Power-Controlled Reverse Channel Training in a Multiuser TDD-MIMO Spatial Multiplexing System With CSIR

B. N. Bharath and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—This paper considers the design of a power-controlled *reverse* channel training (RCT) scheme for spatial multiplexing (SM)-based data transmission along the dominant modes of the channel in a time-division duplex (TDD) multiple-input and multiple-output (MIMO) system, when channel knowledge is available at the receiver. A channel-dependent power-controlled RCT scheme is proposed, using which the *transmitter* estimates the beamforming (BF) vectors required for the forward-link SM data transmission. Tight approximate expressions for 1) the mean square error (MSE) in the estimate of the BF vectors, and 2) a capacity lower bound (CLB) for an SM system, are derived and used to optimize the parameters of the training sequence. Moreover, an extension of the channel-dependent training scheme and the data rate analysis to a multiuser scenario with M user terminals is presented. For the single-mode BF system, a closed-form expression for an upper bound on the average sum data rate is derived, which is shown to scale as $((L_c - L_{B,\tau})/L_c) \log\log M$ asymptotically in M, where L_c and $L_{B,\tau}$ are the channel coherence time and training duration, respectively. The significant performance gain offered by the proposed training sequence over the conventional constant-power orthogonal RCT sequence is demonstrated using Monte Carlo simulations.**

***Index Terms*—Capacity lower bound (CLB), channel estimation, mean square error (MSE), reciprocal multiple-input–multiple-output (MIMO) system, spatial multiplexing (SM), training sequence design.**

I. Introduction

The use of multiple antennas is one of the promising technologies for multiuser wireless communication (e.g., IEEE 802.16a, IEEE 802.20, and 4G protocol candidates), as it offers significant benefits in terms of reliability and throughput. These benefits are realizable only when both the transmitter [base station ($\mathbf{BS}$)] and receivers [user terminals ($\mathbf{UT}$s)] have accurate and up-to-date channel state information (CSI). Thus, one of the important problems in designing multiple-antenna systems is the fast and accurate acquisition of CSI both at the $\mathbf{BS}$ and $\mathbf{UT}$s. CSI can be obtained at the $\mathbf{UT}$s simultaneously, by sending a known training sequence in the forward link from the BS [2]. In time-division duplex (TDD) systems, exploiting the reciprocity of the channel, CSI at the transmitter (CSIT) can be acquired by sending a known training signal in the reverse link, which is also known as reverse channel training (RCT). However, as the number of transmit antennas and/or the number of users becomes large, the overhead due to training can become prohibitive, particularly in vehicular or mobile communications, where the channel is relatively fast varying, since the training duration is proportional to the sum of the number of antennas at all the $\mathbf{UT}$s. When CSI is available at the receiver (CSIR), one can potentially exploit it to design the RCT sequence and selectively feedback only the required part of the CSI to the $\mathbf{BS}$. This could result in faster and/or more accurate acquisition of the CSI at the $\mathbf{BS}$, leading to an improvement in the effective downlink data rate; this is the focus of this paper.

The main body of the existing literature on CSIT acquisition in single- and multiuser TDD systems focuses on orthogonal RCT [3]–[7], where an orthogonal training sequence, such as the scaled identity matrix, is employed. The method employed in [8] and [9] to acquire CSIT is to feedback a scaled version of the received forward-link training signal, from which the transmitter estimates the entire channel matrix. Although this outperforms orthogonal RCT, it has the disadvantage that the transmitter estimates the entire channel matrix, which is not required for certain types of data transmission schemes such as beamforming (BF) or spatial multiplexing (SM) along the dominant modes of the channel. Data-aided blind estimation of the dominant BF vector is proposed in [10]–[12]. Reference [13] proposes a two-stage protocol consisting of conventional RCT followed by quantized CSI transmission in the reverse link. In our past work, we have studied the design and performance of channel-dependent RCT in [14]–[17]. The structure of the RCT scheme we consider here is different from the past work, and it allows for both spatial and temporal allocation of training power. Due to this, the design considerations and the corresponding performance analysis are also different. A channel-dependent RCT scheme was studied in [18], in a single-input–multiple-output channel, and the fraction of energy spent for training was optimized with respect to an approximate expression for the forward-link data signal-to-noise ratio (SNR).

In this paper, we consider an SM system with equal power allocation across m dominant modes of the channel during data

Manuscript received January 2, 2013; revised March 7, 2013 and April 17, 2013; accepted May 24, 2013. Date of publication June 7, 2013; date of current version November 6, 2013. This work was supported in part by research projects funded by the Aerospace Network Research Consortium and the India-UK Advanced Technology Center (IU-ATC). A portion of this paper appeared in [1]. The review of this paper was coordinated by Prof. Y. Zhou.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: bharath@ece.iisc.ernet.in; cmurthy@ece.iisc.ernet.in).

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

Digital Object Identifier 10.1109/TVT.2013.2266923

transmission with perfect CSIR and noisy CSIT obtained via RCT. Equal power allocation across modes is known to be nearly optimal for all but low data SNR [19]. In addition, as shown in [20], with perfect CSIT, using a fixed value of m based on the average SNR and equal power allocation across the m modes is only marginally inferior, in terms of the ergodic capacity, compared with optimally adapting m based on channel instantiation, for Rayleigh fading channels. However, adapting m based on the CSI entails the additional overhead of feeding back the value of m to the $\mathbf{BS}$. Hence, in this paper, we assume that m is predetermined and fixed for all channel instantiations and is known to both the $\mathbf{UT}$ and the $\mathbf{BS}$. The perfect CSIR assumption is required for analytical tractability. It also helps to isolate the effect of estimation errors in the RCT on the performance. This assumption is common in studies that focus on the achievable data rate or outage probability [5], [7], [13], [20]–[23]. Note that our proposed RCT scheme is applicable in practical systems, even when the CSIR is imperfect.

In this paper, we consider SM-based data transmission over $m < n_A$ dominant modes of the channel, where n_A is the number of antennas at the $\mathbf{BS}$/transmitter. The following are our main contributions.

- *Proposed RCT*: We propose an RCT that allows the $\mathbf{BS}$ to directly estimate the dominant eigenmodes of the channel required for data transmission. Further, the proposed RCT allows one to allot the power both spatially (across modes) and temporally (across time), while satisfying an average training power constraint. Both the spatial power allocation matrix D and the temporal power control parameter $\phi_c > 0$ are optimized using the following two performance metrics: 1) the mean square error (MSE) in the estimated BF matrix at the transmitter; and 2) a capacity lower bound (CLB) on the downlink data transmission (see Section III).
- *Optimal RCT with approximate MSE as a metric*: With the approximate MSE in the estimated singular vectors as the performance metric, we obtain an analytical solution for the optimal D and ϕ_c, as a function of the channel singular values. For example, in a 3×4 multiple-input–multiple-output (MIMO) system, the proposed training scheme offers an improvement of over 15 dB in the training power required to achieve the same MSE, compared with the orthogonal RCT (see Section III-A).
- *Optimal RCT with approximate CLB as a metric*: Here, we analytically optimize spatio-temporal power allocation of the RCT scheme to maximize the approximate CLB. In the 3×4 example previously mentioned, using the optimal D outperforms using equal spatial power allocation by approximately 1 bit per use of the channel at around 16 dB of training and data power,[1] whereas temporal power allocation offers only a marginal improvement in the data rate over equal temporal power allocation (see Section III-B).
- *Multiuser case*: We extend the proposed channel-dependent RCT to a multiuser downlink scenario with M UTs. We use the approximate CLB as the metric for user scheduling. In the case of BF with $m = 1$ mode, we show that an upper bound on the sum data rate scales as $(L_c - L_{B,\tau}/L_c) \log\log M$ asymptotically in the number of users M, where L_c and $L_{B,\tau}$ are the coherence time of the channel and the training duration, respectively, in the number of symbols (see Section IV).

Our proposed RCT scheme can lead to significant performance improvements over orthogonal RCT in terms of both the achievable data rate and the MSE in BF vector estimation, particularly when the channel is fast varying. Moreover, from a system designer's point of view, it is useful to know that spatial allocation of the available training power is much more beneficial than temporal allocation, as will be illustrated through simulations in Section V.

We use the following notations. Capital letters and small boldface letters are used for matrices and vectors, respectively. $\mathbb{E}[\cdot]$ denotes the expected value of $[\cdot]$. We use $\mathbb{E}_{|A}[\cdot]$ to denote the expected value of $[\cdot]$ conditioned on A. $(\cdot)^H$ denotes the transpose-conjugate (Hermitian), and $|\cdot|$ denotes the determinant or the absolute value, depending on the context. I_m is the $m \times m$ identity matrix, and $I_{m\times n}$, with $n \le m$, represents the first n columns of the $m \times m$ identity matrix. $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius norm and the Euclidean norm, respectively. The real and imaginary parts of complex number c are represented as $\Re(c)$ and $\Im(c)$, respectively. For $a \in \mathbb{R}$, we use $(a)^+$ to mean $\max\{0, a\}$. We use $\mathbf{x} = \mathcal{O}(\mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ to mean that the entries of $\mathbf{x}$ are less than the corresponding entries of $c\mathbf{y}$ for some $0 < c < \infty$.

II. System Model

The system model consists of a single-cell multiuser system with a BS, which is denoted by $\mathbf{BS}$, and M active UTs, which are denoted by $\mathbf{UT}_1, \ldots, \mathbf{UT}_M$. The $\mathbf{BS}$ has n_A antennas, and each $\mathbf{UT}$ has n_B antennas. Denote the MIMO channel from $\mathbf{BS}$ to $\mathbf{UT}_k$ by $H_k \in \mathbb{C}^{n_B \times n_A}$. Let $H_k = U^{(k)}\Sigma^{(k)}(V^{(k)})^H$ be the singular value decomposition (SVD) of H_k, where the diagonal entries of $\Sigma^{(k)} \in \mathbb{R}^{n_B \times n_A}$, which are denoted by $\sigma_{1,k}, \ldots, \sigma_{n,k}$, are the singular values of H_k, with $n \triangleq \mathrm{rank}(H_k)$, which equals $\min(n_A, n_B)$ almost surely. Moreover, $U^{(k)} \in \mathbb{C}^{n_B \times n_B}$ and $V^{(k)} \in \mathbb{C}^{n_A \times n_A}$ are unitary matrices whose columns are the eigenvectors of $H_k H_k^H$ and $H_k^H H_k$, respectively. The channel is assumed to remain constant for a frame of duration that is equal to the channel coherence time of L_c symbols, and evolve in an independent and identically distributed (i.i.d.) fashion from frame to frame. We assume a TDD mode of operation with perfect reciprocity [10], [24]–[26], and thus, without loss of generality, the reverse-link channel of the kth user is H_k^H (see [10]). The transmission protocol consists of the following three phases.

- **Phase I**: This phase consists of downlink training followed by user scheduling. The downlink training is performed by sending a known pilot sequence from the $\mathbf{BS}$ to all the $\mathbf{UT}$s. Using this, each $\mathbf{UT}$ computes an estimate of their respective channels. Here, we assume that the resulting estimate is error-free, as in, for example, [22] and [23]. This facilitates the derivation of a CLB and its analytically tractable tight approximation in closed form,

[1] Signal power is expressed in a decibel scale relative to noise power in the signal bandwidth.

with respect to which the RCT can be optimized. We consider a scheduling scheme where each user computes a metric (see Section IV) as a function of its CSI. The user with the highest metric is scheduled for data transmission by the $\mathbf{BS}$. The selection of the user with the highest metric can be efficiently implemented using decentralized algorithms such as splitting [27], [28] and timer-based schemes [29], which incurs very low overhead[2] in terms of power and delay. We assume that one of these schemes is used to pick the best user, and we ignore the overhead involved in user selection. In this setting, we focus on the problem of RCT sequence design to convey the CSI of the selected user to the $\mathbf{BS}$.

- **Phase II**: In this phase, the scheduled UT, for example, $\mathbf{UT}_k$, transmits training sequence $X_{B,\tau}^{(k)} \in \mathbb{C}^{n_B \times L_{B,\tau}}$ of duration $L_{B,\tau}$ symbols in the uplink direction. Here, subscript τ stands for "training." The baseband equivalent of the received training signal at the $\mathbf{BS}$, denoted by $Y_{A,\tau} \in \mathbb{C}^{n_A \times L_{B,\tau}}$, is given by

$$\textbf{Reverse-link training}: \quad Y_{A,\tau} = H_k^H X_{B,\tau}^{(k)} + W_{A,\tau}. \tag{1}$$

  The entries of the noise $W_{A,\tau} \in \mathbb{C}^{n_A \times L_{B,\tau}}$ are assumed to be i.i.d. complex circularly symmetric standard Gaussian distributed, which are denoted by $\mathcal{CN}(0,1)$. From $Y_{A,\tau}$, the $\mathbf{BS}$ computes an estimate of $V_m^{(k)}$, the first m columns of matrix $V^{(k)}$, which is subsequently used for downlink data transmission over the dominant modes of the channel. Denote the estimate of $V_m^{(k)}$ by $\hat{V}_m^{(k)}$.
- **Phase III**: This phase consists of data transmission from the $\mathbf{BS}$ to the scheduled user. For the data transmission scheme, we assume SM of data with equal power allocation. The $\mathbf{BS}$ sends $m \geq 1$ i.i.d. data streams, $\mathbf{x}_{A,d}^{(k)} \in \mathbb{C}^m$, multiplied by the estimate of the m dominant right singular vectors of H_k obtained in **Phase II** [2], [20]. Here, the subscript d stands for "data." The corresponding received signal at $\mathbf{UT}_k$, denoted by $\mathbf{y}_{B,d}^{(k)} \in \mathbb{C}^{n_B \times 1}$, is given by

$$\textbf{Forward-link data}: \quad \mathbf{y}_{B,d}^{(k)} = \rho_m H_k \hat{V}_m^{(k)} \mathbf{x}_{A,d}^{(k)} + \mathbf{w}_{A,d}^{(k)}. \tag{2}$$

  In the above equation, $\rho_m \triangleq \sqrt{P_{A,d}/m}$, $\mathbb{E}[\mathbf{x}_{A,d}^{(k)} (\mathbf{x}_{A,d}^{(k)})^H] = I_m$, and $(\hat{V}_m^{(k)})^H (\hat{V}_m^{(k)}) = I_m$ ensure that the data signal satisfies an average power constraint of $P_{A,d}$. The entries of the noise vector $\mathbf{w}_{B,d}^{(k)} \in \mathbb{C}^{n_B}$ are assumed to be i.i.d. $\mathcal{CN}(0,1)$. Note that, $m = 1$ corresponds to pure-BF based data transmission using the estimated dominant right singular vector of the channel. Thus, the scheme considered here encompasses BF as a special case.

With the above transmission and scheduling scheme, in the following section, we consider the problem of designing $X_{B,\tau}$, with the following two performance metrics: 1) MSE in $\hat{V}_m^{(k)}$, and 2) an achievable downlink data rate.

[2] In fact, the time overhead in best user selection is bounded irrespective of the number of users [29].

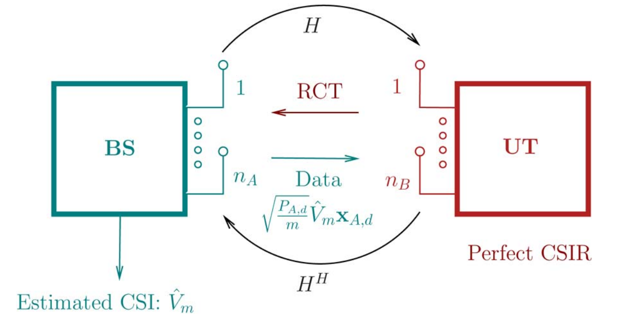

Fig. 1. MIMO $n_B \times n_A$ SM system, showing the RCT and the forward-link data transmission.

III. Design and Optimization of Reverse Channel Training for the Scheduled User

In this section, we assume that the best user has been scheduled for data transmission using the procedure in Phase I of the protocol described in the previous section, and focus on the design of the RCT sequence to efficiently convey the CSI to the $\mathbf{BS}$ (see Fig. 1). Since the index of the scheduled user does not directly enter the expressions, for ease of presentation, in this section, we drop user index k. For SM with equal power allocation, the $\mathbf{BS}$ requires knowledge of $V_m \triangleq [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m]$, the first m columns of the right singular matrix V of the channel H. We propose the following *channel-dependent power-controlled RCT sequence* that enables the $\mathbf{BS}$ to directly estimate V_m:

$$X_{B,\tau} = \sqrt{P_{B,\tau} L_{B,\tau}} \sqrt{\phi_c} U D \tag{3}$$

where $P_{B,\tau}$ is the training power, and $L_{B,\tau}$ is the training duration in the number of symbols. The temporal power control parameter $\phi_c \in \mathbb{R}^+$, the unitary matrix $U \in \mathbb{C}^{n_B \times n_B}$, and the spatial power allocation matrix $D \in \mathbb{R}^{n_B \times m}$ are adapted at the $\mathbf{UT}$ based on the CSIR. For mathematical tractability, $D \in \mathbb{R}^{n_B \times m}$ is restricted to be a diagonal matrix with nonnegative entries d_i, $i = 1, 2, \ldots, m$, satisfying $\|D\|_F^2 \leq 1$. This, along with $\mathbb{E}\phi_c \leq 1$, ensures that $X_{B,\tau}$ satisfies the average power constraint $\mathbb{E}\|X_{B,\tau}\|_F^2 \leq P_{B,\tau} L_{B,\tau}$. The RCT scheme in (3) has the following desirable features: 1) for a given channel realization, D allows the $\mathbf{UT}$ to selectively allot greater or lesser power for training the different channel eigenmodes (spatial power allocation); and 2) power control parameter ϕ_c enables the $\mathbf{UT}$ to perform temporal power allocation.

Note that, in the case of BF, i.e., when $m = 1$, the transmitter requires knowledge of only $\mathbf{v}_1$. Using the proposed RCT scheme, $\mathbf{v}_1$ can be conveyed to $\mathbf{BS}$ using only one training symbol. In contrast, orthogonal RCT requires at least n_B training symbols. Hence, the proposed scheme offers savings in terms of the minimum required training duration.

Now, from (1), the received training signal at the $\mathbf{BS}$, normalized by $\sqrt{P_{B,\tau} L_{B,\tau}}$, is

$$\bar{Y}_{A,\tau} \triangleq \frac{Y_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}} = V \Sigma^H D \sqrt{\phi_c} + \frac{W_{A,\tau}}{\sqrt{P_{B,\tau} L_{B,\tau}}}. \tag{4}$$

Denote the kth columns of $\bar{Y}_{A,\tau}$ and $W_{A,\tau}/\sqrt{P_{B,\tau} L_{B,\tau}}$ by $\bar{\mathbf{y}}_{k,A,\tau}$ and $\mathbf{w}_{k,A,\tau}$, respectively. Note that, in the noiseless case, one can obtain the kth column of V_m by simply normalizing

$\bar{\mathbf{y}}_{k,A,\tau}$. Motivated by this, even in the presence of noise, using $\bar{Y}_{A,\tau}$, the **BS** estimates V_m as $\hat{V}_m \triangleq [\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_m]$, where

$$\hat{\mathbf{v}}_k = \frac{\bar{\mathbf{y}}_{k,A,\tau}}{\bar{\mathbf{y}}_{k,A,\tau_2}}, \qquad 1 \le k \le m. \tag{5}$$

The choice of the matrix D and the parameter ϕ_c determine the allocation of power to different modes and realizations of the channel, which can be used to control the estimation accuracy of $\hat{V}_m$. The design of the RCT sequence involves jointly optimizing matrix D and ϕ_c. In particular, we wish to solve the following two optimization problems.

- *Minimizing the MSE:* The goal here is to minimize the MSE in the estimate of the BF vectors. That is, the optimization problem is given by

$$\min_{\phi_c, D} \mathbb{E}\|V_m - \hat{V}_m\|_F^2 = \min_{\phi_c, D} \mathbb{E} \sum_{k=1}^{m} \|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2 \tag{6}$$

  such that $D \in \mathbb{R}^{n_B \times m}$ is diagonal and nonnegative, $\|D\|_F^2 \le 1$, and $\phi_c > 0$, $\mathbb{E}\phi_c \le 1$.
- *Maximizing a CLB:* The goal here is to maximize a lower bound on the capacity of the system that uses the estimated BF vectors as a precoding matrix during the data transmission phase. That is, the optimization problem is given by

$$\max_{L_{B,\tau}>0, \phi_c>0, D \in \mathbb{R}^{n_B \times m}} C_{\mathrm{LB}} \tag{7}$$

  such that D is diagonal and nonnegative, $\|D\|_F^2 \le 1$, $\mathbb{E}\phi_c \le 1$, and $1 \le L_{B,\tau} \le L_c$. In the above equation, C_{LB} is a lower bound on the capacity, which represents an achievable data rate. We derive an explicit expression for C_{LB} in Section III-B.

Note that, in conventional orthogonal RCT, one uses the training sequence

$$X_{B,\tau}^{(\mathrm{conv})} = \sqrt{\frac{P_{B,\tau} L_{B,\tau}}{n_B}} Q \tag{8}$$

where Q is any $n_B \times L_{B,\tau}$ matrix with orthonormal rows. Further, the transmitter uses the received training signal to estimate the channel matrix $\hat{H}$ using either least squares or minimum MSE (MMSE) estimation,[3] and then employs the m dominant right singular vectors of the estimated channel as the BF vectors for SM data transmission.

We now solve the above optimization problems.

A. Minimizing the MSE

In this subsection, we present the solution to (6). First, for analytical tractability, we present an approximate expression for the MSE, $\mathbb{E}\|V_m - \hat{V}_m\|_F^2$. Then, we use the approximate MSE to optimize the parameters of the training sequence.

1) Approximate MSE in the Estimate of V_m:

Theorem 1: Let the columns of $\hat{V}_m$ be given by (5), and suppose $\phi_c > 0$. Then, there exists a $\hat{V}_{m,\mathrm{approx}} \triangleq V_m + E \in \mathbb{C}^{n_A \times m}$ such that

$$\hat{V}_m = \hat{V}_{m,\mathrm{approx}} + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right) \tag{9}$$

with the columns of $E \triangleq [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m] \in \mathbb{C}^{n_A \times m}$ defined as follows:

$$\mathbf{e}_k \triangleq \frac{-\Re\left\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\right\}}{\sigma_k d_k \sqrt{\phi_c}} \mathbf{v}_k + \frac{1}{\sigma_k d_k \sqrt{\phi_c}} \mathbf{w}_{k,A,\tau}. \tag{10}$$

Further

$$\left|\mathbb{E}\|V_m - \hat{V}_m\|_F^2 - \mathbb{E}\|V_m - \hat{V}_{m,\mathrm{approx}}\|_F^2\right| = \mathcal{O}\left(\frac{1}{x^2}\right) \tag{11}$$

where $x \triangleq P_{B,\tau} L_{B,\tau}$, and the expectation on the left-hand side is with respect to the channel singular values. Also, an approximate expression for the MSE is given by

$$\mathbb{E}\|V_m - \hat{V}_{m,\mathrm{approx}}\|_F^2 = \left(\frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}}\right) \mathbb{E} \sum_{k=1}^{m} \frac{1}{\sigma_k^2 d_k^2 \phi_c}. \tag{12}$$

Proof: See Appendix A. ■

Remark 1: The approximate MSE in (12) differs from $\mathbb{E}\|V_m - \hat{V}_m\|_F^2$ only in the second-order terms of the training power and duration, due to the $(P_{B,\tau} L_{B,\tau})^2$ term in the denominator of (11). In Section V, we illustrate the tightness of the approximate MSE expression at practical training powers.

Remark 2: From the given equation, when $d_k = 1/\sqrt{m}$ and $\phi_c = 1$, i.e., with equal spatio-temporal power allocation, we have

$$\mathbb{E}\|V_m - \hat{V}_{m,\mathrm{approx}}\|_F^2 = \frac{m(2n_A - 1)}{2P_{B,\tau} L_{B,\tau}} \mathbb{E}\left(\sum_{i=1}^{m} \sigma_i^{-2}\right). \tag{13}$$

2) MSE Optimal RCT Design: Now, let us return to the problem of minimizing the MSE in (12) with respect to spatio-temporal parameters D and ϕ_c, i.e.,

$$\min_{\phi_c>0, d_i \ge 0} \left(\frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}}\right) \mathbb{E} \sum_{k=1}^{m} \frac{1}{\sigma_k^2 d_k^2 \phi_c} \tag{14}$$

such that $\sum_{i=1}^{m} d_i^2 \le 1$, and $\mathbb{E}\phi_c \le 1$. The solution is given by the following Lemma.

Lemma 1: The optimal D and ϕ_c that solve (14) are

$$d_k = \sqrt{\frac{\sigma_k^{-1}}{\sum_{i=1}^{m} \sigma_i^{-1}}}, \quad \text{and} \quad \phi_c = \frac{\sum_{i=1}^{m} \sigma_i^{-1}}{\mathbb{E} \sum_{i=1}^{m} \sigma_i^{-1}}. \tag{15}$$

The approximate MSE with $\phi_c = 1$ and optimal D is

$$\mathbb{E}\|V_m - \hat{V}_{m,\mathrm{approx}}\|_F^2 = \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \mathbb{E}\left(\sum_{i=1}^{m} \sigma_i^{-1}\right)^2. \tag{16}$$

The approximate MSE with the jointly optimal D and ϕ_c is

$$\mathbb{E}\|V_m - \hat{V}_{m,\mathrm{approx}}\|_F^2 = \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \left(\mathbb{E} \sum_{i=1}^{m} \sigma_i^{-1}\right)^2. \tag{17}$$

[3] Note that the estimate of the BF vectors obtained using the least squares and MMSE criteria are the same, as they are constant multiples of each other.

Proof: See Appendix B. ■

Comparing the MSE with constant-power training ($\phi_c = 1$) in (16) and the MSE with power-controlled training in (17), it is clear that the power-controlled training outperforms the constant-power training, since $(\mathbb{E}\sum_{i=1}^m \sigma_i^{-1})^2 \leq \mathbb{E}(\sum_{i=1}^m \sigma_i^{-1})^2$ by Jensen's inequality. Thus, spatio-temporal power allocation during training improves the accuracy of the estimate. Note that the given solution is valid if $\mathbb{E}\sum_{i=1}^m \sigma_i^{-1} < \infty$, which is true, for example, when the channel is Rayleigh fading, and $n_A \neq n_B$, since [30]

$$\begin{aligned}\mathbb{E}\sigma_i^{-1} &\leq \sqrt{\mathbb{E}\sigma_i^{-2}} < \sqrt{\mathbb{E}\sum_{i=1}^n \sigma_i^{-2}} \\ &= \sqrt{\mathbb{E}\mathrm{Trace}\{(H^H H)^{-1}\}} = \sqrt{\frac{n_B}{|n_A - n_B|}} < \infty. \quad (18)\end{aligned}$$

Remark 3: For a given channel instantiation, the MSE in estimating the dominant BF vector is small, compared with the modes with smaller gain. Hence, using the inverse of the channel singular values to determine power allocation, as given by (15), is intuitively satisfying. Similarly, across time, it is reasonable to allot power that is proportional to $\sum_{i=1}^m \sigma_i^{-1}$, which, roughly speaking, measures the "goodness" of the channel.

In the above, we optimized power allocated across space and time with the MSE as the performance metric. In the following, we design the RCT to maximize the data rate.

B. Maximizing a CLB

In this subsection, we present the solution to (7). First, we discuss the data transmission scheme that uses the estimated BF vectors as the precoding matrix and derive a lower bound on the capacity. Then, we use the CLB to optimize the parameters of the RCT sequence.

1) Data Transmission and CLB: As mentioned earlier, in Phase III, the data vector $\mathbf{x}_{A,d} \in \mathbb{C}^{m\times 1}$ is premultiplied by $\hat{V}_m \in \mathbb{C}^{n_A\times m}$ obtained from (5), and transmitted to the $\mathbf{UT}$. Since the $\mathbf{UT}$ has perfect CSI, it premultiplies the received data signal by U_m^H, where $U_m \in \mathbb{C}^{n_B\times m}$ is the first m columns of matrix U. From (2), the corresponding received data signal is

$$\mathbf{y}_{B,d} = \sqrt{\frac{P_{A,d}}{m}}\Sigma_m V^H \hat{V}_m \mathbf{x}_{A,d} + U_m^H \mathbf{w}_{A,d}, \quad (19)$$

where $\Sigma_m \in \mathbb{R}^{m\times n_A}$ is the first m rows of matrix $\Sigma \in \mathbb{R}^{n_B\times n_A}$. Note that the distribution of the entries of $\tilde{\mathbf{w}}_{B,d} \triangleq U_m^H \mathbf{w}_{B,d}$ is the same as the entries of $\mathbf{w}_{B,d}$, since U_m^H has orthonormal columns. We rewrite (19) as

$$\mathbf{y}_{B,d} = \sqrt{\frac{P_{A,d}}{m}} G\mathbf{x}_{A,d} + \tilde{\mathbf{w}}_{\mathrm{eff}}, \quad (20)$$

where $G \triangleq \Sigma_{m,m} - \Sigma_m V^H \mathbb{E}_{|H}\{V_e\}$, $V_e \triangleq V_m - \hat{V}_m$, and

$$\begin{aligned}\tilde{\mathbf{w}}_{\mathrm{eff}} \triangleq &\sqrt{\frac{P_{A,d}}{m}}\Sigma_m V^H \mathbb{E}_{|H}\{V_e\}\mathbf{x}_{A,d} \\ &- \sqrt{\frac{P_{A,d}}{m}}\Sigma_m V^H V_e \mathbf{x}_{A,d} + \tilde{\mathbf{w}}_{B,d}. \quad (21)\end{aligned}$$

Here, $\Sigma_{m,m} \in \mathbb{R}^{m\times m}$ is the $m \times m$ principal submatrix of Σ, and (20) is obtained by adding and subtracting the first term in (21). Note that G in the first term in (20) is a deterministic function of H, and it is easy to see that the effective noise term $\tilde{\mathbf{w}}_{\mathrm{eff}}$ is uncorrelated with the data given the channel, i.e., $\mathbb{E}_{|H}\{\tilde{\mathbf{w}}_{\mathrm{eff}}\mathbf{x}_{A,d}^H\} = 0$. Hence, for the system in (20), the worst case noise theorem [31] is applicable, which results in the following lower bound C_{LB} on the channel capacity:

$$C_{\mathrm{LB}} \triangleq \alpha\mathbb{E}_H \log_2 \left| I_m + P_{A,d}\frac{GG^H}{\mathbb{E}_{|H}\{\|\tilde{\mathbf{w}}_{\mathrm{eff}}\|_F^2\}}\right|, \quad (22)$$

where $\alpha \triangleq (L_c - L_{B,\tau})/L_c$. Here, $\mathbb{E}_H\{\cdot\}$ outside the log function denotes the expectation with respect to H, and $\mathbb{E}_{|H}\{\cdot\}$ in the denominator term of (22) and in the definition of G denotes the expectation with respect to the distribution of additive noise in the training and data phases given H.

The derivation of the CLB in (22) is independent of the training scheme used to estimate the dominant BF vectors. Hence, it is also valid for orthogonal RCT.

It turns out that directly optimizing the training sequence to maximize (22) is analytically intractable due to the $\mathbb{E}_{|H}\{V_e\}$ term in the denominator, which is hard to analyze. Hence, we derive an approximate expression for the CLB, and use it to optimize the training sequence. The solution obtained by optimizing the approximate CLB becomes asymptotically accurate as the data and training power become large.

Theorem 2: Let

$$C_{\mathrm{LB,a}} \triangleq \alpha\mathbb{E}\log_2\left| I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{m,m}\Sigma_{m,m}^H}{1+\sigma_{\mathrm{eff}}^2}\right| \quad (23)$$

with

$$\sigma_{\mathrm{eff}}^2 \triangleq \frac{P_{A,d}}{P_{B,\tau}L_{B,\tau}m^2}\sum_{i=1}^m \frac{\beta_i}{d_i^2\phi_c}, \quad \beta_i \triangleq \frac{1}{2} + \frac{\sum_{j=1,j\neq i}^m \sigma_j^2}{\sigma_i^2}. \quad (24)$$

Then, $C_{\mathrm{LB,a}}$ is an approximation to the CLB given by (22), in the sense that

$$|C_{\mathrm{LB}} - C_{\mathrm{LB,a}}| \to 0 \quad \text{as} \quad P_{A,d}, P_{B,\tau} \to \infty$$

such that $P_{A,d}/P_{B,\tau} \leq \mu$, $\mu > 0$.

Proof: See Appendix C. ■

Note that the effect of channel estimation errors on the data rate is captured via σ_{eff}^2 in the denominator of the approximate expression in (23). When the training power is large, the loss in the data rate is small, as expected. The effect of the spatio-temporal power allocation parameters of the training sequence on the data rate is also captured through the σ_{eff}^2 term.

Next, we optimize the training sequence to maximize (23).

2) CLB Optimal RCT Design: We maximize $C_{\mathrm{LB,a}}$, given by (23), subject to $\|D\|_F^2 \leq 1$ and $\mathbb{E}\phi_c \leq 1$. This is equivalent to first optimizing over D for a given ϕ_c, and then finding the optimal functional ϕ_c, as follows:

$$\begin{aligned}\mathbf{P1}: \quad &\max_{\phi_c}\max_{D} \quad C_{\mathrm{LB,a}} \\ &\text{such that} \quad \|D\|_F^2 \leq 1, \quad \text{and} \quad \mathbb{E}\phi_c \leq 1. \quad (25)\end{aligned}$$

The solution is given by the following theorem.

Theorem 3: For a given $L_{B,\tau}$, the optimal $(d_1, d_2, \ldots, d_m)$ that solves $\mathbf{P1}$ in (25) is given by

$$d_i = \sqrt{\frac{\sqrt{\beta_i}}{\sum_{j=1}^{m}\sqrt{\beta_j}}}, \quad 1 \le i \le m \tag{26}$$

where β_i is as defined in (24). Further, the optimal temporal power allocation ϕ_c^* satisfies

$$\lambda = \mathcal{H}(\phi_c^*) \triangleq \left(\frac{1}{\psi + \phi_c^*}\right)\sum_{k=1}^{m}\frac{P_{A,d}\sigma_k^2\psi}{(P_{A,d}\sigma_k^2 + m)\phi_c^* + m\psi} \tag{27}$$

where $\psi \triangleq (P_{A,d}/(P_{B,\tau}L_{B,\tau}m^2))(\sum_{k=1}^{m}\sqrt{\beta_k})^2$, and λ is a Lagrange multiplier, which is chosen such that $\mathbb{E}\phi_c^* = 1$, where the expectation is over the distribution of $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$.

Proof: See Appendix D. ■

Note that, the optimal power control is available only implicitly, as given by (27). The value of λ has to be numerically computed. A procedure for finding λ is described in Section V.

In the special case of pure BF, i.e., when $m = 1$, we have the following simple closed-form solution for the power control policy and the resulting CLB.

Corollary 1: For $m = 1$, the CLB with the optimal power control policy is given by

$$C_{\text{LB,BF}} = \alpha\mathbb{E}\log_2\left(1 + P_{A,d}\frac{\sigma_1^2}{1 + \sigma_{\text{eff,BF}}^2}\right), \tag{28}$$

where $\sigma_{\text{eff,BF}}^2 \triangleq P_{A,d}/(2P_{B,\tau}L_{B,\tau}\phi_c^*)$ with

$$\phi_c^* = \left(\frac{-\psi_2 + \sqrt{\psi_2^2 - 4\psi_1\psi_3}}{2\psi_1}\right)^+.$$

In the above equation, $\psi_1 \triangleq 1 + P_{A,d}\sigma_1^2$, $\psi_2 \triangleq 2\psi + \psi P_{A,d}\sigma_1^2$, and $\psi_3 \triangleq \psi^2 - (P_{A,d}\sigma_1^2\psi/\lambda_{\text{bf}})$. Here, λ_{bf} is chosen to satisfy the average power constraint, and ψ is as defined earlier, with $m = 1$.

Remark 4: The only parameter of the training sequence that remains to be optimized is training duration $L_{B,\tau}$. The optimal $L_{B,\tau}$ can be obtained using a simple off-line search over $\{1, 2, \ldots, L_c\}$. We relegate the details to Section V.

In this subsection, we optimized the spatial and temporal power allocation of the RCT to maximize the approximate CLB. In the following section, we consider the design of the RCT in a multiuser setting and its effect on the data rate.

IV. Multiuser Scenario

In this section, we use the previously derived approximate CLB as a metric to schedule a single user for data transmission and analyze the data rate of the system. If the kth user is scheduled for data transmission, from Theorem 2, with $\phi_c = 1$, the approximate data rate achieved by it is given by

$$R_k \triangleq \alpha\log_2\left|I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{k,m,m}\Sigma_{k,m,m}^H}{1 + \sigma_{k,\text{eff}}^2}\right| \tag{29}$$

where

$$\sigma_{k,\text{eff}}^2 \triangleq \frac{P_{A,d}}{P_{B,\tau}L_{B,\tau}m^2}\left(\sum_{i=1}^{m}\sqrt{\beta_{i,k}}\right)^2$$
$$\beta_{i,k} \triangleq \frac{1}{2} + \frac{\sum_{l=1,l\neq i}^{m}(\sigma_{l,k})^2}{(\sigma_{i,k})^2} \tag{30}$$

and $\Sigma_{k,m,m}$ is the $m \times m$ principal submatrix of the singular value matrix of H_k. For simplicity, we assume that data power, RCT power, and RCT duration are the same for all users. Further, we assume that the channels are i.i.d. across users. The user is selected for data transmission using max-rate scheduling based on R_k. Since R_k can be locally computed at each receiver (user), the user with the highest metric can be efficiently selected using splitting or a timer-based scheme, as explained in Section II.[4] With max-rate scheduling, the average sum data rate is given by

$$R_{\text{avg}} \triangleq \mathbb{E}_{H_1,\ldots,H_M}\left[\max\{R_1, R_2, \ldots, R_M\}\right]. \tag{31}$$

Unfortunately, a closed-form expression for (31) is hard to find when $m > 1$, as it involves finding the distribution of complicated terms involving $\beta_{i,k}$. However, when the channels between the $\mathbf{UT}$s and the $\mathbf{BS}$ are Rayleigh flat fading, and in the case of pure BF, i.e., when $m = 1$, a simplified expression involving a single integral can be found, as shown in the following theorem.

Theorem 4: When $m = 1$ and the channels between the $\mathbf{UT}$s and the $\mathbf{BS}$ are i.i.d. standard Rayleigh distributed, R_{avg} is given by

$$R_{\text{avg}} = \int_0^{\infty}\left(1 - \left(\Pr\left\{\sigma_{1,1}^2 \le \omega\right\}\right)^M\right)dx, \tag{32}$$

where

$$\omega \triangleq \frac{1}{P_{A,d}}\left(2^{\frac{x}{\alpha}} - 1\right)\left(1 + \frac{P_{A,d}}{2P_{B,\tau}L_{B,\tau}}\right)$$

and

$$\Pr\{\sigma_{1,1}^2 \le \omega\} \triangleq \frac{e^{-\omega}}{\prod_{k=1}^{n_B}(n_B - k)!(n_A - k)!}\sum_{j=1}^{n_B}\sum_{p=p_l}^{p_u}\mathcal{S}_{p,j}, \tag{33}$$

where $p_l \triangleq n_A - n_B$, $p_u \triangleq (n_A + n_B)j - 2j^2$, and $\mathcal{S}_{p,j} \triangleq c_{j,p}p![\sum_{s=0}^{p}((-1)^j\omega^{p-s})/(p-s)!]$. The constants $c_{j,p}$ are the coefficients of the $e^{-jy}y^p$ term in the probability density function (pdf) of $\sigma_{1,1}^2$, which is the largest eigenvalue of $H_1^H H_1$ [32, Tab. I].

Proof: See Appendix E. ■

Now, although the given result provides a simple easy-to-compute integral, using which one can analyze the performance of the proposed RCT scheme, it is hard to obtain insight into the system behavior as the number of users, RCT power, data

[4]Note that using the MMSE in the estimate of the BF vectors as the criterion for user selection coincides with max-rate scheduling, when $m = 1$. Both schemes select the user whose channel matrix has the largest dominant singular value.

power, etc., are varied. In the following theorem, we derive an upper bound on R_{avg}, and show that it scales as $((L_c - L_{B,\tau})/L_c)\log\log M$ asymptotically in the number of users M. We illustrate the tightness of the bound in Section V.

Theorem 5: When $m = 1$ and the channels between the $\mathbf{UT}$s and the $\mathbf{BS}$ are i.i.d. standard Rayleigh distributed, an upper bound on R_{avg}, which is denoted by $R^{\text{u}}_{\text{avg}}$, is given by

$$\alpha\log_2\left(1 + \frac{P_{A,d}}{1+\sigma^2_{\text{BF,eff}}}\inf_{s\in(0,1)}\frac{\log M - n_A n_B\log(1-s)}{s}\right) \tag{34}$$

where $\sigma_{\text{BF,eff}} \triangleq (P_{A,d}/(2P_{B,\tau}L_{B,\tau}))$. Further

$$\lim_{M\to\infty}\frac{R^{\text{u}}_{\text{avg}}}{\alpha\log_2\log M} = 1.$$

Proof: See Appendix F. ■

It is interesting to note that a similar data rate scaling of $\log_2\log M$ was also observed in the case of a multiuser multiple-input–single-output downlink channel with perfect CSIR and perfect CSIT [33], [34]. In addition, note that the proposed RCT scheme offers better performance compared with orthogonal RCT, due to the dependence of the data rate on factor α, which captures the training duration overhead.

V. Simulation Results

In this section, we illustrate the performance of the proposed RCT scheme and validate our analytical development using Monte Carlo simulations. The simulation setup consists of an $n_B \times n_A = 3\times 4$ MIMO SM system with $P_{A,d} = P_{B,\tau}$, $L_{B,\tau} = 3$ symbols, and $L_c = 100$ symbols. The MIMO channel is assumed to be a Rayleigh flat-fading channel with i.i.d. coefficients drawn from $\mathcal{CN}(0,1)$. The additive white Gaussian noise is also modeled as having i.i.d. $\mathcal{CN}(0,1)$ components. We calculate the MSE and the CLB by averaging over 10^5 independent channel and training noise instantiations. With conventional orthogonal RCT, we employ the training sequence in (8), as in [3], to obtain an estimate of the channel, from which the m dominant BF vectors are computed using SVD. The resulting error in the estimated BF vectors is used to compute the MSE, and the CLB is computed using the expression in (22). The computed CLB is used to schedule the user for RCT and data transmission, and hence, with conventional estimation also, only the selected user sends the RCT signal. We compare the performance of the proposed RCT scheme with orthogonal RCT in terms of both the MSE and the CLB in the following sections.

A. MSE

Fig. 2 shows the performance of the training scheme proposed in Section III-A in terms of the MSE in the estimate of V_m versus $P_{B,\tau}$, with $m = 3$ modes. Since the proposed RCT scheme has multiple parameters, it is of interest to see the gain offered by optimizing each of the components. Toward this, we plot the MSE with the following settings: 1) $X_{B,\tau}$ in (3) with $D = I_{n_B\times m}$ and $\phi_c = 1$ (fixed power RCT); 2) $X_{B,\tau}$ in (3) with optimal D and $\phi_c = 1$ (RCT with optimal spatial power control); 3) $X_{B,\tau}$ in (3) with jointly optimal D and ϕ_c (RCT with the optimal spatio-temporal power control); and 4) conventional orthogonal training $(X^{(\text{conv})}_{B,\tau}$ in (8)) [3]. From the plot, we observe that the approximate theoretical expression in (17) is tight, and that the proposed scheme significantly outperforms orthogonal RCT at all average training SNRs. Using the optimal spatial power allocation during training offers a gain of approximately 1 dB at $P_{B,\tau} = 15$ dB compared with using $D = I_{n_B\times m}$. However, temporal power allocation does not further significantly improve the performance compared with pure spatial power allocation.

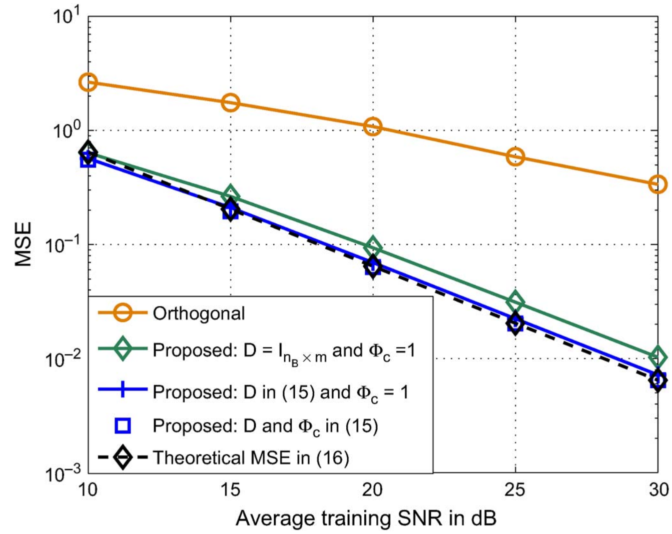

Fig. 2. MSE versus average training SNR for a 3×4 MIMO system with $m = 3$.

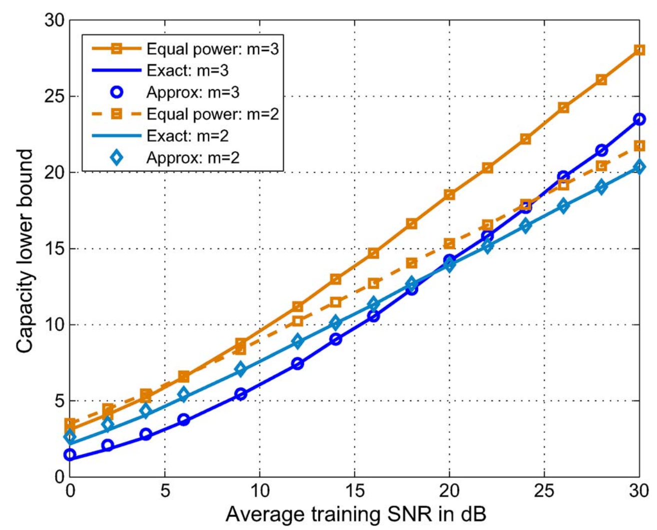

Fig. 3. Tightness of the proposed approximation in (23) for the CLB in (22) for a 3×4 MIMO system with the average data SNR of $P_{A,d} = P_{B,\tau}$, $L_{B,\tau} = 3$ symbols.

B. Data Rate With a Single User

Fig. 3 shows a plot of the CLB in (22) and its approximate expression in (23) versus the average training SNR. It is clear that the approximate expression for the CLB is tight at all average training SNRs. The best value of m, in terms of the

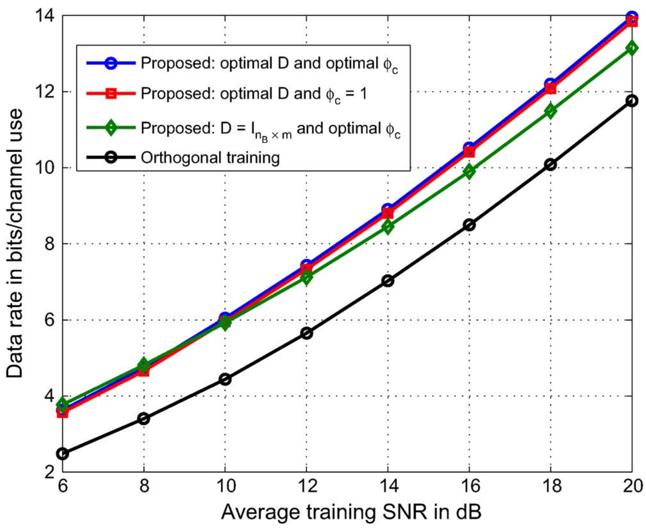

Fig. 4. CLB in (22) for a 3 × 4 MIMO system versus average training SNR $P_{B,\tau}$, with $P_{A,d} = P_{B,\tau}$, $L_{B,\tau} = 3$ symbols, and $m = 3$ modes.

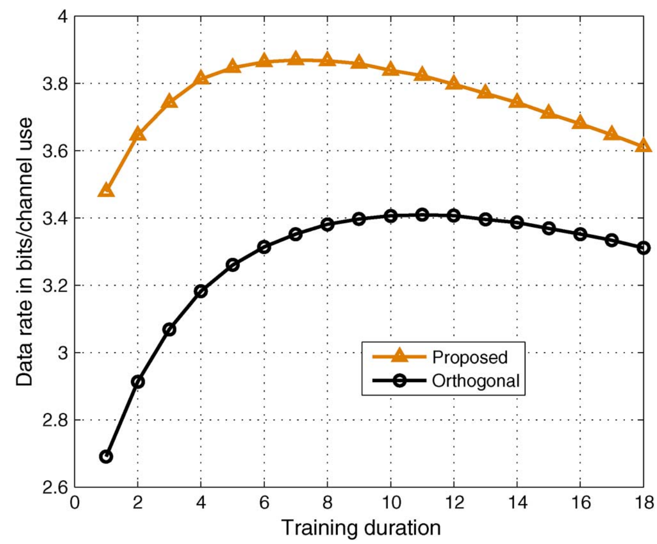

Fig. 5. CLB in (22) for a 3 × 4 MIMO system versus training duration with $P_{A,d} = P_{B,\tau} = 6$ dB.

achievable rate, depends on the average SNR. That is, at low SNRs, $m = 1$ achieves the best data rate. As the average SNR is increased, due to the higher slope of the curves, larger values of m offer the best data rate, and at very high SNRs, $m = \min\{n_A, n_B\}$ is optimal. This observation has been also made in [20]. Hence, in Fig. 3, up until an average training SNR of 20 dB, using $m = 2$ modes offers a better data rate compared with using three modes. However, beyond 20 dB, it is better to use three modes. The figure also shows the capacity performance of SM with equal power allocation over $m = 2$ and $m = 3$ modes, in the genie-aided case where perfect CSI is available at the **BS**. In contrast to the imperfect-CSIT case, the crossover happens at 6 dB in the genie-aided case. Further, the data rate corresponding to $m = 2$ outperforms $m = 3$, i.e., using higher number of modes is not always optimal when the training overhead for acquiring CSIT is taken into account.

For the optimal spatio-temporal power allocation, the value of the Lagrange multiplier $\lambda > 0$ in (27) is required to plot the CLB. To compute this, we start with some $\lambda > 0$. We generate a large number of channel instantiations and compute the power control function ϕ_c^* for each instantiation by inverting (27), as it is a monotone function of ϕ_c. We then compute the average of the values of the ϕ_c's so obtained. Due to the monotonicity of function $\mathcal{H}(\phi_c^*)$, if the average exceeds unity, we increment λ; otherwise, we decrement λ by a small step. Repeating this procedure until the average value is sufficiently close to 1 yields the desired λ and, consequently, the optimal power control function.

Fig. 4 shows a plot of the exact CLB in (22) versus the average training SNR. The proposed training offers an improvement of about 2 bits per use of the channel over orthogonal training. Also, optimal spatial power allocation during training outperforms the proposed scheme with $D = I_{n_B \times m}$ by approximately 1 bit per use of the channel at $P_{B,\tau} = 20$ dB. On the other hand, temporal power allocation during training only offers a marginal data rate improvement.

Fig. 5 shows the CLB in (22) versus training duration with $P_{A,d} = P_{B,\tau} = 6$ dB for the proposed and orthogonal training schemes. The figure shows that training for the minimum duration of one symbol is not always optimal. For the proposed training scheme, the optimal training duration is eight symbols, whereas for the orthogonal training scheme, it is 12 symbols. Thus, the analysis presented in this paper can be used to determine the training duration that optimally trades off the estimation accuracy with the time overhead due to training.

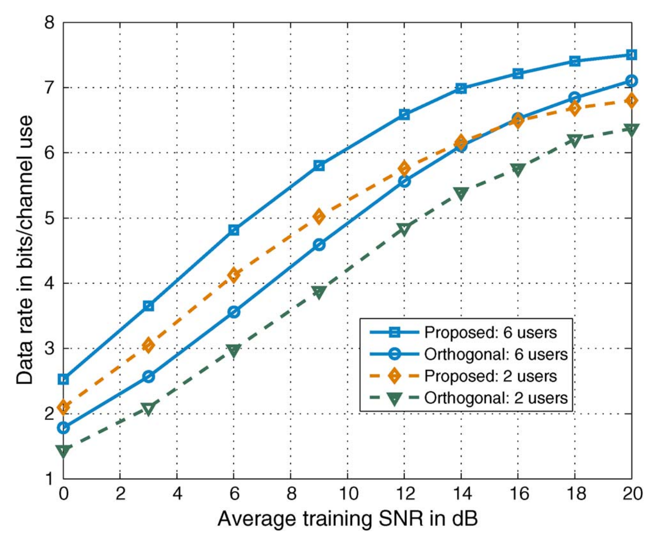

Fig. 6. CLB in (22) for a 3 × 4 multiuser MIMO system versus average training SNR $P_{B,\tau}$, with the data power $P_{A,d} = P_{B,\tau}$ and for the scheduling scheme described in Section IV.

C. Data Rate With Multiple Users

Fig. 6 shows a comparison of the proposed RCT scheme with orthogonal RCT in terms of the average data rate versus $P_{B,\tau}$ for a multiuser system with max-rate-based user scheduling, $m = 3$ modes, and $M = 2$ and 6 users. We use the CLB expression in (22) while evaluating the average data rate in (31). The proposed RCT scheme can lead to a reduction of 2–3 dB in the average training power compared with orthogonal RCT, for achieving the same data rate.

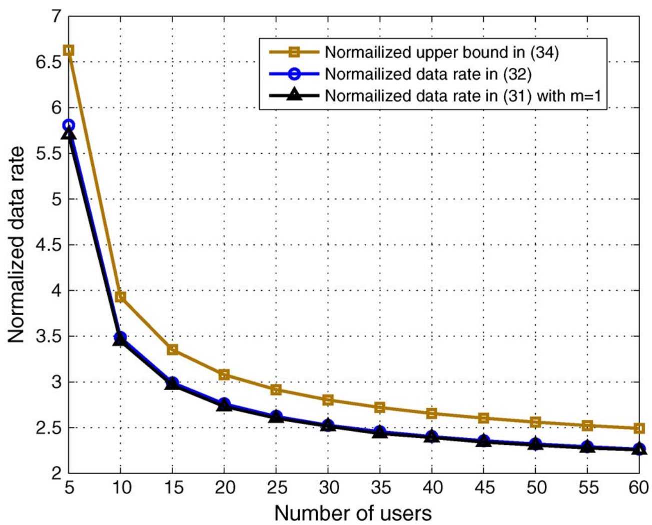

Fig. 7. Normalized data rate versus number of users M for a 2×2 multiuser MIMO BF system with $P_{A,d} = P_{B,\tau} = 6$ dB. Here, the data rate is normalized by $\log_2 \log M$.

Finally, in Fig. 7, we study the behavior of the average data rate as a function of the number of users. We consider a 2×2 multiuser BF system with $L_{B,\tau} = 1$ and plot the normalized data rate, which is defined as $R_{\text{avg}}/\log_2 \log M$, where R_{avg} is as in (31), versus M. We see that the approximate expression in (32) matches well with the exact expression in (31). Further, the upper bound in (34) also captures the $\log_2 \log M$ scaling of the average data rate very well.

VI. Conclusion

In this paper, we have considered a multiuser SM-based TDD-MIMO system with perfect CSIR. First, for a single-user system, a novel *power-controlled* RCT scheme that adapts to the time-varying channel was proposed. This was used by the **BS** to estimate the dominant BF vectors of the channel. The spatial and temporal allocation of the training matrix was optimized using the following two metrics: 1) a CLB, and 2) the MSE, subject to an average power constraint. We then extended the training scheme and the data rate analysis to a multiuser case. Further, for a BF system, we derived a closed-form expression for the average sum data rate and its upper bound. We showed that the upper bound scales as $((L_c - L_{B,\tau})/L_c) \log \log M$ asymptotically in the number of users M, where L_c and $L_{B,\tau}$ are the channel coherence time and the training duration, respectively. Using simulations, we demonstrated the significant performance gain offered by the proposed training sequence over conventional orthogonal RCT.

Appendix A
Proof of Theorem 1

To derive an approximate expression for $\sum_{k=1}^{m} \mathbb{E}\|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2$, we first find the Taylor series expansion of $\hat{\mathbf{v}}_k$, as follows. Substituting for $\bar{\mathbf{y}}_{k,A,\tau}$, the kth column of $\bar{Y}_{A,\tau}$, in (4), the estimate of the kth singular vector in (5) becomes

$$\hat{\mathbf{v}}_k = \frac{(\sigma_k d_k \sqrt{\phi_c} \mathbf{v}_k + \mathbf{w}_{k,A,\tau})}{\left\|(\sigma_k \sqrt{\phi_c} d_k \mathbf{v}_k + \mathbf{w}_{k,A,\tau})\right\|_2} \tag{35}$$

$$= \frac{\mathbf{v}_k + \frac{\mathbf{w}_{k,A,\tau}}{\sigma_k d_k \sqrt{\phi_c}}}{\sqrt{1+x}} \tag{36}$$

$$= \mathbf{v}_k - \frac{\Re\left\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\right\} \mathbf{v}_k}{\sigma_k d_k \sqrt{\phi_c}} + \frac{\mathbf{w}_{k,A,\tau}}{\sigma_k d_k \sqrt{\phi_c}} + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right), \tag{37}$$

where the last expression follows by using $1/\sqrt{1+x} = 1 - x/2 + \mathcal{O}(x^2)$, where $x \triangleq (2\Re\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\}/\sigma_k d_k \sqrt{\phi_c}) + (\mathbf{w}_{k,A,\tau}^H \mathbf{w}_{k,A,\tau}/\sigma_k^2 d_k^2 \phi_c)$, and retaining only the terms of the order strictly less than $\mathcal{O}(1/(P_{B,\tau} L_{B,\tau}))$. Thus, $\hat{V}_m = \hat{V}_{m,\text{approx}} + \mathcal{O}(1/(P_{B,\tau} L_{B,\tau}))$, with $\hat{V}_{m,\text{approx}} = V_m + E$, where E is as defined in the theorem. Let $\mathbf{e}_k$ denote the kth column of E. Let $\hat{\mathbf{v}}_{k,\text{approx}}$ be the kth column of matrix $\hat{V}_{m,\text{approx}}$. From (37), we have

$$\mathbb{E}\left\{\|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2\right\} = \mathbb{E}\left\|\mathbf{e}_k + \mathcal{O}\left(\frac{1}{P_{B,\tau} L_{B,\tau}}\right)\right\|_2^2 \tag{38}$$

$$= \mathbb{E}\left\{\|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2\right\} + \mathcal{O}\left(\frac{1}{(P_{B,\tau} L_{B,\tau})^2}\right). \tag{39}$$

To obtain the above equation, recall from the statement of the theorem that $\mathbf{e}_k \triangleq ((-\Re\{\mathbf{v}_k^H\ \mathbf{w}_{k,A,\tau}\})/(\sigma_k d_k \sqrt{\phi_c}))\ \mathbf{v}_k + (1/(\sigma_k d_k \sqrt{\phi_c}))\ \mathbf{w}_{k,A,\tau}$. Hence, $\mathbb{E}\mathbf{e}_k^H x^2$ contains a term of the order $\mathcal{O}(1/(P_{B,\tau} L_{B,\tau})^2)$ arising from the first term in both x and $\mathbf{e}_k$, plus terms with odd moments of a Gaussian distribution, which are equal to zero, due to which $\mathbb{E}\{\mathbf{e}_k^H x^2\} = \mathcal{O}(1/(P_{B,\tau} L_{B,\tau})^2)$. Now, subtracting $\mathbb{E}\{\|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2\}$ on both sides of (39), summing over k, and taking the absolute value, we get

$$\left|\mathbb{E}\|V_m - \hat{V}_m\|_F^2 - \mathbb{E}\|V_m - \hat{V}_{\text{approx}}\|_F^2\right| = \left|\sum_{k=1}^{m} \mathbb{E}\left\{\|\mathbf{v}_k - \hat{\mathbf{v}}_k\|_2^2\right\} - \mathbb{E}\left\{\|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2\right\}\right| = \mathcal{O}\left(\frac{1}{(P_{B,\tau} L_{B,\tau})^2}\right). \tag{40}$$

To complete the proof, we evaluate $\sum_{k=1}^{m} \mathbb{E}\{\|\mathbf{v}_k - \hat{\mathbf{v}}_{k,\text{approx}}\|_2^2\} = \sum_{i=1}^{m} \mathbb{E}\|\mathbf{e}_k\|_2^2$, as follows:

$$\mathbb{E}\|\mathbf{e}_k\|_2^2 = \mathbb{E}\left[\frac{\mathbb{E}\|\mathbf{w}_{k,A,\tau}\|_2^2 - \mathbb{E}\left|\Re\left\{\mathbf{v}_k^H \mathbf{w}_{k,A,\tau}\right\}\right|^2}{\sigma_k^2 d_k^2 \phi_c}\right] \tag{41}$$

$$= \frac{2n_A - 1}{2P_{B,\tau} L_{B,\tau}} \mathbb{E} \frac{1}{\sigma_k^2 d_k^2 \phi_c}. \tag{42}$$

The facts that $\mathbb{E}|\Re\{\mathbf{v}_k^H\ \mathbf{w}_{k,A,\tau}\}|^2 = 1/(2P_{B,\tau} L_{B,\tau})$ and $\mathbb{E}\|\mathbf{w}_{k,A,\tau}\|_2^2 = n_A/(P_{B,\tau} L_{B,\tau})$ have been used to obtain the given equation. This completes the proof. ∎

APPENDIX B
PROOF OF LEMMA 1

Recall that the approximate MSE $\mathbb{E}\|V_m - \hat{V}_{m,\text{approx}}\|_F^2 = [(2n_A - 1)/(2P_{B,\tau}L_{B,\tau})]\mathbb{E}[(1/\phi_c)\sum_{i=1}^m(1/(\sigma_i^2 d_i^2))]$. Now, the problem in (14) can be rewritten as

$$\min_{\phi_c>0, d_i\geq 0} \mathbb{E}\|V_m - \hat{V}_{m,\text{approx}}\|_F^2 \tag{43}$$

such that $\sum_{i=1}^m d_i^2 = 1$, and $\mathbb{E}\phi_c = 1$. Now, without loss of optimality, we can first optimize d_i for a given ϕ_c, substitute the solution into the objective function, and then optimize ϕ_c. Using the Lagrange multiplier method, the optimal d_i is given by

$$d_i = \sqrt{\frac{\sigma_i^{-1}}{\sum_{k=1}^m \sigma_k^{-1}}}, \tag{44}$$

and the corresponding approximate MSE is given by $[(2n_A - 1)/(2P_{B,\tau}L_{B,\tau})]\mathbb{E}[(\sum_{i=1}^m \sigma_i^{-1})/(\sqrt{\phi_c})^2]$. Note that, when $\phi_c = 1$, this corresponds to the MSE for temporally constant power training. Now, we solve the following problem:

$$\min_{\phi_c:\mathbb{E}\phi_c\leq 1} \mathbb{E}\left\{\frac{\left(\sum_{i=1}^m \sigma_i^{-1}\right)^2}{\phi_c}\right\}. \tag{45}$$

Due to the convexity of the problem, using variational calculus [35], the solution is

$$\phi_c = \frac{\sum_{i=1}^m \sigma_i^{-1}}{\mathbb{E}\sum_{i=1}^m \sigma_i^{-1}}, \tag{46}$$

and the corresponding approximate MSE is given by

$$\mathbb{E}\|V_m - \hat{V}_{m,\text{approx}}\|_F^2 = \frac{2n_A - 1}{2P_{B,\tau}L_{B,\tau}}\left(\mathbb{E}\sum_{i=1}^m \sigma_i^{-1}\right)^2. \tag{47}$$

■

APPENDIX C
PROOF OF THEOREM 2

We need to show that as $P_{B,\tau}, P_{A,d} \to \infty$ with $P_{A,d}/P_{B,\tau} = \mu$, $|C_{\text{LB}} - C_{\text{LB,a}}| \to 0$, where C_{LB} and $C_{\text{LB,a}}$ are as defined in (22) and (23), respectively. From (21), $(1/m)\mathbb{E}_{|H}\{\|\tilde{\mathbf{w}}_{\text{eff}}\|_F^2\}$ can be written as

$$1 + \frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\left\{\|\Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e\|_F^2\right\}. \tag{48}$$

Using the above equation and the definition $G \triangleq \Sigma_{m,m} - \Sigma_m V^H \mathbb{E}_{|H}\{V_e\}$, $C_f \triangleq C_{\text{LB}}/\alpha$ can be written as

$$C_f = \mathbb{E}\log_2\left|I_m + \frac{P_{A,d}}{m}\frac{GG^H}{1+g^2}\right| \tag{49}$$

$$= \mathbb{E}\log_2\left|I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{m,m}^H\Sigma_{m,m}}{1+g^2} + \frac{\Gamma}{1+g^2}\right| \tag{50}$$

$$= \mathbb{E}\log_2\left|I_m + \left(I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{m,m}^H\Sigma_{m,m}}{1+g^2}\right)^{-1}\frac{\Gamma}{1+g^2}\right| + \mathbb{E}\log_2\left|I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{m,m}^H\Sigma_{m,m}}{1+g^2}\right|, \tag{51}$$

where

$$g \triangleq \sqrt{\frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\left\{\|\Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e\|_F^2\right\}},$$

$$\Gamma \triangleq \frac{P_{A,d}}{m}\Big\{-2\Re\left[\Sigma_{m,m}\mathbb{E}_{|H}\left\{V_e^H\right\}V\Sigma_m^H\right] + \Sigma_m V^H \mathbb{E}_{|H}\{V_e\}\mathbb{E}_{|H}\left\{V_e^H\right\}V\Sigma_m^H\Big\}. \tag{52}$$

Substituting for g, and using

$$X \triangleq \frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\left\{\|\Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e\|_F^2\right\}I_m + \frac{P_{A,d}}{m}\Sigma_{m,m}\Sigma_{m,m}^H \tag{53}$$

leads to the following expression:

$$C_f = \mathbb{E}\log_2\left|I_m + \frac{\frac{P_{A,d}}{m}\Sigma_{m,m}\Sigma_{m,m}^H}{1 + \frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\|G_{\text{denom}}\|_F^2}\right| + \mathbb{E}\log_2\left|I_m + (I_m + X)^{-1}\Gamma\right|, \tag{54}$$

where $G_{\text{denom}} \triangleq \Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e$. In the above equation, $\mathbb{E}_{|H}\{\cdot\}$ denotes the expectation conditioned on channel H. Now, the proof would be complete if we show that the following claims hold.

1) **Claim 1**

$$\frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\|\Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e\|_F^2 \to \sigma_{\text{eff}}^2 \tag{55}$$

as $P_{A,d}, P_{B,\tau} \to \infty$, with $P_{A,d}/P_{B,\tau} = \mu > 0$.

2) **Claim 2**

$$\mathbb{E}\log_2\left|I_m + (I_m + X)^{-1}\Gamma\right| \to 0 \tag{56}$$

as $P_{A,d}, P_{B,\tau} \to \infty$, with $P_{A,d}/P_{B,\tau} = \mu > 0$.

1) Proof of Claim 1: Recall from (9) in Theorem 1 that $V_e \triangleq V_m - \hat{V}_m = E + \mathcal{O}(1/(P_{B,\tau}L_{B,\tau}))$. Since $\mathbb{E}_{|H}\{E\} = 0$, we have $\mathbb{E}_{|H}\{V_e\} = \mathcal{O}(1/(P_{B,\tau}L_{B,\tau}))$. Using this in the expression in **Claim 1**, we get

$$\frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\|\Sigma_m V^H \mathbb{E}_{|H}\{V_e\} - \Sigma_m V^H V_e\|_F^2 = \frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\|\Delta\|_F^2 + \frac{P_{A,d}}{m^2}\mathcal{O}\left(\frac{1}{(P_{B,\tau}L_{B,\tau})^{3/2}}\right) \tag{57}$$

where $\Delta \triangleq \Sigma_m V^H E \in \mathbb{C}^{m\times m}$. First, we compute an expression for $(P_{A,d}/m^2)\mathbb{E}_{|H}\{\|\Delta\|_F^2\}$ as follows. The (i,j)th entry of Δ is $\Delta_{ij} = \sigma_i \mathbf{v}_i^H \mathbf{e}_j$, with $\mathbf{e}_j$ representing the jth column of

E defined in Theorem 1. Substituting for $\mathbf{e}_j$, the (i,j)th element of Δ can be written as

$$\Delta_{ij} = \begin{cases} \frac{1}{d_i\sqrt{\phi_c}}\sqrt{-1}\Im\{\mathbf{v}_i^H\mathbf{w}_{i,A,\tau}\}, & i=j, 1\le i,j\le m \\ \frac{\sigma_i}{\sigma_j d_j\sqrt{\phi_c}}\mathbf{v}_i^H\mathbf{w}_{j,A,\tau}, & i\neq j, 1\le i,j\le m. \end{cases} \quad (58)$$

From (58)

$$\mathbb{E}_{|H}\{|\Delta_{ij}|^2\} = \begin{cases} \frac{1}{2P_{B,\tau}L_{B,\tau}d_j^2\phi_c}, & i=j, 1\le i,j\le m \\ \frac{1}{P_{B,\tau}L_{B,\tau}}\frac{\sigma_i^2}{\sigma_j^2 d_j^2\phi_c}, & i\neq j, 1\le i,j\le m. \end{cases} \quad (59)$$

From the above equation, it follows that $(P_{A,d}/m^2)\mathbb{E}_{|H}\{\|\Delta\|_F^2\} = \sigma_{\text{eff}}^2$. Since

$$\frac{P_{A,d}}{m}\mathcal{O}\left(\frac{1}{(P_{B,\tau}L_{B,\tau})^{3/2}}\right) = \mathcal{O}\left(\frac{\mu}{\sqrt{P_{B,\tau}}L_{B,\tau}^{3/2}}\right) \to 0$$

as $P_{A,d}, P_{B,\tau} \to \infty$ such that $P_{A,d}/P_{B,\tau} = \mu > 0$, we get

$$\frac{P_{A,d}}{m^2}\mathbb{E}_{|H}\left\{\left\|-\Sigma_m V^H\mathbb{E}_{|H}\{V_e\} + \Sigma_m V^H V_e\right\|_F^2\right\} \to \sigma_{\text{eff}}^2$$

where $\sigma_{\text{eff}}^2 \triangleq (\mu/(L_{B,\tau}m^2))\sum_{i=1}^m(\beta_i/(d_i^2\phi_c))$, which proves **Claim 1**.

2) Proof of Claim 2: First, note that Γ converges to a finite constant as $P_{A,d}, P_{B,\tau} \to \infty$ such that $P_{A,d}/P_{B,\tau} = \mu > 0$ since $\mathbb{E}_{|H}\{V_e\} = \mathcal{O}(1/(P_{B,\tau}L_{B,\tau}))$. From **Claim 1** and from the definition of X, it is clear that $(I_m + X)^{-1} \to 0$, as $P_{A,d}, P_{B,\tau} \to \infty$. Thus, we have $(I_m + X)^{-1}\Gamma \to 0$ at the rate of $1/P_{A,d}$. Since $\log_2|\cdot|$ is continuous, from (54), we have

$$\left|C_f - \mathbb{E}\log_2\left|I_m + \frac{P_{A,d}}{m}\frac{\Sigma_{m,m}\Sigma_{m,m}^H}{1+\sigma_{\text{eff}}^2}\right|\right| \to 0 \quad (60)$$

as $P_{A,d}, P_{B,\tau} \to \infty$, with $P_{A,d}/P_{B,\tau} = \mu > 0$, which completes the proof. ■

Appendix D
Proof of Theorem 3

Maximizing $C_{\text{LB,a}}$ given in Theorem 2 with respect to D is equivalent to solving the following optimization problem:

$$\min_{d_i\ge 0:\sum_{i=1}^m d_i^2\le 1}\sum_{i=1}^m\frac{\beta_i}{d_i^2}. \quad (61)$$

The solution in (26) now follows directly by noting that (61) is convex in $(d_1,\ldots,d_m)$ and using the Lagrangian multiplier method. The resulting expression for $C_{\text{LB,a}}$ is given by

$$C_{\text{LB,a}} = \frac{L_c - L_{B,\tau}}{L_c}\sum_{i=1}^m\mathbb{E}\log_2\left(1 + \frac{P_{A,d}}{m}\frac{\sigma_i^2\phi_c}{\psi+\phi_c}\right), \quad (62)$$

where $\psi \triangleq (P_{A,d}/(P_{B,\tau}L_{B,\tau}m^2))(\sum_{k=1}^m\sqrt{\beta_k})^2$. Since the objective functional in (62) is concave in ϕ_c and the constraint is convex, we get the necessary and sufficient condition in (27) by differentiating the Lagrangian, equating it to zero, and solving for λ. ■

Appendix E
Proof of Theorem 4

Substituting for R_1 to R_M from (29) with $m=1$, $\bar{R}_{\max} \triangleq \mathbb{E}\max\{R_1,\ldots,R_M\}$ can be written as

$$\begin{aligned} \bar{R}_{\max} &= \int_0^\infty \Pr\left\{\log_2\left[1 + \frac{P_{A,d}}{1+\sigma_{\text{eff}}^2}\max_{1\le i\le M}\{\sigma_{1,i}^2\}\right] > \frac{x}{\alpha}\right\}dx \\ &= \int_0^\infty \Pr\left\{\max_{1\le i\le M}\sigma_{1,i}^2 > \omega\right\}dx \quad (63) \\ &= \int_0^\infty \left(1 - \left(\Pr\{\sigma_{1,1}^2\le\omega\}\right)^M\right)dx, \quad (64) \end{aligned}$$

where ω is as defined in Theorem 4, and the last equality follows since the singular values are i.i.d. across users. Now, we need to find an expression for $\Pr\{\sigma_{1,1}^2 \le \omega\}$. From [32], for Rayleigh flat-fading channels, the pdf of $\sigma_{1,1}^2$, which is denoted by $f_{\sigma_{1,1}^2}(y)$, is given by

$$f_{\sigma_{1,1}^2}(y) = \frac{1}{\prod_{k=1}^{n_B}(n_B-k)!(n_A-k)!}\sum_{j=1}^{n_B}\sum_{p=n_A-n_B}^{(n_A+n_B)j-2j^2}\mathcal{T}_{j,p}, \quad (65)$$

where $\mathcal{T}_{j,p} \triangleq c_{j,p}p!e^{-jy}y^p$, and coefficients $c_{j,p}$ are as described in [32]. Using the identity [36]

$$\int e^{ax}x^q dx = e^{ax}\sum_{l=0}^q(-1)^l\frac{q!x^{q-l}}{(q-l)!a^{l+1}}, \quad (66)$$

we get the desired result. ■

Appendix F
Proof of Theorem 5

Substituting for R_k from (29) and with $m=1$, $\bar{R}_{\max} \triangleq \mathbb{E}\max\{R_1,\ldots,R_M\}$ can be written as

$$\begin{aligned} \bar{R}_{\max} &\triangleq \alpha\mathbb{E}\log_2\left(1 + \frac{P_{A,d}}{1+\sigma_{\text{eff}}^2}\max_{1\le j\le M}\{\sigma_{1,j}^2\}\right) \quad (67) \\ &\le \alpha\log_2\left(1 + \frac{P_{A,d}}{1+\sigma_{\text{eff}}^2}\mathbb{E}\max_{1\le j\le M}\{\sigma_{1,j}^2\}\right) \quad (68) \end{aligned}$$

where σ_{eff}^2 is as defined in (30), and (68) follows from Jensen's inequality. Now, we find an upper bound on $\mathbb{E}\max\{\sigma_{1,1}^2,\ldots,\sigma_{1,M}^2\}$ as follows. Let $\bar{\sigma}_{\max}^2 \triangleq \mathbb{E}\max_{1\le i\le M}\sigma_{1,i}^2$. Pick $s>0$, and consider

$$\begin{aligned} \exp\{s\bar{\sigma}_{\max}^2\} &\overset{(a)}{\le} \mathbb{E}\exp\left\{s\max_{1\le i\le M}\sigma_{1,i}^2\right\} \\ &= \int_0^\infty \Pr\left\{\exp\left\{s\max_{1\le i\le M}\sigma_{1,i}^2\right\} > x\right\}dx \quad (69) \\ &\overset{(b)}{\le} \int_0^\infty\sum_{k=1}^M\Pr\{\exp\{s\sigma_{1,k}^2\} > x\}dx \quad (70) \\ &\overset{(c)}{=} M\int_0^\infty\Pr\{\exp\{s\sigma_{1,1}^2\} > x\}dx \quad (71) \\ &\le M\mathbb{E}\{\exp\{s\sigma_{1,1}^2\}\} \quad (72) \\ &\overset{(d)}{\le} M\mathbb{E}\{\exp\{s\|H_1\|_F^2\}\}. \quad (73) \end{aligned}$$

In the above equations, (a) follows from Jensen's inequality, (b) follows from the union bound, (c) follows from the fact that $\sigma_{1,i}$ are i.i.d., and finally, (d) follows from the fact that $\sigma_{1,1}^2 \leq \|H_1\|_F^2$. Now, we evaluate $\mathbb{E}\{\exp\{s\|H_1\|_F^2\}\}$. Using the fact that $\|H\|_2^2$ is a chi-square random variable with $2n_An_B$ degrees of freedom when the channel is Rayleigh flat fading, we have

$$\mathbb{E}\left\{e^{\left\{s\|H_1\|_F^2\right\}}\right\} = \frac{1}{(n_An_B-1)!}\int_0^\infty x^{n_Bn_B-1}e^{-(1-s)x}dx$$
$$= \frac{1}{(1-s)^{n_An_B}}, \quad s \in (0,1). \tag{74}$$

Using the above equation in (73), and taking the logarithm on both sides, we get

$$\mathbb{E}\max_{1\leq i\leq M}\sigma_{1,i}^2 \leq \inf_{s\in(0,1)}\left[\frac{\log M}{s} - \frac{n_An_B\log(1-s)}{s}\right]. \tag{75}$$

Substituting the given equation in (68), we get (34). Choosing $s = 1/2$ in (34), we get $R_{\text{avg}} \leq R_{\text{avg}}^{\text{u}}$, where

$$R_{\text{avg}}^{\text{u}} \triangleq \alpha\log_2\left(1 + \frac{P_{A,d}}{1+\sigma_{\text{eff}}^2}\left[2\log M + 2n_An_B\log 2\right]\right). \tag{76}$$

For large M, it is easy to see that $\lim_{M\to\infty} R_{\text{avg}}^{\text{u}}/(\alpha\log_2\log M) = 1$. This completes the proof. ■

References

[1] B. N. Bharath and C. R. Murthy, "Channel estimation at the transmitter in a reciprocal MIMO spatial multiplexing system," in *Proc. NCC*, 2012, pp. 1–5.

[2] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 4, pp. 585–595, Dec. 1999.

[3] R. Venkataramani and T. L. Marzetta, "Reciprocal training and scheduling protocol for MIMO systems," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Oct. 2003, p. 304.

[4] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[5] T. L. Marzetta and B. M. Hochwald, "Learning the channel at the transmitter," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2004, pp. 1160–1169.

[6] C. Steger, A. Khoshnevis, A. Sabharwal, and B. Aazhang, "The case for transmitter training," in *Proc. IEEE Int. Symp. Inf. Theory*, Washington, DC, USA, Sep. 2007.

[7] D. Samardzija and N. Mandayam, "Impact of pilot design on achievable data rates in multiple antenna multiuser TDD systems," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1370–1379, Sep. 2007.

[8] L. P. Withers, R. M. Taylor, and M. W. David, "Echo-MIMO: A two-way channel training method for matched cooperative beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4419–4432, Sep. 2008.

[9] J. Chen and D. Slock, "Comparison of two analog feedback schemes for transmit side MIMO channel estimation," in *Proc. PIMRC*, 2007, pp. 1–5.

[10] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2424–2431, Sep. 2004.

[11] S. Gazor and K. AlSuhaili, "Communications over the best singular mode of a reciprocal MIMO channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1993–2001, Jul. 2010.

[12] R. Prasad, B. N. Bharath, and C. R. Murthy, "Joint data detection and dominant singular mode estimation in time varying reciprocal MIMO systems," in *Proc. ICASSP*, Prague, Czech Republic, May 2011, pp. 3240–3243.

[13] U. Salim, D. Gesbert, and D. Slock, "Combining training and quantized feedback in multiantenna reciprocal channels," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1383–1396, Mar. 2012.

[14] B. N. Bharath and C. R. Murthy, "Reverse channel training for reciprocal MIMO systems with spatial multiplexing," in *Proc. ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 2673–2676.

[15] B. N. Bharath and C. R. Murthy, "Channel training signal design for reciprocal multiple antenna systems with beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 140–151, Jan. 2013.

[16] B. N. Bharath and C. R. Murthy, "On the DMT of TDD-SIMO systems with channel-dependent reverse channel training," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3332–3341, Nov. 2012.

[17] B. N. Bharath and C. R. Murthy, "Power controlled reverse channel training achieves an infinite diversity order in a TDD-SIMO system with perfect CSIR," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1800–1803, Nov. 2012.

[18] X. Zhou, T. Lamahewa, P. Sadeghi, and S. Durrani, "Two-way training: Optimal power allocation for pilot and data transmission," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 564–569, Feb. 2010.

[19] E. Martinian, *Waterfilling Gains O(1/SNR) at High SNR*, 2004. [Online]. Available: citeseer.ist.psu.edu/646493.html

[20] J. C. Roh and B. D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[21] T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.

[22] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[23] C. Chen and L. Wang, "Performance analysis of scheduling in multiuser MIMO systems with zero-forcing receivers," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1435–1445, Sep. 2007.

[24] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. Signal Process. Appl.*, 2005, pp. 403–406.

[25] T. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2006, pp. 359–363.

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

[27] V. Shah, N. Mehta, and R. Yim, "Splitting algorithms for fast relay selection: Generalizations, analysis, and a unified view," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1525–1535, Apr. 2010.

[28] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. 23rd Annu. Joint Conf. IEEE Comput. Commun. Soc.*, 2004, vol. 3, pp. 1662–1672.

[29] V. Shah, N. Mehta, and R. Yim, "Optimal timer based selection schemes," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1814–1823, Jun. 2010.

[30] A. Lozano, A. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[31] B. Hassibi and B. Hochwald, "How much training is needed in multi-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[32] P. Dighe, R. Mallik, and S. Jamuar, "Analysis of transmit–receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 694–703, Apr. 2003.

[33] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[34] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.

[35] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[36] E. W. Beyer, *Standard Mathematical Tables and Formulae*. London, U.K.: CRC, 1991.

Authors' photographs and biographies not available at the time of publication.