Scaling-law Optimal Training and Scheduling in the SIMO Uplink

Sugumar Murugesan, Elif Uysal-Biyikoglu, Philip Schniter*
Dept. ECE, The Ohio State University, 2015 Neil Ave, Columbus, OH 43210
{murugess, elif, schniter}@ece.osu.edu

***Abstract*— In fading MIMO channels, there is a tradeoff between the time (or energy) spent gathering CSI and the remaining time in which to transmit data before the channel loses coherence. The tradeoff is more pronounced in multiuser systems as the number of users- hence the number of channel vectors to be estimated- increases, and is inherently coupled with multiuser scheduling. We consider a multiple access block fading channel with coherence time T, n independent users, each with one transmit antenna and the same average power constraint ρ_{avg}, and a base station with M receive antennas and no a priori channel state information. We construct a training-based communication scheme and jointly optimize the training and user selection: we find the optimal number of users to be trained, L_{opt}, and the optimal number to be scheduled for transmission out of those trained, in order to maximize sum rate. The optimal duration of training is shown to be equal to L_{opt} symbol times. We also show the necessary and sufficient condition for training sequences to satisfy. This optimized training-based scheme achieves the same scaling law with increasing SNR as the non-coherent capacity of a single user $n \times M$ MIMO channel: $L_{\text{opt}}\big(1 - \frac{L_{\text{opt}}}{T}\big)\log_2(\rho_{\text{avg}}) + O(1)$ as $\rho_{\text{avg}} \to \infty$, where $L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. We show this is also the scaling law of the sum capacity of the associated non-coherent SIMO uplink, hence our scheme is scaling-law optimal. Finally, the asymptotic behavior of sum rate and throughput per user under increasing n, M or T is explored.**

***Index Terms*– non-coherent capacity, training, muliple-access channel, multiuser scheduling.**

I. Introduction

It is important for multiple-input multiple-output (MIMO) transceivers to be robust to varying degrees of channel state information (CSI.) While large capacity gains are possible with MIMO architectures when the channel response is known at the receiver (see, *e.g.* [1]–[3],) learning the channel often requires the transmitters to allocate some time and energy to send known training sequences to the receiver. When channel variation is slow, hence the coherence time long, learning the channel coefficients may be a good investment of time and energy. On the other hand, when the coherence time is relatively short, there is a tradeoff between how much time (or energy) is used to learn channel coefficients and how much time remains in which to transmit data. This tradeoff has been explored for a single-user MIMO channel by Hassibi and Hochwald [4], where, under some assumptions, the optimal fraction of the coherence interval to be used for training has been found under different values of signal to noise ratio (SNR) and other parameters.

The problem is more challenging in *multiuser* MIMO channels, where training is inherently tied to user selection (scheduling.) The multiuser setting is of practical interest for the design of existing and proposed communication networks such as broadband wireless described by the IEEE 802.16 standard. Thus for concreteness, we will describe the problem in the context of a wireless multiple-antenna uplink, while the results could equally well apply to other MIMO multiple access channels.

Specifically, in this paper, we will address the joint optimization of training and scheduling in a multiple access channel with n users where each user has an average power constraint ρ_{avg}. Each user (transmitter) has a single antenna and the base station (BS) has M antennas. We assume block-fading with a coherence time of T, where the BS knows the channel statistics but has no a priori information about current realizations. We ask the following broad set of questions: *For a given M and T, how much time should be spent on training and how many users should be trained within a coherence interval? How many of those trained should be selected to transmit data? How does the sum capacity scale with the number of users and SNR?*

Our approach is constructive: we design a training scheme where each coherence interval is divided

*This work supported by NSF CAREER grant 237037.

into two phases. In the training phase, a (randomly) selected group of L users send training symbols, upon reception of which the BS estimates their channel vectors. In the data transmission phase, a subset of size $K \leq L$ out of the trained users are scheduled to transmit data. We consider the maximization of sum rate by optimally setting parameters such as the time and power allocated to each phase, and the values of L and K. In order to do this, we obtain a lower bound on the sum rate by extending to the multiple access SIMO channel a non-coherent channel capacity lower bound introduced in [5] and also used in [4].

The high SNR regime is one where a training-based scheme performs best, and consequently this regime is of interest to us. We will show that setting $L = K = L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$ is optimal, resulting in a sum rate (bits/channel use) of $L_{\text{opt}}\left(1 - \frac{L_{\text{opt}}}{T}\right) \log_2(\rho_{\text{avg}}) + \mathrm{O}(1)$ as $\rho_{\text{avg}} \to \infty$. This sum rate has the same rate of increase in SNR as that in a non-coherent, single user, $n \times M$ channel [6]. Also, we show that the sum capacity of the non-coherent uplink also scales at the same rate with SNR, implying that our scheme is scaling-law optimal. The *prelog factor*, $L_{\text{opt}}\left(1 - \frac{L_{\text{opt}}}{T}\right)$, has the physical interpretation as the number of parallel, non-interfering point-to-point channels available for data communication, and happens to be equal to the degrees of freedom of the non-coherent single user $n \times M$ channel [6]. Thus we prove that the non-coherent $n \times M$ uplink channel has the same degrees of freedom as its single user counterpart.

At high SNR, and as the coherence time of the channel grows, we will find that the prelog factor of the sum rate of our scheme approaches $\min(n, M)$, the degrees of freedom available to a multiple access channel with perfect CSI at the receiver. Again, as the SNR, the number of users n or the number of BS antennas grows, the scheduling gain vanishes (i.e., $L = K$ becomes optimal) due to several factors that will be discussed. In this regime, the optimal number of users to be trained (and allowed to transmit) will be shown to be $L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$.

Meanwhile, we will observe that the throughput per user *strictly decreases* with n. This is in contrast to the coherent SIMO multiple access channel where the per user throughput remains a constant for $n \leq M$ and drops with n only for $n > M$.

The organization of the rest of the paper is as follows. We begin by describing the problem setup more precisely in Section II. Next, in Section III, we derive the sum capacity lower bound, to be used as the main performance metric. We discuss the optimal design of various parameters involved in our scheme in Section IV. Section V contains the asymptotic results. This is followed by conclusions in Section VI.

II. Problem Setup

Channel Model: There are n users, each with one antenna and the same average power constraint, ρ_{avg}, and a base station with M antennas. The fading coefficients linking the users to the BS antennas are i.i.d. $\mathcal{CN}(0, 1)$. The channel is block-fading, i.e., the channel coefficients remain constant for a discrete coherence interval $T \geq 2$ after which it changes to an independent realization. The BS does not know the realization of H, but knows its distribution. Noise is Gaussian and independent across receive antennas and time.

We shall restrict our attention to a *training-based* non-coherent communication scheme consistent with the scheme adopted in [4] for a *single user* MIMO channel. According to this scheme, within every coherence interval T, there are two phases: training, followed by transmission. Let $c \in \mathbb{Z}$ be the index of a coherence interval.

Training Phase: In coherence interval c, $L \leq n$ users are allowed to train. Since the BS does not have any information about the current channel state, it chooses the L users on a random or round-robin basis and these users transmit for T_τ symbol times (we assume the existence of a feedback channel on which the BS can inform the users of the selection using negligible time and power.)

Each user transmits a vector of length T_τ, so the vectors transmitted by all L users can be summarized as the training symbol matrix $S_{\tau,c} \in \mathbb{C}^{T_\tau \times L}$ such that $\mathrm{tr}[S_{\tau,c}^* S_{\tau,c}] \leq LT_\tau$ (A^* indicates the Hermitian of the matrix A throughout this paper.) Received signals at each of the M antennas for the duration of training can be written in the form of a matrix $X_{\tau,c} \in \mathbb{C}^{T_\tau \times M}$:

$$X_{\tau,c} = \sqrt{\rho_\tau} S_{\tau,c} H_{\tau,c} + V_{\tau,c}, \qquad (1)$$

where $V_{\tau,c} \in \mathbb{C}^{T_\tau \times M}$ is an AWGN matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries, independent of $H_{\tau,c}$. ρ_τ is the training power level of each of the L active users, giving the total training energy spent by all the active users, in any coherence time, as $\rho_\tau L T_\tau$.

At the end of the training phase, the BS finds the minimum mean square error (MMSE) estimate of $H_{\tau,c}$ as follows:

$$\hat{H}_{\tau,c} = \sqrt{\frac{1}{\rho_\tau}}\Big(\frac{I_L}{\rho_\tau} + S^*_{\tau,c}S_{\tau,c}\Big)^{-1} S^*_{\tau,c}X_{\tau,c}, \qquad (2)$$

with $\tilde{H}_{\tau,c} := H_{\tau,c} - \hat{H}_{\tau,c}$ being the zero mean channel estimation error.

Data Transmission Phase: Having found the channel estimate $\hat{H}_{\tau,c}$, the BS uses it as if it were accurate during the data transmission phase and treats the estimation error as additive noise. It chooses a subset of K users from these L users (according to a performance criterion to be introduced soon.) Let this subset be indexed by i and ρ_d be the data power level of each of the K active users. Then, the received signal on all M antennas during the data transmission phase of length $T_d := T - T_\tau$ can be written as a matrix $X^i_{d,c} \in \mathbb{C}^{T_d\times M}$:

$$\begin{aligned} X^i_{d,c} &= \sqrt{\rho_d}S^i_{d,c}H^i_{d,c} + V^i_{d,c} \\ &= \sqrt{\rho_d}S^i_{d,c}\hat{H}^i_{d,c} + \underbrace{\sqrt{\rho_d}S^i_{d,c}\tilde{H}^i_{d,c} + V^i_{d,c}}_{\bar{V}^i_{d,c}}, \qquad (3) \end{aligned}$$

where $H^i_{d,c} \in \mathbb{C}^{K\times M}$ is constructed from the rows of $H_{\tau,c}$ corresponding to these K users, $S^i_{d,c} \in \mathbb{C}^{T_d\times K}$ is the data symbol matrix that satisfies $\mathrm{E}\,\mathrm{tr}[S^{i*}_{d,c}S^i_{d,c}] \leq KT_d$, and $V^i_{d,c} \in \mathbb{C}^{T_d\times M}$ is an AWGN matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. In (3), $X^i_{d,c}$ has been explicitly written in terms of the MMSE estimate $\hat{H}^i_{d,c} \in \mathbb{C}^{K\times M}$ of (the corresponding portion of) the channel matrix, and $\tilde{H}^i_{d,c} = H^i_{d,c} - \hat{H}^i_{d,c}$, which is the zero-mean channel estimation error. Since all statistical quantities are stationary across the coherence intervals, the suffix c will hereafter be dropped w.l.o.g.

Note that the total data energy spent by all the K active users in any coherence time is $\rho_d K T_d$. Since ρ_{avg} is the average power constraint of each user, with equal total energy (data and training of all users) allotted to all coherence times, the total energy spent in any coherence time is $\rho_{\text{avg}} nT$ thus giving the relation $\rho_{\text{avg}} nT = \rho_d T_d K + \rho_\tau T_\tau L$. Also, by the symmetry of the random/round-robin selection of users, each user ends up spending the same average power, ρ_{avg}.

Note that using the channel estimate as if it were perfect is not necessarily an optimal approach. Nevertheless, the scheme we described, which is an extension of the single-user training-based scheme of [4], is interesting because it is practical, analyzable, and, as will be shown, scaling-law optimal.

In the next section, a capacity lower bound will be presented. This bound will serve as a performance metric upon which we shall study the effects of various parameters like the training sequence (S_τ), the training period (T_τ), power allocation between the training and data phases, the number of users to be trained (L) and the number of users to be allowed to transmit data (K).

III. Performance Metric

The performance metric we will use is a lower bound on the sum capacity of the non-coherent uplink, C_{sum}, which is a straightforward extension of the non-coherent channel capacity lower bound first introduced in [5] and applied to the MIMO channel in [4].

Consider the channel in (3) for one symbol time given by

$$\boldsymbol{x}^i_d = \sqrt{\rho_d}\boldsymbol{s}^i_d\hat{H}^i_d + \bar{\boldsymbol{v}}^i_d. \qquad (4)$$

where $\boldsymbol{s}^i_d$, $\boldsymbol{x}^i_d$ and $\bar{\boldsymbol{v}}^i_d$ correspond to one row (i.e., one channel use) of S^i_d, X^i_d and $\bar{V}^i_d$ in (3) respectively. Let I^i be the mutual information between $\boldsymbol{s}^i_d$ and $\boldsymbol{x}^i_d$ given $\hat{H}^i_d$, i.e., $I(\boldsymbol{s}^i_d;\boldsymbol{x}^i_d|\hat{H}^i_d)$. Then the lower bound is given by, (see Appendix I in [7] for its derivation):

$$\begin{aligned} C_{\text{LB}}(R_{\bar{\boldsymbol{v}}^i_d}, R_{\boldsymbol{s}^i_d}) &= \mathrm{E} \inf_{p_{\bar{\boldsymbol{v}}^i_d},\forall i} \sup_{p_{\boldsymbol{s}^i_d},\forall i} \max_i \frac{T-T_\tau}{T}. \\ &\qquad I^i(p_{\bar{\boldsymbol{v}}^i_d}, p_{\boldsymbol{s}^i_d}, R_{\bar{\boldsymbol{v}}^i_d}, R_{\boldsymbol{s}^i_d}) \qquad (5) \end{aligned}$$

The mutual information I^i has been written as a function of the signal and noise PDFs and also explicitly as a function of the respective correlation matrices $R_{\boldsymbol{s}^i_d}$ and $R_{\bar{\boldsymbol{v}}^i_d}$ which are derived as follows: The signal correlation matrix is $R_{\boldsymbol{s}^i_d} = \mathrm{E}[\boldsymbol{s}^{i*}_d\boldsymbol{s}^i_d]$. Since the users cannot cooperate and since we do not perform power control across space or time (other than multiuser scheduling), $R_{\boldsymbol{s}^i_d} = I_K, \forall i$ (where I_K is the $K \times K$ identity matrix). The correlation matrix of the zero-mean noise, $\bar{\boldsymbol{v}}^i_d$, is given by

$$\begin{aligned} R_{\bar{\boldsymbol{v}}^i_d} &= \mathrm{E}(\sqrt{\rho_d}\boldsymbol{s}^i_d\tilde{H}^i_d + \boldsymbol{v}^i_d)^*(\sqrt{\rho_d}\boldsymbol{s}^i_d\tilde{H}^i_d + \boldsymbol{v}^i_d) \\ &= \rho_d\,\mathrm{E}[\tilde{H}^{i*}_d\tilde{H}^i_d] + I_M. \qquad (6) \end{aligned}$$

For brevity we will refer to this capacity lower bound as C_{LB} hereafter. For the channel in (4), by identifying the best case signal and worst case additive noise (considering the mutual information I^i) as Gaussian ([4], [7]), it can be proved that, with $I_{\text{lb}}^i = \frac{T-T_\tau}{T}\log\det(I_M + \rho_d R_{\bar{\boldsymbol{v}}_d^i}^{-1}\hat{H}_d^{i*}\hat{H}_d^i)$,

$$C_{\text{LB}} \;=\; \mathrm{E}\max_i I_{\text{lb}}^i \tag{7}$$

Also, we show in the Appendix I of [7] that C_{LB} is also a lower bound on the maximum sum rate achievable within the two-phased training scheme described earlier, under worst noise and best signal design conditions. Let us call this rate $R_{\text{worst}}^{\max}$, and record this fact below.

$$C_{\text{LB}} \leq R_{\text{worst}}^{\max} \leq C_{\text{sum}} \tag{8}$$

Note that C_{LB} is influenced by the training sequence used, the energy shared between the training and the data transmission phases and the duration of training. We consider the roles of these parameters and how to set them in the following section.

IV. Parameter Design

Within the training based scheme described in Section II, the following three are design choices: Training sequence S_τ, Training power ρ_τ, and Training period T_τ.

In the light of the analysis in the preceding section, it is tempting to choose these parameters to maximize C_{LB}. However, from (7), the effect of these parameters on the capacity lower bound is highly convoluted. For analytical tractability, we relax the objective function and limit consideration to a certain solution space. In particular, we do the following:

- From (7), $C_{\text{LB}} = \mathrm{E}\max_i I_{\text{lb}}^i \geq \mathrm{E}\, I_{\text{lb}}^q$, for any fixed q. In the rest of this section, $\mathrm{E}\, I_{\text{lb}}^q$, for a fixed q, will be the objective function.
- S_τ is restricted to the class of training sequences that render symmetry in the estimation error variance across the user subsets, i.e., $\sigma^2_{\tilde{H}_d^i} = \sigma^2_{\tilde{H}_d^j}, \forall i,j$ where $\sigma^2_{\tilde{H}_d^i} := \frac{1}{MK}\mathrm{E}\,\mathrm{tr}[\tilde{H}_d^{i*}\tilde{H}_d^i], \forall i$.

Training Sequence, S_τ: We now design the training sequence by identifying an effective SNR term that affects the objective function. Specifically, we proceed by normalizing the noise correlation and the channel estimate matrices in I_{lb}^q as follows: Define $\ddot{R}_{\bar{\boldsymbol{v}}_d^q} := \frac{1}{\sigma^2_{\bar{\boldsymbol{v}}_d^q}} R_{\bar{\boldsymbol{v}}_d^q}$, where

$$\sigma^2_{\bar{\boldsymbol{v}}_d^q} := \frac{1}{M}\mathrm{tr}[R_{\bar{\boldsymbol{v}}_d^q}] = 1 + K\rho_d\sigma^2_{\tilde{H}_d^q}. \tag{9}$$

Let $\ddot{H}_d^q := \frac{1}{\sigma_{\hat{H}_d^q}}\hat{H}_d^q$ with $\sigma^2_{\hat{H}_d^i} := \frac{1}{MK}\mathrm{E}\,\mathrm{tr}[\hat{H}_d^{i*}\hat{H}_d^i], \forall i$. Therefore,

$$\mathrm{E}\, I_{\text{lb}}^q = \frac{T-T_\tau}{T}\mathrm{E}\log\det(I_M + \rho_{\text{eff}}^q \ddot{R}_{\bar{\boldsymbol{v}}_d^q}^{-1}\ddot{H}_d^{q*}\ddot{H}_d^q), \tag{10}$$

where ρ_{eff}^q is the effective SNR for the subset q given by

$$\rho_{\text{eff}}^q = \frac{\rho_d\sigma^2_{\hat{H}_d^q}}{1+K\rho_d\sigma^2_{\tilde{H}_d^q}} = \frac{1}{K}\Big[\frac{1+K\rho_d}{1+K\rho_d\sigma^2_{\tilde{H}_d^q}} - 1\Big] \tag{11}$$

since $\sigma^2_{\tilde{H}_d^q} + \sigma^2_{\hat{H}_d^q} = \sigma^2_{H_d^q} = 1$. As argued in [4], since the training sequence primarily affects the objective function in (10) through ρ_{eff}^q, we choose to maximize ρ_{eff}^q by minimizing $\sigma^2_{\tilde{H}_d^q}$. Let,

$$\sigma^2_{\tilde{H}_\tau} := \frac{1}{ML}\mathrm{E}\,\mathrm{tr}[\tilde{H}_\tau^*\tilde{H}_\tau] = \frac{1}{ML}\sum_{a=1}^{L}\sum_{b=1}^{M}\mathrm{var}[\tilde{H}_\tau]_{a,b} \tag{12}$$

where $\mathrm{var}[\tilde{H}_\tau]_{a,b}$ indicates the variance of the $(a,b)^{th}$ element of $\tilde{H}_\tau$. Observe that, if Q is the number of subsets of K users formed from the L trained users, i.e. $Q = \begin{pmatrix} L \\ K \end{pmatrix}$, then $\sum_{i=1}^{Q}\sigma^2_{\tilde{H}_d^i}MK$ has QKM entries made up of variances of LM elements in the $\tilde{H}_\tau$ matrix. Since the subsets we form are symmetric with respect to all the users and hence to all $\tilde{H}_\tau$ entries, each element has $\frac{QK}{L}$ representations in this summation. Therefore, we have

$$\sum_{i=1}^{Q}\sigma^2_{\tilde{H}_d^i}MK \;=\; \sum_{a=1}^{L}\sum_{b=1}^{M}\mathrm{var}[\tilde{H}_\tau]_{a,b}\frac{QK}{L} \tag{13}$$

and (12) becomes,

$$\sigma^2_{\tilde{H}_\tau} = \frac{1}{Q}\sum_{i=1}^{Q}\sigma^2_{\tilde{H}_d^i} = \sigma^2_{\tilde{H}_d^q}, \tag{14}$$

where the last equality arises from our assumption on S_τ that ensures $\sigma^2_{\tilde{H}_d^i} = \sigma^2_{\tilde{H}_d^j}$ for all i,j. We therefore minimize $\sigma^2_{\tilde{H}_d^q}$ by minimizing $\sigma^2_{\tilde{H}_\tau}$. The following condition on the training sequence is necessary and sufficient for minimizing $\sigma^2_{\tilde{H}_\tau}$ (see Appendix II in [7]):

$$S_\tau^* S_\tau \;=\; T_\tau I_L. \tag{15}$$

Observe that we need $T_\tau \geq L$ to achieve (15). This constraint is intuitive because, during training, every transmission gives us M equations. There are LM

unknowns, thus at least L transmissions are needed in the training phase. With $T_\tau \geq L$, we can prove that,

$$R_{\tilde{H}_d^q} := \mathrm{E}\left[(\mathrm{vec}\,\tilde{H}_d^q)(\mathrm{vec}\,\tilde{H}_d^q)^*\right] = \frac{1}{1+\rho_\tau T_\tau} I_{KM}$$

$$\sigma^2_{\tilde{H}_d^q} = \frac{1}{1+\rho_\tau T_\tau} \qquad (16)$$

Also, since $R_{\tilde{H}_d^q} + R_{\hat{H}_d^q} = I_{KM}$, $R_{\hat{H}_d^q} := \mathrm{E}\left[(\mathrm{vec}\,\hat{H}_d^q)(\mathrm{vec}\,\hat{H}_d^q)^*\right] = \frac{\rho_\tau T_\tau}{1+\rho_\tau T_\tau} I_{KM}$. Thus, $\ddot{H}_d^q = \frac{1}{\sigma_{\hat{H}_d^q}} \hat{H}_d^q$ has independent $\mathcal{CN}(0,1)$ entries. We will use this property later. From (6) and (9),

$$\ddot{R}_{\bar{\boldsymbol{v}}_d^q} = \frac{1}{\sigma^2_{\bar{\boldsymbol{v}}_d^q}}\Big[\frac{\rho_d K}{1+\rho_\tau T_\tau} I_M + I_M\Big] = I_M. \qquad (17)$$

$$\rho_{\text{eff}}^q = \frac{\rho_d \rho_\tau T_\tau}{1+\rho_\tau T_\tau + K\rho_d}. \qquad (18)$$

Note that (16) applies to all q, thus S_τ renders symmetry in the estimation error variance across the user subsets. This is consistent with the assumption we made in the beginning of this section on S_τ. In fact, all the equations from (14) through (18) apply equally well to any subset i leading to $\mathrm{E}\,I^i_{\text{lb}} = \mathrm{E}\,I^j_{\text{lb}}, \forall i,j$. Defining $I_{\text{lb}} := I^q_{\text{lb}}$, the objective function can be rewritten as the following where $\rho_{\text{eff}} = \rho^i_{\text{eff}}, \forall i$:

$$\mathrm{E}\,I_{\text{lb}} = \frac{T-T_\tau}{T}\,\mathrm{E}\log\det(I_M + \rho_{\text{eff}}\ddot{H}_d^{i*}\ddot{H}_d^i) \qquad (19)$$

Power allocation, α: The energy consumed by the active users in any coherence time is composed of the energy used in the training phase and that in the data transmission phase. It is possible to maximize ρ_{eff} by appropriate power allocation between these phases. In each coherence time, the total energy consumed by all the users is $\rho_{\text{avg}} nT = \rho_\tau T_\tau L + \rho_d T_d K$, where, recall that ρ_{avg} is the average power constraint of each user. Let $\rho_d T_d K = \alpha\rho_{\text{avg}} nT$ for some $\alpha \in (0,1]$. Then

$$\rho_{\text{eff}} = \frac{(\rho_{\text{avg}} nT)^2}{T_d K}\frac{\alpha(1-\alpha)}{L+\rho_{\text{avg}} nT - \alpha\rho_{\text{avg}} nT(1-\frac{L}{T_d})} \qquad (20)$$

The value of α that maximizes ρ_{eff} is derived in Appendix III of [7] to be the following:

$$\alpha_{\text{opt}} = \begin{cases} \frac{1}{2} & T_d = L \\ \gamma - \sqrt{\gamma(\gamma-1)} & T_d > L \\ \gamma + \sqrt{\gamma(\gamma-1)} & T_d < L \end{cases}, \quad \text{where}$$

$$\gamma = \frac{L+\rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1-\frac{L}{T_d}]} \qquad (21)$$

The intuition behind (21) will be apparent after we discuss the design of the training period.

Training period design, T_τ: We now derive the training period T_τ that maximizes $\mathrm{E}\,I_{\text{lb}}$. It can be proven (see Appendix IV of [7]) that $\mathrm{E}\,I_{\text{lb}}$ monotonically increases with T_d for $0 < T_d \leq T - L$. From this, combined with the fact that $T_\tau \geq L$ (from the argument following (15)), we conclude the value of T_τ that maximizes $\mathrm{E}\,I_{\text{lb}}$ is $T_{\tau,\text{opt}} = L$.

With $T_\tau = L$, using the result in (21), it can easily be proven that $\rho_\tau L > \rho_{\text{avg}} n > \rho_d K$ when $T_d > L$ and $\rho_\tau L < \rho_{\text{avg}} n < \rho_d K$ when $T_d < L$, thus giving the intuitive physical interpretation that, when more time is spent on data transmission relative to training, less total power should be spent on data and vice versa.

With this, we have optimized the parameters for our scheme with the exception of L and K. To summarize,

Signal Design: Gaussian symbols, i.i.d. across space and time, with variance ρ_d.
Training Period: $T_\tau = L$, where L is the number of users trained.
Training Sequence: Designed such that $S_\tau^* S_\tau = T_\tau I_L$. Since $T_\tau = L$, the standard L dimensional basis vectors (scaled by $\sqrt{\rho_\tau T_\tau}$) can be used as training sequences for the L users. This gives the interesting physical interpretation that, during training phase, each participating user gets exactly one channel use to train its channel.
Power Share: The total energy spent on data is $\rho_d T_d K = \alpha\rho_{\text{avg}} nT$ and the total energy spent on training is $\rho_\tau T_\tau L = (1-\alpha)\rho_{\text{avg}} nT$, where $\alpha = \alpha_{\text{opt}}$ is given in (21).
User Selection Protocol:

- In each coherence time, during the training phase, L users are selected either randomly or by a round-robin technique to train their channel.
- At the BS, after training is complete, a subset, $i^{\max}$, of users is chosen such that $i^{\max} =$

$\arg\max_i I_{lb}^i$ (or to maximize the mutual information if the signal and additive noise distributions are known and non-Gaussian) and scheduled to transmit data, over a low rate feedback channel.

Due to the inherent symmetry established by this protocol, each user gets the same ergodic rate. Since we may be dealing with possibly short coherence times, interleaving of data symbols across coherence intervals may be necessary to achieve the promised ergodic rate. Thus each user maintains a codebook of rate $\frac{T}{n(T-L)} C_{\rm LB}$ and interleaves its codewords across the coherence intervals in which it transmits data.

Before concluding this section, using the designed parameters, we update $C_{\rm LB}$ and $\rho_{\rm eff}$ as follows,

$$C_{\rm LB} = \frac{T-L}{T} \mathrm{E} \max_i \log\det(I_M + \rho_{\rm eff} \ddot{H}_d^{i*} \ddot{H}_d^i) \tag{22}$$

with $\rho_{\rm eff}$ in (20) rewritten as,

$$\rho_{\rm eff} = \begin{cases} \frac{(\rho_{\rm avg} n)^2}{K(1+2\rho_{\rm avg} n)} & T = 2L \\ \frac{\rho_{\rm avg} nT}{K(T-2L)} (\sqrt{\gamma} - \sqrt{\gamma - 1})^2 & T > 2L \\ \frac{\rho_{\rm avg} nT}{K(2L-T)} (\sqrt{-\gamma} - \sqrt{1-\gamma})^2 & T < 2L \end{cases}$$
$$\text{where} \quad \gamma = \frac{L + \rho_{\rm avg} nT}{\rho_{\rm avg} nT} \frac{T-L}{T-2L}. \tag{23}$$

The remaining question is: What are the optimum numbers of users to be trained (L) and allowed to transmit data (K)? We explore this in the following section.

V. Asymptotic Analysis

In this section, we address the design of L and K in regimes where various parameters such as the SNR ($\rho_{\rm avg}$), the number of users in the system (n) and the number of receive antennas (M) are large. We also derive the scaling-law (w.r.t SNR) for the sum capacity of the non-coherent multiuser channel and prove that our scheme is scaling-law optimal.

Theorem 1: With T, n, M fixed,

$$C_{\rm LB} = \frac{T-L}{T} \min(K, M) \log(\rho_{\rm avg}) + O(1) \quad \text{as} \quad \rho_{\rm avg} \to \infty \tag{24}$$

and this rate of increase is maximized when $L = K = L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$.

Proof: We omit the details of the proof (see [7]) in the interest of space. The proof proceeds by performing an Eigen value decomposition of the argument of $\log\det$ in (22) and then bounding the resulting terms w.r.t $\rho_{\rm avg}$. ■

Note that the non-coherent capacity (C) of a single user $n \times M$ channel is derived in [6] as,

$$C = \frac{T-n^*}{T} n^* \log(\rho_{\rm avg}) + O(1) \quad \text{as} \quad \rho_{\rm avg} \to \infty \tag{25}$$

with $n^* = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Also, coding across antennas is not ruled out in deriving the non-coherent capacity of this single user MIMO channel. Therefore C acts as an upper bound to the sum capacity of our multiple access SIMO channel where users *cannot* cooperate. Thus we have the following corollary to Theorem 1.

Corollary 1: With $C_{\rm LB}$ acting as a lower bound and C as an upper bound to the sum capacity ($C_{\rm sum}$) of the non-coherent, multiple access SIMO channel, from Theorem 1 and [6],

$$C_{\rm sum} = \frac{T-n^*}{T} n^* \log(\rho_{\rm avg}) + O(1) \quad \text{as} \quad \rho_{\rm avg} \to \infty \tag{26}$$

giving the non-coherent multiple access SIMO channel the same degrees of freedom as the non-coherent single user MIMO channel. Note that our scheme is thus scaling-law optimal with the same prelog factor as $C_{\rm sum}$.

Now we proceed to analyze how the multiuser scheduling gain behaves as SNR grows. If $C_{\rm LB}(L, K)$ indicates the lower bound in (22), then the baseline case (i.e., no mulituser scheduling) occurs with $L = K$ as,

$$C_{\rm LB}(L, L) = \frac{T-L}{T} \mathrm{E} \log\det(I_M + \rho_{\rm eff} \ddot{H}_d^{i*} \ddot{H}_d^i) \tag{27}$$

with $i = 1$ since we have only one subset now. Following the proof of Theorem 1, we can see that

$$C_{\rm LB}(L, L) = \left(\frac{T - L_{\rm opt}}{T}\right) L_{\rm opt} \log(\rho_{\rm avg}) + O(1) \quad \text{as} \quad \rho_{\rm avg} \to \infty \tag{28}$$

with $L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Thus we see that,

$$\lim_{\rho_{\rm avg} \to \infty} \frac{\max_{L,K} C_{\rm LB}(L, K)}{\max_L C_{\rm LB}(L, L)} = 1 \tag{29}$$

An intuitive explanation for this is: at high SNR, the power gain obtained by exploiting the statistical diversity available within the trained group of users (i.e., with $K < L$) shows *inside* the log function. This gain could not compensate for the loss in the prelog factor (due to $K < L$). Thus as SNR grows, trying to tap the scheduling gain in the system and hence selecting a subset of trained users to transmit data is suboptimal. Hence $K = L$ becomes optimal at high SNR.

Theorem 2: With n, M fixed,

$$C_{\mathrm{LB}} = \min(n, M)\log(\rho_{\mathrm{avg}}) + O(1) \quad \text{as} \quad \rho_{\mathrm{avg}} \to \infty, T \to \infty \tag{30}$$

and $L = K = L_{\mathrm{opt}} = n$.

For details of the proof see [7]. Note that (30) has the same prelog factor as that of the capacity expression of the coherent multiuser uplink [6], [8], i.e., capacity under perfect channel knowledge. As coherence time increases, the sum rate of our scheme approaches the coherent sum rate. This is because, as T grows, the finite training overhead (recall $L \leq n$) becomes negligible. This is illustrated by Fig.1(a). In fact, using an argument similar to that of Corollary 1, we have the quite intuitive result that as $T \to \infty$, and $\rho_{\mathrm{avg}} \to \infty$, the non-coherent sum capacity increases at the same rate as coherent capacity.

$$C_{\mathrm{sum}} = \min(n, M)\log(\rho_{\mathrm{avg}}) + O(1) \quad \text{as} \quad \rho_{\mathrm{avg}} \to \infty, T \to \infty. \tag{31}$$

Theorem 3: With T, M, ρ_{avg} fixed,

$$C_{\mathrm{LB}} = \frac{T-L}{T}\min(K, M)\log(n) + O(1) \quad \text{as} \quad n \to \infty \tag{32}$$

and C_{LB} is maximized when $L = K = L_{\mathrm{opt}} = \min(M, \lfloor\frac{T}{2}\rfloor)$ giving a prelog factor equal to the degrees of freedom of the non-coherent uplink obtained in Corollary 1.

The proof ([7]) follows that of Theorem 1. An interesting physical interpretation is, at high values of n, every time the number of users in the system doubles, the sum rate, in bits per channel use, increases by the channel's degrees of freedom. This is illustrated in Fig.1(b). This is because every additional user to the system brings along its own average power constraint, thus effectively increasing the total SNR. This is unlike the case of a downlink with a total power constraint at the BS that does not increase with the number of users.

The increase in the sum rate with n is not without cost: the per-user throughput monotonically decreases in the number of users, n. The result is made precise in the following theorem (proof found in [7]).

Theorem 4: For fixed M and T, as $\rho_{\mathrm{avg}} \to \infty$, $\frac{C_{\mathrm{LB}}}{n}$, monotonically decreases with n. Similarly, the per user capacity $\frac{C_{\mathrm{sum}}}{n}$ also decreases with n. Since the per user rate of our scheme is sandwiched between $\frac{C_{\mathrm{LB}}}{n}$ and $\frac{C_{\mathrm{sum}}}{n}$, it also decreases with n.

It is instructive to compare this result with the coherent channel case. Here, as $\rho_{\mathrm{avg}} \to \infty$, the per user capacity is $\frac{\min(n,M)}{n}\log(\rho_{\mathrm{avg}}) + O(1)$ [8], [6], which remains constant for $n \leq M$ and starts to decrease with n only when $n > M$. The cost of learning the channel is the sole reason for the monotonic decrease in non-coherent per user capacity versus n. Note also that, as the coherence period (T) of the channel grows, at high SNR, the non-coherent channel's per user capacity resembles that of the coherent channel.

Theorem 5: With T, n, ρ_{avg} fixed,

$$C_{\mathrm{LB}} = \Big(\frac{T-L}{T}\Big)K\log(M) + O(1) \quad \text{as} \quad M \to \infty \tag{33}$$

with the maximum at $L = K = L_{\mathrm{opt}} = \min(n, \lfloor\frac{T}{2}\rfloor)$ giving a prelog factor which is the same as the available degrees of freedom of the non-coherent uplink channel

Proof: The proof (for further details see [7]) proceeds by using the fact that the mutual information of the i^{th} user subset, I^i_{lb} (recall from (7)), converges (in distribution) to a Gaussian [9], as $M \to \infty$, i.e.,

$$I^i_{\mathrm{lb}} \overset{d}{=} \mathcal{N}\Bigg(\Big(\frac{T-L}{T}\Big)K\log(1+\rho_{\mathrm{eff}}M), \Big(\frac{T-L}{T}\Big)^2\frac{K}{M}\log_2^2 e\Bigg), \ \forall i, \ \text{as } M \to \infty \tag{34}$$

Since $C_{\mathrm{LB}} = \mathrm{E}\max_i I^i_{\mathrm{lb}}$, with further manipulation, we prove (33). ■

Note that every time the number of antennas at the receiver doubles, the sum rate (in bits per channel use) increases by the channel's degrees of freedom, as illustrated in Fig.2(a). Also note that as $M \to \infty$, from (34), the variance of the mutual information associated with any subset goes to zero (*channel hardening* [9]) and consequently the scheduling gain disappears (see Fig.2(b)). That is, as M grows, $\max_{L,K} C_{\rm LB}(L,K) - \max_L C_{\rm LB}(L,L)$ converges to zero, where $\max_L C_{\rm LB}(L,L)$ corresponds to the case with no multiuser scheduling. It is interesting to compare this result with the case when $\rho_{\rm avg} \to \infty$ (Theorem 1). There the scheduling gain was still present with increasing SNR, but we found that exploiting it is suboptimal.

VI. Conclusion

We designed a training based communication scheme for a non-coherent SIMO multiple access channel wherein training and user selection are jointly optimized. We established that the non-coherent SIMO multiple access channel has the same degrees of freedom as the non-coherent single user MIMO channel given by $L_{\rm opt}\left(1-\frac{L_{\rm opt}}{T}\right)$, where $L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Further, we proved that our training-based scheme has a prelog factor equal to the above degrees of freedom of the non-coherent SIMO multiple access channel. This implies that our training based scheme is scaling-law optimal. We studied the behavior of the scheme in the asymptotic regime, i.e., when SNR, the number of users or the number of BS antennas grows. The multiuser scheduling gain should not be exploited as SNR grows, whereas as M grows, the scheduling gain vanishes due to channel hardening effects. Consequently, as SNR or the number of BS antennas is high, all the users that are trained must be allowed to transmit, this optimum number being $L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. We also observed that doubling n or M acts in the same way as a 3dB increase in SNR, resulting in an increase in the rate (bits/channel use) by the channel's degrees of freedom. Interestingly, at high SNR, the degrees of freedom available per user in a non-coherent channel monotonically decreases with n for all $n \geq 1$, whereas for a coherent channel, the per user degrees of freedom remains a constant for $n \leq M$ and drops with n only for $n > M$.

Finally, we would like to note that our model contains mathematical similarities to the problem of communication in non-coherent wideband channels: Dividing a wideband channel into many narrowband slots, one can ask questions about how many slots to learn and how many to transmit in. The sub-optimality of spending energy to learn too large a number of subchannels is well known: Medard et al. have shown [5], [10], [11] that non-coherent channel capacity decays due to energy being spread over a wide bandwidth. More recently, Agarwal and Honig [12] considered optimizing the number of frequency slots to train and the power allocation to maximize the rate achievable with a training-based scheme. It may be possible to transport our results and techniques for the SIMO multiple access channel to non-coherent wideband links: for example, insights about the optimum number of users to train and select for transmission may lead to insights in the wideband problem about optimal number of subbands to train and use.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[3] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," in *IEEE Journal on Selected Areas in Communications*, vol. 21, (Invited Paper), pp. 684–702, June 2003.

[4] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple - antenna wireless links?," *IEEE Trans. on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.

[5] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. on Information Theory*, vol. 46, pp. 933–946, May 2000.

[6] L. Zheng and D. N. C. Tse, "Communication on the grassman manifold: A geometric approach to the noncoherent multi-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.

[7] S. Murugesan, E. Uysal-Biyikoglu, and P. Schniter, "Optimization of training and scheduling in the non-coherent MIMO multiple-access channel," submitted for publication, June 2006.

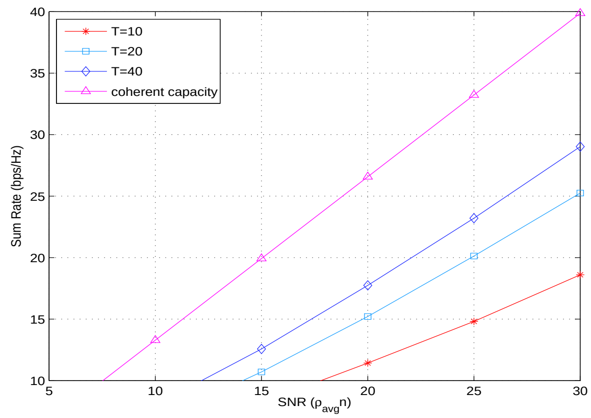
[8] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[9] B. M. Hochwald, T. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Information Theory*, vol. 50, pp. 1893–1909, Sep. 2004.

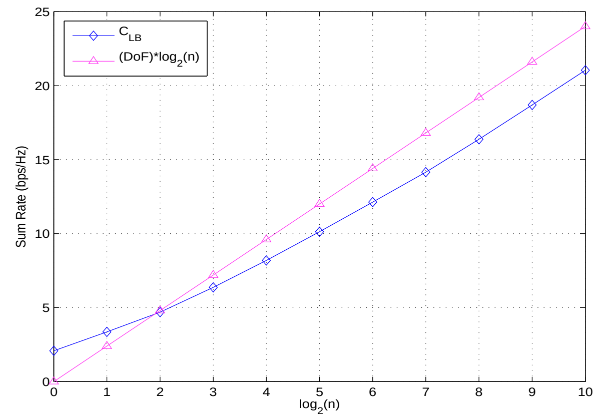
[10] M. Medard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. on Information Theory*, vol. 36, April 2002.

[11] M. Medard and D. N. C. Tse, "Spreading in block-fading channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 1598–1602, 2000.

[12] M. Agarwal and M. L. Honig, "Wideband fading channel capacity with training and limited feedback," in *Proc. Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL, USA), October 2005.

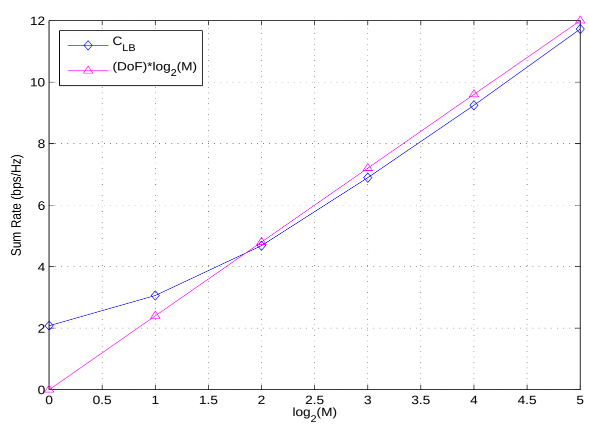


(a)

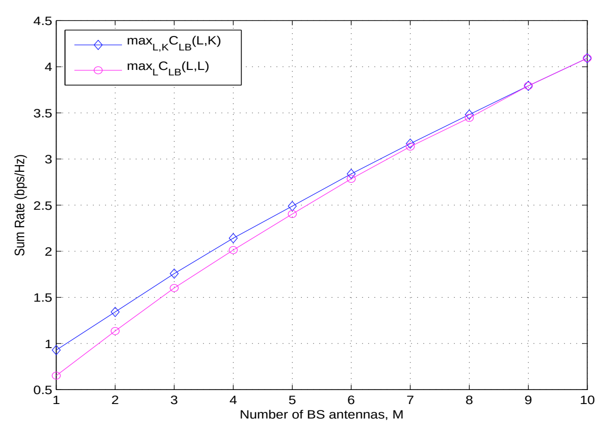


(b)

Fig. 1. (a) Illustration to show how the slope of sum rate achieved by the training-based scheme approaches that of coherent channel capacity when $n = M = 4$. (b) $C_{\rm LB}$ increases by the channel's degrees of freedom (DoF) every time the number of users in the system doubles. $T = 10$, $M = 4$, $\rho_{\rm avg} = 3$ dB used.



(a)



(b)

Fig. 2. (a) As $M \to \infty$, $C_{\rm LB}$ (in bits per channel use) increases by the channel's degrees of freedom (DoF) every time the number of receive antennas at the base station doubles. $T = 10$, $n = 4$, $\rho_{\rm avg} = 3$ dB. (b) Comparison of $\max_{L,K} C_{\rm LB}(L, K)$ with $\max_L C_{\rm LB}(L, L)$, where there is no multiuser scheduling. The scheduling gain vanishes as M grows due to channel hardening effects. Here $n = 8$, $T = 50$, total power $\rho_{\rm avg} n$=1.