users. In addition, the DPC technique is computationally challenging to implement in practice. There has been significant research focus on reducing the computational complexity at the base station and the users. In this regard, different precoding schemes with low complexity have been proposed. This body of work [25]–[29] demonstrates that sum rates close to sum capacity can be achieved with much lower computational complexity. However, these results assume perfect CSI at the base station and the users.

¹There is no exchange of channel state information among base stations.

The problem of lack of channel CSI is usually studied by considering one of the following two settings. As discussed before, in the first setting, CSI at users is assumed to be available and a limited feedback link is assumed to exist from the users to the base station. In [3], [5]–[8], [30] such a setting is considered. In [5], the authors show that at high signal to noise ratios (SNRs), the feedback rate required per user must grow linearly with the SNR (in dB) in order to obtain the full MIMO BC multiplexing gain. The main result in [6] is that the extent of CSI feedback can be reduced by exploiting multi-user diversity. In [7] it is shown that nonrandom vector quantizers can significantly increase the MIMO downlink throughput. In [8], the authors design a joint CSI quantization, beamforming and scheduling algorithm to attain optimal throughput scaling. In the next setting, time division duplex systems are considered and channel training and estimation error are accounted for in the net achievable rate. This approach is used in [9], [10], [31], [32]. In [9], the authors give a lower bound on sum capacity and demonstrate that it is always beneficial to increase the number of antennas at the base station. In [10], the authors study a heterogeneous user setting and present scheduling and precoding methods for this setting. In [31], the authors consider two-way training and propose two variants of linear MMSE precoders as alternatives to linear zero-forcing precoder used in [9]. Single-cell analysis of TDD systems are also provided in [32].

Given this extensive body of literature in single-cell systems, the main contribution of this paper is in understanding multi-cell systems with channel training. Its emphasis is on TDD systems, which are arguably poorly studied compared to FDD systems. Specifically, the main contributions are to demonstrate the pilot contamination problem associated with uplink training, understand its impact on the operation of multi-cell MIMO TDD cellular systems, and develop a new precoding method to mitigate this problem.

B. Notation

We use bold font variables to denote matrices and vectors. $(\cdot)^T$ denotes the transpose and $(\cdot)^\dagger$ denotes the Hermitian transpose, $\text{tr}\{\cdot\}$ denotes the trace operation, $(\cdot)^{-1}$ denotes the inverse operation, and $\|\cdot\|$ denotes the two-norm. $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with diagonal entries equal to the components of \mathbf{a} . $\mathbb{E}[\cdot]$ and $\text{var}\{\cdot\}$ stand for expectation and variance operations, respectively.

C. Organization

The rest of this paper is organized as follows. In Section II, we describe the multi-cell system model. In Section III, we explain the communication scheme and the technique to obtain achievable rates. We analyze the effect of pilot contamination in Section IV, and give the details of the new precoding method in Section V. We present numerical results in Section VI. Finally, we provide our concluding remarks in Section VII.

II. MULTI-CELL TDD SYSTEM MODEL

We consider a cellular system with L cells numbered $1, 2, \dots, L$. Each cell consists of one base station with M

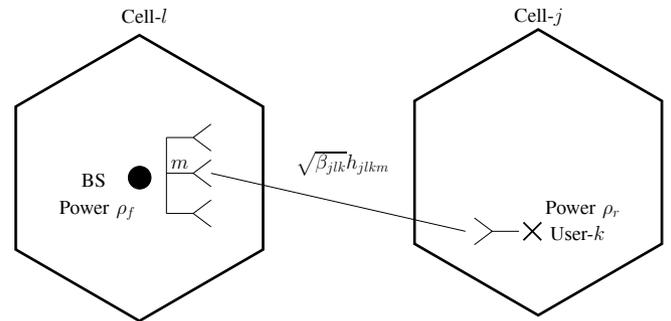


Fig. 2. System model showing the base station in l -th cell and the k -th user in j -th cell.

antennas and $K (\leq M)$ single-antenna users². Let the average power (during transmission) at the base station be p_f and the average power (during transmission) at each user be p_r . The propagation factor between the m -th base station antenna of the l -th cell and the k -th user of the j -th cell is $\sqrt{\beta_{jlk}} h_{jlk m}$, where $\{\beta_{jlk}\}$ are non-negative constants and assumed to be known to everybody, and $\{h_{jlk m}\}$ are independent and identically distributed (i.i.d.) zero-mean, circularly-symmetric complex Gaussian $\mathcal{CN}(0, 1)$ random variables and known to nobody³. This system model is shown in Figure 2. The above assumptions are fairly accurate and justified due to the following reason. The $\{\beta_{jlk}\}$ values model path-loss and shadowing that change slowly and can be learned over long period of time, while the $\{h_{jlk m}\}$ values model fading that change relatively fast and must be learned and used very quickly. Since the cell layout and shadowing are captured using the constant $\{\beta_{jlk}\}$ values, for the purpose of this paper, the specific details of the cell layout and shadowing model are irrelevant. In other words, any cell layout and any shadowing model can be incorporated with the above abstraction.

We assume channel reciprocity for the forward and reverse links, i.e., the propagation factor $\sqrt{\beta_{jlk}} h_{jlk m}$ is same for both forward and reverse links, and block fading, i.e., $\{h_{jlk m}\}$ remains constant for a duration of T symbols. Note that we allow for a constant factor variation in forward and reverse propagation factors through the different average power constraints at the base stations and the users. The additive noises at all terminals are i.i.d. $\mathcal{CN}(0, 1)$ random variables. The system equations describing the signals received at the base station and the users are given in the next section.

III. COMMUNICATION SCHEME

The communication scheme consists of two phases: uplink training and data transmission. Uplink training phase consists of users transmitting training pilots, and base stations obtaining channel estimates. Data transmission phase consists of base stations transmitting data to the users through transmit precoding. Next, we describe these phases briefly and provide a set of achievable data rates using a given precoding method.

²These are the users of interest for applying the precoding framework.

³For compact notation, we do not separate the subscript or superscript indices using commas throughout the paper.

A. Uplink Training

At the beginning of every coherence interval, all users (in all cells) transmit training sequences, which are τ length column vectors⁴. Let $\sqrt{\tau}\boldsymbol{\psi}_{jk}$ (normalized such that $\boldsymbol{\psi}_{jk}^\dagger\boldsymbol{\psi}_{jk} = 1$) be the training vector transmitted by the k -th user in the j -th cell. Consider the base station of the l -th cell. The τ length column vector received at the m -th antenna of this base station is

$$\mathbf{y}_{lm} = \sum_{j=1}^L \sum_{k=1}^K \sqrt{p_r\tau\beta_{jlk}} h_{jlkM} \boldsymbol{\psi}_{jk} + \mathbf{w}_{lm}, \quad (1)$$

where \mathbf{w}_{lm} is the additive noise. Let $\mathbf{Y}_l = [\mathbf{y}_{l1} \mathbf{y}_{l2} \dots \mathbf{y}_{lM}]$ ($\tau \times M$ matrix), $\mathbf{W}_l = [\mathbf{w}_{l1} \mathbf{w}_{l2} \dots \mathbf{w}_{lM}]$ ($\tau \times M$ matrix), $\boldsymbol{\Psi}_j = [\boldsymbol{\psi}_{j1} \boldsymbol{\psi}_{j2} \dots \boldsymbol{\psi}_{jK}]$ ($\tau \times K$ matrix), $\mathbf{D}_{jl} = \text{diag}\{\beta_{j1l} \beta_{j2l} \dots \beta_{jKl}\}$, and

$$\mathbf{H}_{jl} = \begin{bmatrix} h_{j1l1} & \dots & h_{j1lM} \\ \vdots & \ddots & \vdots \\ h_{jKl1} & \dots & h_{jKlM} \end{bmatrix}.$$

From (1), the signal received at this base station can be expressed as

$$\mathbf{Y}_l = \sqrt{p_r\tau} \sum_{j=1}^L \boldsymbol{\Psi}_j \mathbf{D}_{jl}^{\frac{1}{2}} \mathbf{H}_{jl} + \mathbf{W}_l. \quad (2)$$

The MMSE estimate of the channel \mathbf{H}_{il} given \mathbf{Y}_l in (2) is

$$\hat{\mathbf{H}}_{jl} = \sqrt{p_r\tau} \mathbf{D}_{jl}^{\frac{1}{2}} \boldsymbol{\Psi}_j^\dagger \left(\mathbf{I} + p_r\tau \sum_{i=1}^L \boldsymbol{\Psi}_i \mathbf{D}_{il} \boldsymbol{\Psi}_i^\dagger \right)^{-1} \mathbf{Y}_l. \quad (3)$$

This MMSE estimate in (3) follows from standard results in estimation theory (for example see [33]). We denote the MMSE estimate of the channel between this base station and all users by $\hat{\mathbf{H}}_l = [\hat{\mathbf{H}}_{l1} \hat{\mathbf{H}}_{l2} \dots \hat{\mathbf{H}}_{lL}]$. This notation is used later in Section IV.

B. Downlink Transmission

Consider the base station of the l -th cell. Let the information symbols to be transmitted to users in the l -th cell be $\mathbf{q}_l = [q_{l1} q_{l2} \dots q_{lK}]^T$ and the $M \times K$ linear precoding matrix be $\mathbf{A}_l = f(\hat{\mathbf{H}}_l)$. The function $f(\cdot)$ corresponds to the specific (linear) precoding method performed at the base station. The signal vector transmitted by this base station is $\mathbf{A}_l \mathbf{q}_l$. We consider transmission symbols and precoding methods such that $\mathbb{E}[\mathbf{q}_l] = \mathbf{0}$, $\mathbb{E}[\mathbf{q}_l \mathbf{q}_l^\dagger] = \mathbf{I}$, and $\text{tr}\{\mathbf{A}_l^\dagger \mathbf{A}_l\} = 1$. These (sufficient) conditions imply that the average power constraint at the base station is satisfied.

Now, consider the users in the j -th cell. The noisy signal vector received by these users is

$$\mathbf{x}_j = \sum_{l=1}^L \sqrt{p_f} \mathbf{D}_{jl}^{\frac{1}{2}} \mathbf{H}_{jl} \mathbf{A}_l \mathbf{q}_l + \mathbf{z}_j, \quad (K \times 1 \text{ vector}) \quad (4)$$

where \mathbf{z}_j is the additive noise. From (4), the signal received by the k -th user can be expressed as

$$x_{jk} = \sum_{l=1}^L \sum_{i=1}^K \sqrt{p_f \beta_{jlk}} [h_{jlk1} \dots h_{jlkM}] \mathbf{a}_{li} q_{li} + z_{jk}, \quad (5)$$

⁴We assume that there is time synchronization present in the system for coherent uplink transmission.

where \mathbf{a}_{li} is the i -th column of the precoding matrix \mathbf{A}_l and z_{jk} is the k -th element of \mathbf{z}_j .

C. Achievable Rates

Next, we provide a set of achievable rates using the method suggested in [9]. With the above communication scheme, the base stations have channel estimates while the users do not have any channel estimate. Therefore, the achievable rates we derive have a different structure compared to typical rate expressions. In particular, the effective noise term has channel variations around the mean in addition to typical terms.

Let $g_{li}^{jk} = \sqrt{p_f \beta_{jlk}} [h_{jlk1} \dots h_{jlkM}] \mathbf{a}_{li}$. Now, (5) can be written in the form

$$\begin{aligned} x_{jk} &= \sum_{l=1}^M \sum_{i=1}^K g_{li}^{jk} q_{li} + z_{jk}, \\ &= \mathbb{E} [g_{jk}^{jk}] q_{jk} + \left(g_{jk}^{jk} - \mathbb{E} [g_{jk}^{jk}] \right) q_{jk} \\ &\quad + \sum_{(l,i) \neq (j,k)} g_{li}^{jk} q_{li} + z_{jk}. \end{aligned} \quad (6)$$

In (6), the effective noise is defined as

$$z'_{jk} = \left(g_{jk}^{jk} - \mathbb{E} [g_{jk}^{jk}] \right) q_{jk} + \sum_{(l,i) \neq (j,k)} g_{li}^{jk} q_{li} + z_{jk}. \quad (7)$$

Now, the effective point-to-point channel described by (6) can be written in the familiar form

$$x_{jk} = \mathbb{E} [g_{jk}^{jk}] q_{jk} + z'_{jk}, \quad (8)$$

where q_{jk} is the input, x_{jk} is the output, $\mathbb{E}[g_{jk}^{jk}]$ is the known channel and z'_{jk} is the additive noise. This expectation is known as it only depends on the channel distribution and not the instantaneous channel. However, the additive noise is neither independent nor Gaussian. We use the result in [34] that shows that worst-case uncorrelated additive noise is independent Gaussian noise of same variance to derive the following achievable rates.

Theorem 1: Consider the point-to-point communication channels given by (8). Then, the following set of rates are achievable:

$$R_{jk} = C \left(\frac{|\mathbb{E} [g_{jk}^{jk}]|^2}{1 + \text{var} \{g_{jk}^{jk}\} + \sum_{(l,i) \neq (j,k)} \mathbb{E} [|g_{li}^{jk}|^2]} \right), \quad (9)$$

where $C(\theta) = \log_2(1 + \theta)$.

Proof: Consider complex Gaussian $\mathcal{CN}(0,1)$ distributions for all inputs Q_{jk} ⁵. Then, the resulting mutual information with this (not necessarily maximizing) input distribution $I(Q_{jk}; X_{jk})$ is an achievable rate. However, this does not result in a computable expression. To obtain a computable lower bound to this rate, we observe that input random variable Q_{jk} and effective noise Z'_{jk} given by (7) are uncorrelated based on the following: The input distributions are such that

⁵Upper case symbols are used in this proof to emphasize that these are random variables.

Q_{jk} is clearly independent of Q_{li} for all $(l, i) \neq (j, k)$ and Z_{jk} . Furthermore, Q_{jk} is independent of G_{jk}^{jk} . Therefore,

$$\begin{aligned} \mathbb{E} \left[\left(G_{jk}^{jk} - \mathbb{E} \left[G_{jk}^{jk} \right] \right) |Q_{jk}|^2 \right] &= \\ \mathbb{E} \left[\left(G_{jk}^{jk} - \mathbb{E} \left[G_{jk}^{jk} \right] \right) \right] \mathbb{E} \left[|Q_{jk}|^2 \right] &= 0. \end{aligned}$$

The variance of the effective noise is

$$\mathbb{E} \left[|Z'_{jk}|^2 \right] = \text{var} \left\{ G_{jk}^{jk} \right\} + \sum_{(l,i) \neq (j,k)} \mathbb{E} \left[|G_{li}^{jk}|^2 \right] + 1.$$

Now, from [34] (Theorem 1), we know the result that the channel with independent Gaussian noise \hat{Z}_{jk} with same variance given by

$$\hat{X}_{jk} = \mathbb{E} \left[G_{jk}^{jk} \right] Q_{jk} + \hat{Z}_{jk}$$

is worse, i.e.,

$$\begin{aligned} I(Q_{jk}; X_{jk}) &\geq I(Q_{jk}; \hat{X}_{jk}) \\ &= h(\hat{X}_{jk}) - h(\hat{X}_{jk}|Q_{jk}) \\ &= h(\hat{X}_{jk}) - h(\hat{Z}_{jk}). \end{aligned}$$

This completes the proof. \blacksquare

The set of achievable rates given by (9) is valid for any linear precoding method, and depends on the precoding method through the expectation and variance terms appearing in (9). Similar achievable rates are used in the single-cell setting as well to study and/or compare precoding methods. Next, we perform pilot contamination analysis using these achievable rates.

IV. PILOT CONTAMINATION ANALYSIS

We analyze the pilot contamination problem in the following setting: one user per cell ($K = 1$), same training sequence used by all users ($\psi_{j1} = \psi, \forall j$) and matched-filter (MF) precoding. We consider this setting as it captures the primary effect of pilot contamination which is the correlation between the precoding matrix (vector in this setting) used by the base station in a cell and channel to users in other cells. We provide simple and insightful analytical results in this setting. As mentioned earlier, we emphasize that the pilot contamination problem results from uplink training with non-orthogonal training sequences, and hence, it is not specific to the setting considered here. However, the level of its impact on the achievable rates would vary depending on the system settings.

In order to simplify notation, we drop the subscripts associated with the users in every cell. In this section, \mathbf{H}_{jl} , $\hat{\mathbf{H}}_{jl}$ and \mathbf{A}_l are vectors and we denote these using \mathbf{h}_{jl} , $\hat{\mathbf{h}}_{jl}$ and \mathbf{a}_l , respectively. The matched-filter used at the base station in the l -th cell is given by $\mathbf{a}_l = \hat{\mathbf{h}}_{jl}^\dagger / \|\hat{\mathbf{h}}_{jl}\|$. The user in the j -th cell receives signal from its base station and from other base stations. From (4), this received signal is

$$x_{jk} = \sqrt{p_f \beta_{jj}} \mathbf{h}_{jj} \mathbf{a}_j q_j + \sum_{l \neq j} \sqrt{p_f \beta_{jl}} \mathbf{h}_{jl} \mathbf{a}_l q_l + z_j. \quad (10)$$

We compute first and second order moments of the effective channel gain and the inter-cell interference and use these to

obtain a simple expression for the achievable rate given by (9).

In the setting considered here, the MMSE estimate of \mathbf{h}_{jl} based on \mathbf{Y}_l given by (3) can be simplified using matrix inversion lemma and the fact that $\psi^\dagger \psi = 1$ as follows:

$$\begin{aligned} \hat{\mathbf{h}}_{jl} &= \sqrt{p_r \tau \beta_{jl}} \psi^\dagger \left(\mathbf{I} + \psi \left(p_r \tau \sum_{i=1}^L \beta_{il} \right) \psi^\dagger \right)^{-1} \mathbf{Y}_l, \\ &= \sqrt{p_r \tau \beta_{jl}} \psi^\dagger \left(\mathbf{I} - \frac{\psi \left(p_r \tau \sum_{i=1}^L \beta_{il} \right) \psi^\dagger}{1 + p_r \tau \sum_{i=1}^L \beta_{il}} \right) \mathbf{Y}_l, \\ &= \sqrt{p_r \tau \beta_{jl}} \left(\psi^\dagger - \frac{\left(p_r \tau \sum_{i=1}^L \beta_{il} \right)}{1 + p_r \tau \sum_{i=1}^L \beta_{il}} \psi^\dagger \right) \mathbf{Y}_l, \\ &= \frac{\sqrt{p_r \tau \beta_{jl}}}{\kappa_l} \psi^\dagger \mathbf{Y}_l, \end{aligned}$$

where $\kappa_l = 1 + p_r \tau \sum_{i=1}^L \beta_{il}$.

Remark 1: The above channel estimates clearly suggest the graveness of the pilot contamination impairment. For a given base station, its estimate of every channel is simply a scaled version of the same vector $\psi^\dagger \mathbf{Y}_l$. Thus, it cannot distinguish between the channel to its user and other users, which makes pilot contamination a fundamental problem in multi-cell systems.

Since $\psi^\dagger \mathbf{Y}_l$ is proportional to the MMSE estimate of \mathbf{h}_{jl} for any j , we have

$$\frac{\hat{\mathbf{h}}_{jl}^\dagger}{\|\hat{\mathbf{h}}_{jl}\|} = \frac{\mathbf{Y}_l^\dagger \psi}{\|\psi^\dagger \mathbf{Y}_l\|}, \forall j. \quad (11)$$

Using (11), we obtain

$$\begin{aligned} \mathbf{h}_{jl} \mathbf{a}_l &= \mathbf{h}_{jl} \frac{\hat{\mathbf{h}}_{jl}^\dagger}{\|\hat{\mathbf{h}}_{jl}\|}, \\ &= \|\hat{\mathbf{h}}_{jl}\| + \tilde{\mathbf{h}}_{jl} \frac{\hat{\mathbf{h}}_{jl}^\dagger}{\|\hat{\mathbf{h}}_{jl}\|}, \end{aligned} \quad (12)$$

where $\tilde{\mathbf{h}}_{jl} = \mathbf{h}_{jl} - \hat{\mathbf{h}}_{jl}$. From the properties of MMSE estimation, we know that $\hat{\mathbf{h}}_{jl}$ is independent of $\tilde{\mathbf{h}}_{jl}$, $\hat{\mathbf{h}}_{jl}$ is $\mathcal{CN} \left(\mathbf{0}, \frac{p_r \tau \beta_{jl}}{\kappa_l} \mathbf{I} \right)$, and $\tilde{\mathbf{h}}_{jl}$ is $\mathcal{CN} \left(\mathbf{0}, \frac{1 + p_r \tau \sum_{i \neq j} \beta_{il}}{\kappa_l} \mathbf{I} \right)$. These results are used next.

From (12), we get

$$\begin{aligned} \mathbb{E} [\mathbf{h}_{jl} \mathbf{a}_l] &= \mathbb{E} \left[\|\hat{\mathbf{h}}_{jl}\| \right], \\ &= \sqrt{\frac{p_r \tau \beta_{jl}}{\kappa_l}} \mathbb{E} [\theta], \end{aligned} \quad (13)$$

where $\theta = \sqrt{\sum_{m=1}^M |u_m|^2}$ and $\{u_m\}$ is i.i.d. $\mathcal{CN}(0, 1)$. From (12), we also have

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{h}_{jl} \mathbf{a}_l\|^2 \right] &= \mathbb{E} \left[\|\hat{\mathbf{h}}_{jl}\|^2 \right] + \mathbb{E} \left[\frac{\hat{\mathbf{h}}_{jl}}{\|\hat{\mathbf{h}}_{jl}\|} \tilde{\mathbf{h}}_{jl} \tilde{\mathbf{h}}_{jl}^\dagger \frac{\hat{\mathbf{h}}_{jl}^\dagger}{\|\hat{\mathbf{h}}_{jl}\|} \right], \\ &= \frac{p_r \tau \beta_{jl}}{\kappa_l} \mathbb{E} [\theta^2] + \frac{1 + p_r \tau \sum_{i \neq j} \beta_{il}}{\kappa_l}. \end{aligned} \quad (14)$$

Next, we state two lemmas required to obtain a closed-form expression for the achievable rate.

Lemma 2: The effective channel gain in (10) has expectation

$$\mathbb{E} \left[\sqrt{p_f \beta_{jj}} \mathbf{h}_{jj} \mathbf{a}_j \right] = \left(p_f \beta_{jj} \frac{p_r \tau \beta_{jj}}{1 + p_r \tau \sum_{i=1}^L \beta_{ij}} \right)^{1/2} \mathbb{E}[\theta]$$

and variance $\text{var} \left\{ \sqrt{p_f \beta_{jj}} \mathbf{h}_{jj} \mathbf{a}_j \right\} =$

$$p_f \beta_{jj} \left(\frac{p_r \tau \beta_{jj}}{1 + p_r \tau \sum_{i=1}^L \beta_{ij}} \text{var}\{\theta\} + \frac{1 + p_r \tau \sum_{i \neq j} \beta_{ij}}{1 + p_r \tau \sum_{i=1}^L \beta_{ij}} \right).$$

Proof: The proof follows from (13) and (14). Note that $\text{var}\{\theta\} = \mathbb{E}[\theta^2] - (\mathbb{E}[\theta])^2$ by definition. ■

Lemma 3: For both signal and interference terms in (10), the first and second order moments are as follows:

$$\mathbb{E} \left[\sqrt{p_f \beta_{jl}} \mathbf{h}_{jl} \mathbf{a}_l \mathbf{q}_l \right] = 0,$$

$$\mathbb{E} \left[\left| \sqrt{p_f \beta_{jl}} \mathbf{h}_{jl} \mathbf{a}_l \mathbf{q}_l \right|^2 \right] = p_f \beta_{jl} \left(\frac{p_r \tau \beta_{jl}}{1 + p_r \tau \sum_{i=1}^L \beta_{il}} \mathbb{E}[\theta^2] + \frac{1 + p_r \tau \sum_{i \neq j} \beta_{il}}{1 + p_r \tau \sum_{i=1}^L \beta_{il}} \right).$$

Proof: Since $\mathbb{E}[\mathbf{q}_l] = 0$ and \mathbf{q}_l is independent of \mathbf{h}_{jl} and \mathbf{a}_l , it is clear that

$$\mathbb{E} \left[\sqrt{p_f \beta_{jl}} \mathbf{h}_{jl} \mathbf{a}_l \mathbf{q}_l \right] = 0.$$

The proof of the second order moment follows directly from (14). ■

The main result of this section is given in the next theorem. This theorem provides a closed-form expression for the achievable rates under the setting considered in this section, i.e., one user per cell ($K = 1$), same training sequence used by all users ($\psi_{j1} = \psi, \forall j$) and matched-filter (MF) precoding.

Theorem 4: For the setting considered, the achievable rate of the user in the j -th cell during downlink transmission in (9) is given by

$$C \left(\frac{p_f \beta_{jj} \frac{p_r \tau \beta_{jj}}{\kappa_j} \mathbb{E}^2[\theta]}{\sum_{l \neq j} p_f \beta_{jl} \frac{p_r \tau \beta_{jl}}{\kappa_l} \mathbb{E}[\theta^2] + \zeta} \right), \quad (15)$$

where

$$\zeta = 1 + p_f \beta_{jj} \frac{p_r \tau \beta_{jj}}{\kappa_j} \text{var}\{\theta\} + \sum_{l=1}^L p_f \beta_{jl} \frac{1 + p_r \tau \sum_{i \neq j} \beta_{il}}{\kappa_l},$$

$\kappa_j = 1 + p_r \tau \sum_{i=1}^L \beta_{ij}$, $\mathbb{E}[\theta] = \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)}$, $\mathbb{E}[\theta^2] = M$ and $\text{var}\{\theta\} = M - \mathbb{E}^2[\theta]$. Here, $\Gamma(\cdot)$ is the Gamma function. For large M , the following limiting expression for achievable rate can be obtained:

$$\lim_{M \rightarrow \infty} R_j = C \left(\frac{\frac{\beta_{jj}^2}{1 + p_r \tau \sum_{i=1}^L \beta_{ij}}}{\sum_{l \neq j} \frac{\beta_{jl}^2}{1 + p_r \tau \sum_{i=1}^L \beta_{il}}} \right). \quad (16)$$

Proof: The proof of (15) follows by substituting the results of Lemma 2 and Lemma 3 in (9). Since θ has a scaled (by a factor of $1/\sqrt{2}$) chi distribution with $2M$ degrees of

freedom, it is straightforward to see that $\mathbb{E}[\theta] = \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)}$, $\mathbb{E}[\theta^2] = M$ and $\text{var}\{\theta\} = M - \mathbb{E}^2[\theta]$.

Using the duplication formula

$$\Gamma(z) \Gamma \left(z + \frac{1}{2} \right) = 2^{(1-2z)} \sqrt{\pi} \Gamma(2z)$$

and Stirling's formula

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n n^n} e^{-n}} = 1,$$

we obtain

$$\begin{aligned} & \lim_{M \rightarrow \infty} \frac{1}{\sqrt{M}} \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)} \\ &= \lim_{M \rightarrow \infty} \sqrt{\frac{\pi}{M}} 2^{(1-2M)} \frac{(2M-1)!}{(M-1)!(M-1)!}, \\ &= \lim_{M \rightarrow \infty} \sqrt{\frac{\pi}{M}} 2^{(1-2M)} \frac{\sqrt{2\pi(2M-1)} (2M-1)^{(2M-1)}}{2\pi(M-1)(M-1)^{2(M-1)} e^{-1}}, \\ &= \lim_{M \rightarrow \infty} \sqrt{\frac{2M-1}{2M}} \left(1 + \frac{1}{2(M-1)} \right)^{2M-1} e^{-1}, \\ &= 1. \end{aligned}$$

Therefore, $\lim_{M \rightarrow \infty} \frac{\mathbb{E}^2[\theta]}{M} = 1$ and $\lim_{M \rightarrow \infty} \frac{\text{var}\{\theta\}}{M} = 0$. This completes the proof of (16). ■

For large M , the value of $\text{var}\{\theta\}$ ($\approx 1/4$) is insignificant compared to M . The results of the above theorem show that the performance does saturate with M . Typically, the reverse link is interference-limited, i.e., $p_r \tau \sum_{i=1}^L \beta_{il} \gg 1, \forall j$. The term $\sum_{i=1}^L \beta_{il}$ is the expected sum of squares of the propagation coefficients between the base station in the j -th cell and all users. Therefore, $\sum_{i=1}^L \beta_{il}$ is generally constant with respect to j . Using these approximations in (16), we get

$$R_j \approx C \left(\frac{\beta_{jj}^2}{\sum_{l \neq j} \beta_{jl}^2} \right).$$

This clearly show that the impact of pilot contamination can be very significant if cross gains (between cells) are of the same order of direct gains (within the same cell). It suggests frequency/time reuse and pilot reuse techniques to reduce the cross gains (in the same frequency/time) relative to the direct gains. The benefits of frequency reuse in the limit of an infinite number of antennas were demonstrated in [14].

Remark 2: Our result in Theorem 4 is not an asymptotic result. The expression in (15) is exact for any value of the number of antennas M at the base stations. Hence, this expression can be used to find the appropriate frequency/time reuse scheme for any given value of M and other system parameters. We do not focus on this in this paper, as this would depend largely on the actual system parameters including the cell layout and the shadowing model.

Remark 3: The result in Theorem 4 is for the setting with one user per cell. In the general setting with K users per cell, a similar analysis can be performed, however, it need not simplify to a simple closed-form expression. The achievable rate in (9) can be numerically evaluated in the general setting, and this can be used to numerically study the impact of pilot contamination.

To summarize, the impact of uplink training with non-orthogonal pilots can be serious when the cross-gains are not small compared to the direct gains. This pilot contamination problem is often neglected in theory and even in many large-scale simulations. The analysis in this section shows the need to account for this impact especially in systems with high reuse of training sequences. In addition to uplink training in TDD systems, which is the focus of this paper, the pilot contamination problem would appear in other scenarios as well as it is fundamental to training with non-orthogonal pilots.

Next, we proceed to develop a new precoding method referred to as the multi-cell MMSE-based precoding.

V. MULTI-CELL MMSE-BASED PRECODING

In the previous section, we show that pilot contamination severely impacts the system performance by increasing the inter-cell interference. In particular, we show that the inter-cell interference grows like the intended signal with the number of antennas M at the base stations while using zero-forcing precoding. Therefore, in the presence of pilot contamination, in addition to frequency/time/pilot reuse schemes, it is crucial to account for inter-cell interference while designing a precoding method. Furthermore, since pilot contamination is originating from the non-orthogonal training sequences, it is important to account for the training sequence allocation while designing a precoding method. The approach of accounting for inter-cell interference while designing a precoding method is common, while the approach of accounting for the training sequence allocation is not. Again, the usual approach is to decouple the channel estimation and precoding completely. However, while using non-orthogonal pilots, this is not the right approach. These observations follow from our pilot contamination analysis in the previous section.

The precoding problem cannot be directly formulated as a joint optimization problem as different base stations have different received training signals. In other words, the problem is decentralized in nature. Therefore, one approach is to apply single-cell precoding methods. For example, since we assume orthogonal training sequences in every cell, we can perform zero-forcing on the users in every cell. The precoding matrix corresponding to this zero-forcing approach is given by

$$\mathbf{A}_l = \frac{\hat{\mathbf{G}}_{ll}^\dagger (\hat{\mathbf{G}}_{ll} \hat{\mathbf{G}}_{ll}^\dagger)^{-1}}{\sqrt{\text{tr} \left[(\hat{\mathbf{G}}_{ll} \hat{\mathbf{G}}_{ll}^\dagger)^{-1} \right]}}, \quad (17)$$

where $\hat{\mathbf{G}}_{ll} = \sqrt{p_f} \mathbf{D}_{ll}^{\frac{1}{2}} \hat{\mathbf{H}}_{ll}$. However, this zero-forcing precoding or other single-cell precoding methods do not account for the training sequence allocation, which is potentially the right approach to mitigate the pilot contamination problem. We explore this next.

In order to determine the precoding matrices, we formulate an optimization problem for each precoding matrix. Consider the j -th cell. The signal received by the users in this cell given by (4) is a function of all the precoding matrices (used at all the base stations). Therefore, the MMSE-based precoding methods for single-cell setting considered in [31]

does not extend (directly) to this setting. Let us consider the signal and interference terms corresponding to the base station in the l -th cell. Based on these terms, we formulate the following optimization problem to obtain the precoding matrix \mathbf{A}_l . We use the following notation: $\mathbf{F}_{jl} = \sqrt{p_f} \mathbf{D}_{jl}^{\frac{1}{2}} \mathbf{H}_{jl}$, $\hat{\mathbf{F}}_{jl} = \sqrt{p_f} \mathbf{D}_{jl}^{\frac{1}{2}} \hat{\mathbf{H}}_{jl}$ and $\tilde{\mathbf{F}}_{jl} = \mathbf{F}_{jl} - \hat{\mathbf{F}}_{jl}$ for all j and l . The optimization problem is:

$$\begin{aligned} \min_{\mathbf{A}_l, \alpha_l} \quad & \mathbb{E}_{\tilde{\mathbf{F}}_{jl}, \mathbf{z}_l, \mathbf{q}_l} \left[\|\alpha_l (\mathbf{F}_{ll} \mathbf{A}_l \mathbf{q}_l + \mathbf{z}_l) - \mathbf{q}_l\|^2 + \right. \\ & \left. \sum_{j \neq l} \|\alpha_l \gamma (\mathbf{F}_{jl} \mathbf{A}_l \mathbf{q}_l)\|^2 \right] \left| \hat{\mathbf{F}}_{jl} \right] \quad (18) \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}_l^\dagger \mathbf{A}_l\} = 1. \end{aligned}$$

This objective function is very intuitive. The objective function of the problem (18) consists of two parts: (i) the sum of squares of “errors” seen by the users in the l -th cell, and (ii) the sum of squares of interference seen by the users in all other cells. The parameter γ of the optimization problem “controls” the relative weights associated with these two parts. The real scalar parameter α_l is important as it “virtually” corresponds to the potential scaling that can be performed at the users. The optimal solution to the problem (18) denoted by \mathbf{A}_l^{opt} is the multi-cell MMSE-based precoding matrix.

Next, we obtain a closed-form expression for \mathbf{A}_l^{opt} . The following lemma is required later for obtaining the optimal solution to the problem (18).

Lemma 5: Consider the optimization problem (18). For all j and l ,

$$\mathbb{E} \left[\tilde{\mathbf{F}}_{jl}^\dagger \tilde{\mathbf{F}}_{jl} \left| \hat{\mathbf{F}}_{jl} \right. \right] = \delta_{jl} \mathbf{I}_M,$$

where

$$\delta_{jl} = p_f \text{tr} \left\{ \mathbf{D}_{jl} \left(\mathbf{I}_K + p_r \tau \mathbf{D}_{jl}^{\frac{1}{2}} \Psi_j^\dagger \mathbf{A}_{jl} \Psi_j \mathbf{D}_{jl}^{\frac{1}{2}} \right)^{-1} \right\}, \quad (19)$$

and

$$\mathbf{A}_{jl} = \left(\mathbf{I} + p_r \tau \sum_{i \neq j} \Psi_i \mathbf{D}_{il} \Psi_i^\dagger \right)^{-1}.$$

Proof: Let $\tilde{\mathbf{f}}_{jlm}$ denote the m -th column of $\tilde{\mathbf{F}}_{jl}$. Similarly, we define \mathbf{h}_{jlm} and $\hat{\mathbf{h}}_{jlm}$. From (3), we have

$$\begin{aligned} \tilde{\mathbf{f}}_{jlm} &= \sqrt{p_f} \mathbf{D}_{jl}^{\frac{1}{2}} (\mathbf{h}_{jlm} - \hat{\mathbf{h}}_{jlm}), \\ &= \sqrt{p_f} \mathbf{D}_{jl}^{\frac{1}{2}} \left(\mathbf{h}_{jlm} - \sqrt{p_r} \tau \mathbf{D}_{jl}^{\frac{1}{2}} \Psi_j^\dagger \mathbf{A}_l \mathbf{y}_{lm} \right), \end{aligned}$$

where

$$\mathbf{A}_l = \left(\mathbf{I} + p_r \tau \sum_{i=1}^L \Psi_i \mathbf{D}_{il} \Psi_i^\dagger \right)^{-1},$$

and \mathbf{y}_{lm} is given by (1). For given j and l , it is clear that $\left\{ \tilde{\mathbf{f}}_{jlm} \right\}_{m=1}^M$ is i.i.d. zero-mean \mathcal{CN} distributed. Hence,

$$\mathbb{E} \left[\tilde{\mathbf{F}}_{jl}^\dagger \tilde{\mathbf{F}}_{jl} \middle| \hat{\mathbf{F}}_{jl} \right] = \delta_{jl} \mathbf{I}_M \text{ where}$$

$$\begin{aligned} \delta_{jl} &= \mathbb{E} \left[\tilde{\mathbf{f}}_{jlm}^\dagger \tilde{\mathbf{f}}_{jlm} \middle| \hat{\mathbf{F}}_{jl} \right], \\ &= p_f \text{tr} \left\{ \mathbf{D}_{jl}^{\frac{1}{2}} \left(\mathbf{I}_K - \mathbb{E} \left[\hat{\mathbf{h}}_{jlm} \hat{\mathbf{h}}_{jlm}^\dagger \middle| \hat{\mathbf{F}}_{jl} \right] \right) \mathbf{D}_{jl}^{\frac{1}{2}} \right\}, \\ &= p_f \text{tr} \left\{ \mathbf{D}_{jl}^{\frac{1}{2}} \left(\mathbf{I}_K - p_r \tau \mathbf{D}_{jl}^{\frac{1}{2}} \boldsymbol{\Psi}_j^\dagger \boldsymbol{\Delta}_l \boldsymbol{\Psi}_j \mathbf{D}_{jl}^{\frac{1}{2}} \right) \mathbf{D}_{jl}^{\frac{1}{2}} \right\}, \\ &= p_f \text{tr} \left\{ \mathbf{D}_{jl}^{\frac{1}{2}} \left(\mathbf{I}_K + p_r \tau \mathbf{D}_{jl}^{\frac{1}{2}} \boldsymbol{\Psi}_j^\dagger \boldsymbol{\Lambda}_{jl} \boldsymbol{\Psi}_j \mathbf{D}_{jl}^{\frac{1}{2}} \right)^{-1} \mathbf{D}_{jl}^{\frac{1}{2}} \right\}. \end{aligned}$$

The last step follows from matrix inversion lemma. \blacksquare

The main result of this section is given by the following theorem. This theorem provides a closed-form expression for the multi-cell MMSE-based precoding matrix.

Theorem 6: The optimal solution to the problem (18) is

$$\mathbf{A}_l^{opt} = \frac{1}{\alpha_l^{opt}} \left(\hat{\mathbf{F}}_{ll}^\dagger \hat{\mathbf{F}}_{ll} + \gamma^2 \sum_{j \neq l} \hat{\mathbf{F}}_{jl}^\dagger \hat{\mathbf{F}}_{jl} + \eta \mathbf{I}_M \right)^{-1} \hat{\mathbf{F}}_{ll}^\dagger, \quad (20)$$

where

$$\eta = \delta_{ll} + \gamma^2 \sum_{j \neq l} \delta_{jl} + K,$$

δ_{jl} is given by (19) and α_l^{opt} satisfies $\text{tr}\{(\mathbf{A}_l^{opt})^\dagger \mathbf{A}_l^{opt}\} = 1$.

Proof: First, we simplify the objective function $J(\mathbf{A}_l, \alpha_l)$ of the problem (18) as follows:

$$\begin{aligned} J(\mathbf{A}_l, \alpha_l) &= \mathbb{E} \left[\|\alpha_l (\mathbf{F}_{ll} \mathbf{A}_l \mathbf{q}_l + \mathbf{z}_l) - \mathbf{q}_l\|^2 + \sum_{j \neq l} \|\alpha_l \gamma \mathbf{F}_{jl} \mathbf{A}_l \mathbf{q}_l\|^2 \middle| \hat{\mathbf{F}}_{jl} \right], \\ &= \mathbb{E} \left[\|(\alpha_l \mathbf{F}_{ll} \mathbf{A}_l - \mathbf{I}_K) \mathbf{q}_l\|^2 + \sum_{j \neq l} \|\alpha_l \gamma \mathbf{F}_{jl} \mathbf{A}_l \mathbf{q}_l\|^2 \middle| \hat{\mathbf{F}}_{jl} \right] + \alpha_l^2 K, \\ &= \text{tr} \left\{ \mathbb{E} \left[(\alpha_l \mathbf{F}_{ll} \mathbf{A}_l - \mathbf{I}_K)^\dagger (\alpha_l \mathbf{F}_{ll} \mathbf{A}_l - \mathbf{I}_K) + \sum_{j \neq l} \alpha_l^2 \gamma^2 \mathbf{A}_l^\dagger \mathbf{F}_{jl}^\dagger \mathbf{F}_{jl} \mathbf{A}_l \middle| \hat{\mathbf{F}}_{jl} \right] \right\} + \alpha_l^2 K, \\ &= \text{tr} \left\{ \alpha_l^2 \mathbf{A}_l^\dagger \mathbb{E} \left[\mathbf{F}_{ll}^\dagger \mathbf{F}_{ll} \middle| \hat{\mathbf{F}}_{ll} \right] \mathbf{A}_l - \alpha_l \mathbf{A}_l^\dagger \hat{\mathbf{F}}_{ll}^\dagger - \alpha_l \hat{\mathbf{F}}_{ll} \mathbf{A}_l + \sum_{j \neq l} \alpha_l^2 \gamma^2 \mathbf{A}_l^\dagger \mathbb{E} \left[\mathbf{F}_{jl}^\dagger \mathbf{F}_{jl} \middle| \hat{\mathbf{F}}_{jl} \right] \mathbf{A}_l \right\} + (\alpha_l^2 + 1) K, \\ &= \text{tr} \left\{ \alpha_l^2 \mathbf{A}_l^\dagger \left(\hat{\mathbf{F}}_{ll}^\dagger \hat{\mathbf{F}}_{ll} + \gamma^2 \sum_{j \neq l} \hat{\mathbf{F}}_{jl}^\dagger \hat{\mathbf{F}}_{jl} - \alpha_l \mathbf{A}_l^\dagger \hat{\mathbf{F}}_{ll}^\dagger - \alpha_l \hat{\mathbf{F}}_{ll} \mathbf{A}_l + \left(\delta_{ll} + \gamma^2 \sum_{j \neq l} \delta_{jl} \right) \mathbf{I}_M \right) \mathbf{A}_l \right\} + (\alpha_l^2 + 1) K. \end{aligned}$$

The last step follows from Lemma 5.

Now, consider the Lagrangian formulation

$$L(\mathbf{A}_l, \alpha_l, \lambda) = J(\mathbf{A}_l, \alpha_l) + \lambda \left(\text{tr} \left\{ \mathbf{A}_l^\dagger \mathbf{A}_l \right\} - 1 \right)$$

for the problem (18). Let

$$\mathbf{R} = \hat{\mathbf{F}}_{ll}^\dagger \hat{\mathbf{F}}_{ll} + \gamma^2 \sum_{j \neq l} \hat{\mathbf{F}}_{jl}^\dagger \hat{\mathbf{F}}_{jl} + \left(\delta_{ll} + \gamma^2 \sum_{j \neq l} \delta_{jl} + \frac{\lambda}{\alpha_l^2} \right) \mathbf{I}_M,$$

$\mathbf{U} = \alpha_l \mathbf{R}^{\frac{1}{2}} \mathbf{A}_l$ and $\mathbf{V} = \mathbf{R}^{-\frac{1}{2}} \hat{\mathbf{F}}_{ll}^\dagger$. We have

$$L(\mathbf{A}_l, \alpha_l, \lambda) = \|\mathbf{U} - \mathbf{V}\|^2 - \text{tr} \left\{ \hat{\mathbf{F}}_{ll} \mathbf{R}^{-1} \hat{\mathbf{F}}_{ll}^\dagger \right\} + (\alpha_l^2 + 1) K - \lambda. \quad (21)$$

This can be easily verified by expanding the right hand side. It is clear from (21) that, for any given α_l and λ , $L(\mathbf{A}_l, \alpha_l, \lambda)$ is minimized if and only if $\mathbf{U} = \mathbf{V}$. Hence, we obtain

$$\mathbf{A}_l^{opt} = \frac{1}{\alpha_l} \mathbf{R}^{-1} \hat{\mathbf{F}}_{ll}^\dagger. \quad (22)$$

Let $L(\alpha_l, \lambda) = L(\mathbf{A}_l^{opt}, \alpha_l, \lambda)$. Now, we have

$$L(\alpha_l, \lambda) = -\text{tr} \left\{ \hat{\mathbf{F}}_{ll} \mathbf{R}^{-1} \hat{\mathbf{F}}_{ll}^\dagger \right\} + (\alpha_l^2 + 1) K - \lambda. \quad (23)$$

Note that $\hat{\mathbf{F}}_{ll}^\dagger \hat{\mathbf{F}}_{ll} + \gamma^2 \sum_{j \neq l} \hat{\mathbf{F}}_{jl}^\dagger \hat{\mathbf{F}}_{jl}$ can be factorized in the form $\mathbf{S}^\dagger \text{diag}\{[c_1 \ c_2 \ \dots \ c_M]\} \mathbf{S}$ where $\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}_M$. Let $\delta = \delta_{ll} + \gamma^2 \sum_{j \neq l} \delta_{jl}$. Therefore,

$$\begin{aligned} \mathbf{R}^{-1} &= \left(\mathbf{S}^\dagger \text{diag}\{[c_1 \ c_2 \ \dots \ c_M]\} \mathbf{S} + \left(\delta + \frac{\lambda}{\alpha_l^2} \right) \mathbf{I}_M \right)^{-1}, \\ &= \left(\mathbf{S}^\dagger \text{diag} \left\{ \left[c_1 + \delta + \frac{\lambda}{\alpha_l^2} \ \dots \ c_M + \delta + \frac{\lambda}{\alpha_l^2} \right] \right\} \mathbf{S} \right)^{-1}, \\ &= \mathbf{S}^\dagger \text{diag} \left\{ \left[\left(c_1 + \delta + \frac{\lambda}{\alpha_l^2} \right)^{-1} \ \dots \ \left(c_M + \delta + \frac{\lambda}{\alpha_l^2} \right)^{-1} \right] \right\} \mathbf{S}. \end{aligned} \quad (24)$$

Substituting (24) in (23), we get

$$L(\alpha_l, \lambda) = -\sum_{m=1}^M \frac{d_m}{c_m + \delta + \frac{\lambda}{\alpha_l^2}} + (\alpha_l^2 + 1) K - \lambda, \quad (25)$$

where d_m is the (m, m) -th entry of $\mathbf{S} \hat{\mathbf{F}}_{ll}^\dagger \hat{\mathbf{F}}_{ll} \mathbf{S}^\dagger$. Consider the equations obtained by differentiating (25) w.r.t. α_l and λ and equating to zero:

$$\sum_{m=1}^M \frac{d_m}{\left(c_m + \delta + \frac{\lambda}{\alpha_l^2} \right)^2 \alpha_l^2} = 1, \quad (26)$$

$$-\sum_{m=1}^M \frac{d_m}{\left(c_m + \delta + \frac{\lambda}{\alpha_l^2} \right)^2 \alpha_l^3} + 2\alpha_l K = 0. \quad (27)$$

Substituting (26) in (27), we get

$$\frac{\lambda}{\alpha_l^2} = K. \quad (28)$$

Combining the results in (22), (24), and (28), we get (20), where α_l^{opt} is such that $\|\mathbf{A}_l^{opt}\|_F^2 = 1$. \blacksquare

The precoding described above is primarily suited for maximizing the minimum of the rates achieved by all the users. This is because all users are treated equally without differentiating them based on the channels. Therefore, when the performance metric of interest is sum rate, this precoding can be combined with power control, scheduling, and other similar techniques to enhance the net performance. Since our main concern is the inter-cell interference resulting from pilot contamination, and to avoid too complicated systems, we do not use that possibility in this paper. In the next section, all numerical results and comparisons are performed without power control.

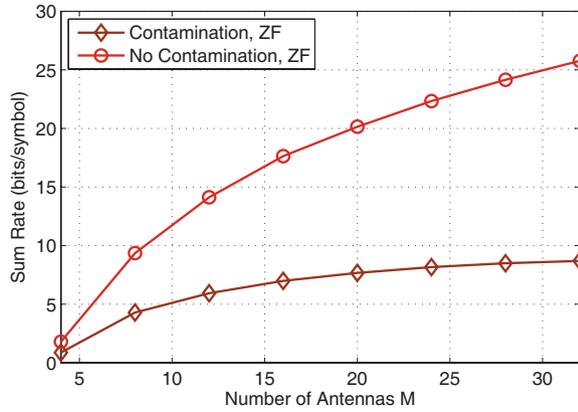


Fig. 3. Comparison of sum rate with and without pilot contamination for the two-cell system; sum rate saturates in the presence of contamination.

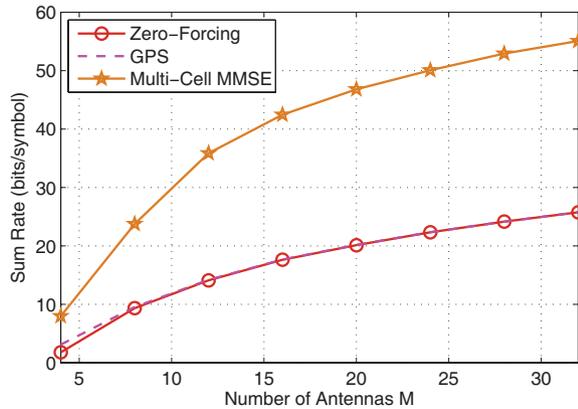


Fig. 4. Comparison of different schemes for the two-cell system; Multi-cell MMSE performs inter-cell interference mitigation leading to improved sum rate.

VI. NUMERICAL RESULTS

Multi-Cell MMSE precoding denotes the new precoding method developed given in (20) with parameter γ set to unity.⁶ ZF precoding denotes the popular zero-forcing precoding given in (17). GPS denotes the single-cell precoding method suggested in [31], which is a special case of the precoding given in (20) with parameter γ set to zero. In all the plots, we average the performance metric over 10^3 i.i.d. channel realizations.

A. Two-cell System

We consider a basic two-cell example to understand the impact of pilot contamination on the total system throughput (sum rate). In particular, we consider a two-cell system with $p_f = 20$ dB, $p_r = 10$ dB and $K = 4$ users in both cells. We set all the direct gains to 1 and all cross gains to a , i.e., for all k , $\beta_{jlk} = 1$ if $j = l$ and $\beta_{jlk} = a$ if $j \neq l$. We capture the main observations using two plots.

First, in Figure 3, we use cross gain value of $a = 0.8$ and plot sum rate versus number of antennas M with zero-forcing for training lengths of $\tau = 4$ (scenario with pilot contamination) and $\tau = 8$ (scenario without contamination). With $\tau = 4$, the orthogonal training sequences used in the 1-st

⁶In our simulations, we observe that the performance is not very sensitive to the value of this parameter.

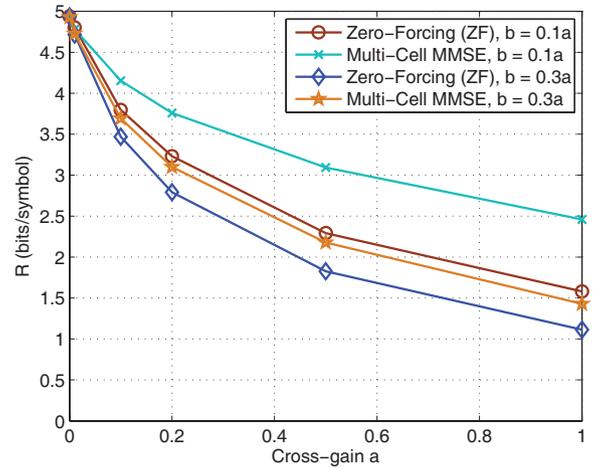


Fig. 5. Comparison of zero-forcing and multi-cell MMSE precoding methods; a and b correspond to different cross-gains and R denotes the minimum rate achieved by all users.

is reused in the 2-nd cell. With $\tau = 8$, all users in the system are given orthogonal pilots. In Figure 3, we can clearly observe the saturation of total throughput in the presence of pilot contamination. Note that both scenarios deal with interference. In many practical systems, we cannot necessarily keep τ large ($\tau = 8$ in this example) as the coherence interval is typically very short, which requires using small τ .

In the presence of pilot contamination ($\tau = 4$), neither GPS nor multi-cell MMSE provide noticeable improvement in total throughput (except for small values of M). This is not surprising as pilot contamination is a very fundamental problem in this example.⁷ For $M = 4$ to $M = 12$, the improvement using GPS and multi-cell MMSE is reasonable. This improvement results from using MMSE instead of zero-forcing, and hence both GPS and multi-cell MMSE provide very similar performance.⁸

In the absence of pilot contamination ($\tau = 8$), the story is different as shown in Figure 4. Multi-cell MMSE outperforms both the single-cell schemes (GPS and zero-forcing) by a huge margin. This is possible as multi-cell MMSE is capable of performing efficient inter-cell interference mitigation. Note that this inter-cell interference mitigation is achieved using the channel estimates obtained in a distributed manner. This clearly shows the advantage of multi-cell MMSE. However, this example only focuses on the scenario without pilot contamination. Therefore, a natural question is whether multi-cell MMSE can provide throughput gains in a mixed scenario, which is addressed next.

B. Multi-cell System

We consider a multi-cell system with $L = 4$ cells, $M = 8$ antennas at all base stations, $K = 2$ users in every cell and training length of $\tau = 4$. We consider $p_f = 20$ dB and $p_r = 10$ dB. Orthogonal training sequences are collectively used within the 1-st and 2-nd cells. The training sequences used in the 1-st (2-nd) cell are reused in the 3-rd (4-th) cell. Thus, we model a scenario where training sequences are reused. We keep the

⁷We can still overcome pilot contamination using frequency/time reuse.

⁸This plot is not provided as it is not the main focus of this paper.

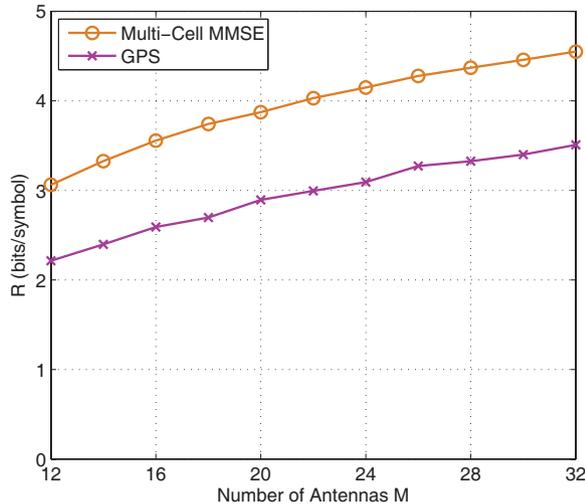


Fig. 6. Comparison of GPS and multi-cell MMSE precoding methods with $a = 0.8$ and $b = 0.1a$; R denotes the minimum rate achieved by all users.

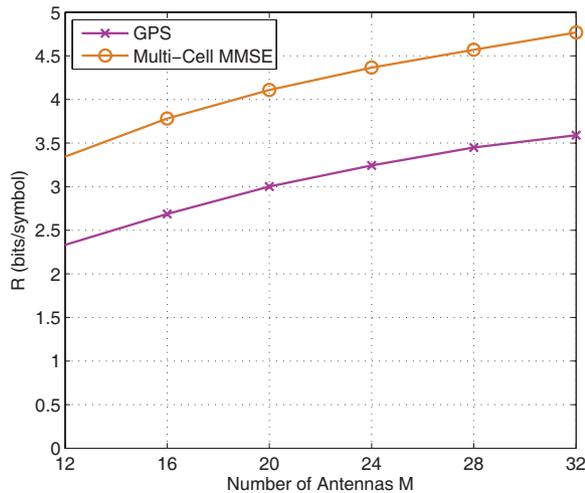


Fig. 7. Comparison of GPS and multi-cell MMSE precoding methods with $a = 0.8$ and $b = 0.1a$; Pilots in cells 1 and 3 (2 and 4) are not reused but rotated by 45 degrees.

propagation factors as follows: for all k , $\beta_{jlk} = 1$ if $j = l$, $\beta_{jlk} = a$ if $(j, l) \in \{(1, 2), (2, 1), (3, 4), (4, 3)\}$, and $\beta_{jlk} = b$ for all other values of j and l . “Frequency reuse” is handled semi-quantitatively by adjusting the cross-gains.

Another performance metric of interest is the minimum rate achieved by all users denoted by $R = \min_{jk} R_{jk}$. In Figure 5, we plot the performance of ZF and multi-cell MMSE precoding methods for different values of a and b . We observe significant advantage of using multi-cell MMSE precoding for wide range of values of a and b . In Figure 6, we plot the performance of GPS and multi-cell MMSE precoding methods as a function of the number of antennas M . We also consider the scenario when pilots in cells 1 and 3 (2 and 4) are not reused but rotated by 45 degrees. This comparison is given in Figure 7. In both cases, we observe significant advantages in using our multi-cell MMSE precoding. Thus, the multi-cell MMSE scheme is capable of handling different scenarios as it utilizes training sequences for precoder design to mitigate inter-cell interference even in the presence of pilot contamination. Note that in both cases the performance of

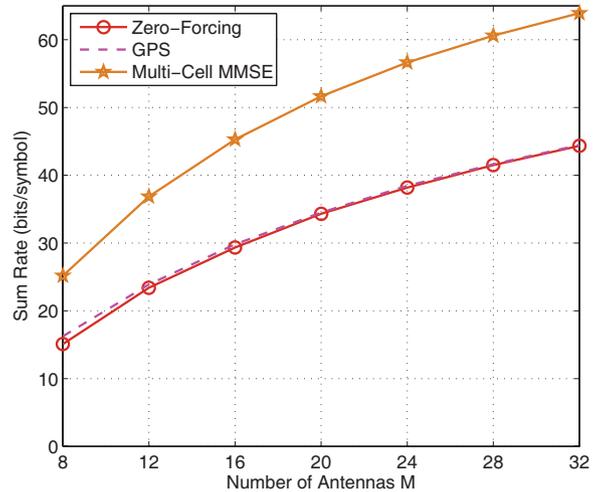


Fig. 8. Comparison of different precoding methods with $a = 0.8$, $b = 0.1a$, $\tau = 8$; Each cell has $K = 4$ users; Sum rate corresponds to total sum throughput of all $L = 4$ cells.

ZF and GPS are almost indistinguishable, and hence we have omitted ZF in these plots.

Finally, we consider a system with reused pilots, $K = 4$ users in every cell and training length of $\tau = 8$. In Figure 8, we plot the total sum throughput of different precoding methods as a function of the number of antennas M . In summary, all the numerical results show that the new multi-cell MMSE precoding offers significant performance gains.

VII. CONCLUSION

In this paper, we characterize the impact of corrupted channel estimates caused by pilot contamination in TDD systems. When non-orthogonal training sequences are assigned to users, the precoding matrix used at a (multiple antenna) base station becomes correlated with the channel to users in other cells (referred to as pilot contamination). For a special setting that captures pilot contamination, we obtain a closed-form expressions for the achievable rates. Using these analytical expressions, we show that, in the presence of pilot contamination, rates achieved by users saturate with the number of base station antennas. We conclude that appropriate frequency/time reuse techniques have to be employed to overcome this saturation effect. The fact that pilot contamination hasn’t surfaced in FDD studies suggests that researchers are assuming partial CSI with independently corrupted noise, and are not fully incorporating the impact of channel estimation.

Next, we develop a multi-cell MMSE-based precoding that depends on the set of training sequences assigned to the users. We obtain this precoding as the solution to an optimization problem whose objective function consists of two parts: (i) the mean-square error of signals received at the users in the same cell, and (ii) the mean-square interference caused at the users in other cells. We show that this precoding method reduce both intra-cell interference and inter-cell interference, and thus is similar in spirit to existing joint-cell precoding techniques. The primary differences between joint-cell precoding and our approach are that a.) our approach is distributed in nature and, b.) we explicitly take into account the set of training sequences assigned to the users. Through numerical results, we show

that our method outperforms popular single-cell precoding methods.

REFERENCES

- [1] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE International Symposium on Information Theory*, July 2009, pp. 2184–2188.
- [2] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, 2010.
- [3] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, pp. 506–522, Feb. 2005.
- [4] P. Ding, D. J. Love, and M. D. Zoltowski, "On the sum rate of channel subspace feedback for multi-antenna broadcast channels," in *Proc. IEEE Global Telecommunications Conference*, Nov. 2005, pp. 2699–2703.
- [5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.
- [6] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1478–1491, 2007.
- [7] A. Ashikhmin and R. Gopalan, "Grassmannian packings for efficient quantization in MIMO broadcast systems," in *Proc. IEEE International Symposium on Information Theory*, June 2007, pp. 1811–1815.
- [8] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3879–3891, 2007.
- [9] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conference on Signals, Systems and Computers*, Oct./Nov. 2006, pp. 359–363.
- [10] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," in *Proc. IEEE International Conference on Communications*, May 2008, pp. 4100–4105.
- [11] B. Vojčić and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.
- [12] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, pp. 1268–1278, 2006.
- [13] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. ISSPA*, 2005.
- [14] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [15] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEEE Proc. Commun.*, vol. 153, pp. 548–555, Aug. 2006.
- [16] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: overcoming intercell interference in indoor wireless systems," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2007, pp. 83–87.
- [17] S. Shamai (Shitz), O. Somekh, and B. M. Zaidel, "Multi-cell communications: an information theoretic perspective," in *Joint Workshop on Communications and Coding*, Oct. 2004.
- [18] H. Zhang and H. Dai, "Co-channel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *European J. Wireless Commun. and Networking*, pp. 222–235, 4th quarter 2004.
- [19] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, 2009.
- [20] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1707, July 2003.
- [21] 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, pp. 1912–1921, Aug. 2003.
- [22] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.
- [23] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.
- [24] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.
- [25] F. Boccardi, F. Tosato, and G. Caire, "Precoding schemes for the MIMO-GBC," in *Int. Zurich Seminar on Communications*, Feb. 2006.
- [26] M. Airy, S. Bhadra, R. W. Heath, Jr., and S. Shakkottai, "Transmit precoding for the multiple antenna broadcast channel," in *Proc. Vehicular Technology Conference*, vol. 3, 2006, pp. 1396–1400.
- [27] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multiuser communication part II: perturbation," *IEEE Trans. Commun.*, vol. 53, pp. 537–544, Jan. 2005.
- [28] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna broadcast channels with linear preprocessing," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2338–2342, Sep. 2006.
- [29] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, pp. 3658–3663, Sep. 2006.
- [30] M. Joham, P. M. Castro, L. Castedo, and W. Utschick, "MMSE optimal feedback of correlated CSI for multi-user precoding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008, pp. 3129–3132.
- [31] K. S. Gomadam, H. C. Papadopoulos, and C.-E. W. Sundberg, "Techniques for multi-user MIMO with two-way training," in *Proc. IEEE International Conference on Communications*, May 2008, pp. 3360–3366.
- [32] 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, June 2010.
- [33] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [34] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.



Jubin Jose is currently a Ph.D. candidate in Electrical and Computer Engineering at The University of Texas at Austin (UT Austin). He received his B.Tech. degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras in 2006 and M.S. degree in Electrical and Computer Engineering from UT Austin in 2008. His industry experience includes positions at NEC labs, Princeton, NJ, Qualcomm Corporate R&D, San Diego, CA and Alcatel-Lucent Bell Labs, Murray Hill, NJ. His current research interests lie in the broad

area of communication networks with emphasis on wireless communication, information theory and network optimization. He received the WNCG student leadership award in 2011.



Alexei Ashikhmin is a member of technical staff in the Communications and Statistical Sciences Department, Bell Laboratories, Murray Hill, New Jersey. He received his Ph.D. degree in Electrical Engineering from the Institute of Information Transmission Problems, Russian Academy of Science, Moscow, Russia, in 1994. In 1995 and 1996, he was with the Mathematics and Computer Science Department, Delft University of Technology, The Netherlands. From 1997 to 1999, he was a Postdoctoral Fellow at Modeling, Algorithms, and Informatics Group of Los Alamos National Laboratory. Since 1999, he has been with Bell Laboratories. In 2005–2007, Alexei Ashikhmin was an adjunct professor at Columbia University. He taught courses "Error Correcting Codes" and "Communications Theory."

Alexei Ashikhmin's research interests include communications theory, the theory of error correcting codes, and classical and quantum information theory. From 2003 to 2006, Dr. Ashikhmin served as an Associate Editor for IEEE Transactions on Information Theory, and is serving again, beginning this year. In 2002, Dr. Ashikhmin received Bell Laboratories President's Gold Award for breakthrough research resulting in the ability to deliver unprecedented wireless bit-rates. In 2005, Dr. Ashikhmin was honored by the IEEE Communications S.O. Rice Prize Paper Award for work on LDPC codes for information transmission with multiple antennas. In 2010, Dr. Ashikhmin received Bell Labs Team Award for Crosstalk Cancellation in DSL systems. Dr. Ashikhmin is a senior member of IEEE Information Theory Society.



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 energy efficiency and spectral efficiency.

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) and for the IEEE TRANSACTIONS ON INFORMATION THEORY

Special Issue on Space-Time Transmission, Reception, Coding, and Signal Design (Oct. 2003). 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.



Sriram Vishwanath received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Madras, India, in 1998, the M.S. degree in electrical engineering from California Institute of Technology (Caltech), Pasadena, in 1999, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2003. His industry experience includes positions at Lucent Bell Labs and National Semiconductor Corporation. He is currently an Associate Professor in the Department of Electrical and Computer Engineering,

University of Texas, Austin. His research interests include network information theory, wireless systems, and mobile systems. Dr. Vishwanath received the NSF CAREER award in 2005 and the ARO Young Investigator Award in 2008. He is the co-recipient of the 2005 IEEE Joint Information Theory Society and Communications Society best paper award. He has served as the general chair of the IEEE Wireless Network Coding Conference (WiNC) in 2010, the local arrangements chair of ISIT 2010, and the Guest Editor-in-Chief of IEEE TRANSACTIONS ON INFORMATION THEORY Special Issue on Interference Networks.