Optimization of Training and Scheduling in the Non-Coherent SIMO Multiple Access Channel

Sugumar Murugesan, Elif Uysal-Biyikoglu, and Philip Schniter

***Abstract*— Channel state information (CSI) is important for achieving large rates in MIMO channels. However, in time-varying MIMO channels, there is a tradeoff between the time/energy spent acquiring channel state information (CSI) and the time/energy remaining for data transmission. This tradeoff is accentuated in the MIMO multiple access channel (MAC), since the number of channel vectors to be estimated increases with the number of users. Furthermore, the problem of acquiring CSI is tightly coupled with the problem of exploiting CSI through multiuser scheduling. This paper considers a block-fading MAC with coherence time T, n uncoordinated users—each with one transmit antenna and the same average power constraint, and a base station with M receive antennas and no a priori CSI. For this scenario, a training-based communication scheme is proposed and the training and multiuser-scheduling aspects of the scheme are jointly optimized. In the high-SNR regime, the sum capacity of the non-coherent SIMO MAC is characterized and used to establish the SNR-scaling-law optimality of the proposed scheme. In the low-SNR regime, the sum-rate of the proposed scheme is found to decay linearly with vanishing SNR when flash signaling is incorporated. Furthermore, this linear decay is shown to be order-optimal through comparison to the low-SNR sum capacity of the non-coherent SIMO MAC. A by-product of these SNR-asymptotic analyses is the observation that non-trivial scheduling (i.e., scheduling a strict subset of trained users) is advantageous at low SNR, but not at high SNR. The sum-rate and per-user throughput are also explored in the large-n and large-M regimes.**

***Index Terms*— non-coherent capacity, training, multiple access channel, multiuser scheduling, opportunistic 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/energy is used to acquire CSI and how much time/energy remains for data transmission. This tradeoff has been explored for the single-user MIMO channel by Hassibi and Hochwald [4], where, under some assumptions, the optimal fraction of the coherence interval dedicated to training is derived as a function of signal-to-noise ratio (SNR) and other parameters.

The problem is more challenging in *multiuser* MIMO channels, where training is inherently tied to the problem of scheduling trained users for data transmission (i.e., channel-realization-based opportunistic scheduling). The multiuser setting is of practical interest for the design of existing and proposed communication networks such as IEEE 802.16 broadband wireless. Thus, for concreteness, we pose the problem in the context of a wireless multiple-antenna uplink, although we acknowledge that the results could apply equally well to other applications as well (e.g., communicating over a wideband channel via a subset of narrowband subchannels).

Specifically, this paper addresses the joint optimization of training and scheduling in a multiple access channel (MAC) with n users, where each user has an average power constraint ρ_{avg}. Each user is assumed to have a single antenna, and the base station (BS) has M antennas (SIMO configuration). The channel gains are assumed to be block-fading with a coherence time of T, and the BS is assumed to know a priori the channel statistics but not the channel state. We are interested in answering the following 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 users should be scheduled for data transmission? How does the sum capacity scale with SNR, the number of users, and the number of BS antennas?*

Our approach is constructive: we design a training scheme where each coherence interval is divided into two phases. In the training phase, a (randomly) selected group of L users send training symbols, upon reception of which the BS estimates the corresponding channel gains. In the data transmission phase, a subset $K \leq L$ trained users are scheduled to transmit data. We consider sum-rate maximization through optimization of design parameters such as the time and power allocated between the training and data phases and the values of L and K. Our optimization is performed in accordance with a sum-rate lower bound which can be considered an extension of non-coherent capacity lower bound introduced in [5] and used in [4].

We also study the scheduling gain of the system and the performance of the proposed scheme in several asymptotic regimes. In the high-SNR regime, we show that setting $L = K = L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$ is optimal, resulting in a sum-rate (in bits/channel-use) of $L_{\text{opt}}\Big(1 - \frac{L_{\text{opt}}}{T}\Big) \log_2(\rho_{\text{avg}}) + \text{O}(1)$ as $\rho_{\text{avg}} \to \infty$. The *pre-log factor*, $L_{\text{opt}}\Big(1 - \frac{L_{\text{opt}}}{T}\Big)$, can be interpreted as the equivalent number of parallel, non-

Manuscript received June 1, 2006; revised Jan 1, 2007. This work was supported by NSF CAREER grant 237037 and the Office of Naval Research.
The authors are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA (e-mail: {murugess, elif, schniter}@ece.osu.edu).
Digital Object Identifier 10.1109/JSAC.2007.070917.

0733-8716/07/$25.00 © 2007 IEEE

interfering point-to-point channels available for data communication. Furthermore, this pre-log factor is shown to equal that for the sum capacity of the non-coherent SIMO MAC, thereby establishing the scaling-law optimality of our scheme. In fact, this pre-log factor coincides with that for the capacity of the non-coherent single-user $n \times M$ channel [6], establishing that the non-coherent SIMO MAC has the same degrees-of-freedom as its single-user counterpart. A by-product of this analysis is the observation that non-trivial scheduling (i.e., $K < L$) is not advantageous at high SNR. The asymptotic behaviors of sum-rate in the high-n and high-M regimes are also examined.

In the low-SNR regime (i.e., as SNR goes to zero), we show that, when flash-signaling is incorporated in the proposed scheme, sum-rate varies linearly with vanishing SNR. Since the coherent capacity of the single-user $n \times M$ MIMO channel is known to decay linearly with vanishing SNR ([7], [8]), we conclude the non-coherent SIMO MAC sum capacity also varies linearly with SNR, making the flash version of our proposed scheme order-optimal in the low-SNR regime. Furthermore, we show that non-trivial scheduling (i.e. $K < L$) leads to sum-rate increase when SNR is low, in contrast to our findings when SNR is high. Finally, by considering the coherent SIMO MAC at low SNR, we argue that scheduling can be used to improve sum capacity and that it is optimal to schedule the single best user for data transmission.

The paper is organized as follows. The problem setup is detailed in Section II, and the sum capacity lower bound used as our principle performance metric is described in Section III. Several of our scheme's design parameters are optimized in Section IV, and asymptotic analyses are conducted in Sections V and VI. Finally, Section VII concludes.

II. Problem Setup

Channel Model: There are n users, each with one antenna and the same average power constraint, $\rho_{\rm avg}$, 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 they change to an independent realization. The BS does not know the realization of H, but instead knows its distribution. The 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 the *single-user* MIMO channel. According to this scheme, within every coherence interval T, there are two phases: training, followed by transmission.

Training Phase: In a coherence interval, $L \leq n$ users are allowed to train. Since the BS does not know 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 (assuming the existence of a feedback channel on which the BS can inform the users of the selection using negligible time and power.)

Since each user transmits a vector of length T_τ, the vectors transmitted by all L users can be collected into the training symbol matrix $S_\tau \in \mathbb{C}^{T_\tau \times L}$ such that $\operatorname{tr}[S_\tau^* S_\tau] \leq LT_\tau$, where A^* will be used to indicate the Hermitian transpose of matrix A throughout the paper. The training signals received at the M receiver antennas can be collected into the matrix $X_\tau \in \mathbb{C}^{T_\tau \times M}$, where

$$X_\tau = \sqrt{\rho_\tau} S_\tau H_\tau + V_\tau \tag{1}$$

and where $V_\tau \in \mathbb{C}^{T_\tau \times M}$ is an AWGN matrix with i.i.d. $\mathcal{CN}(0,1)$ entries, independent of the channel matrix $H_\tau \in \mathbb{C}^{L \times M}$. The training power level for each of the L active users is denoted by ρ_τ, so that the total training energy spent by all the active users in a coherence interval is $\rho_\tau L T_\tau$.

At the end of the training phase, the BS computes the minimum mean-square error (MMSE) estimate of H_τ as

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

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

Data Transmission Phase: During the data transmission phase, the BS uses the channel estimate $\hat{H}_\tau$ as if it were perfect and treats the estimation error as additive noise. Furthermore, the BS chooses a subset of K users from these L users according to a performance criterion that will be introduced shortly. Let this subset be indexed by i and let ρ_d denote the data power level of each of the K active users. Then, the signals received on all M antennas during the data transmission phase of length $T_d := T - T_\tau$ can be collected into the matrix $X_d^i \in \mathbb{C}^{T_d \times M}$, where

$$X_d^i = \sqrt{\rho_d} S_d^i H_d^i + V_d^i = \sqrt{\rho_d} S_d^i \hat{H}_d^i + \underbrace{\sqrt{\rho_d} S_d^i \tilde{H}_d^i + V_d^i}_{\tilde{V}_d^i}. \tag{3}$$

In (3), $H_d^i \in \mathbb{C}^{K \times M}$ is constructed from the rows of H_τ corresponding to these K users, $S_d^i \in \mathbb{C}^{T_d \times K}$ denotes the data symbol matrix and satisfies $\operatorname{E} \operatorname{tr}[S_d^{i*} S_d^i] \leq KT_d$, and $V_d^i \in \mathbb{C}^{T_d \times M}$ denotes an AWGN matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. In (3), X_d^i has been explicitly written in terms of the MMSE estimate $\hat{H}_d^i \in \mathbb{C}^{K \times M}$ of H_d^i, where $\tilde{H}_d^i = H_d^i - \hat{H}_d^i$ denotes the zero-mean channel estimation error.

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

Note that using the channel estimate as if it were perfect is not necessarily the 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 parameter choices, such as the training sequence (S_τ), the training period (T_τ), the training/data power allocation, the number of trained users (L) and the number of data-transmitting users (K).

III. Performance Metric

For our performance metric, we choose a lower bound on the sum capacity of the non-coherent uplink, $C_{\rm sum}$, 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, denoted as

$$\boldsymbol{x}_d^i = \sqrt{\rho_d}\boldsymbol{s}_d^i \hat{H}_d^i + \bar{\boldsymbol{v}}_d^i, \quad (4)$$

where $\boldsymbol{s}_d^i$, $\boldsymbol{x}_d^i$ and $\bar{\boldsymbol{v}}_d^i$ correspond to one row (i.e., one channel-use) of S_d^i, X_d^i and $\bar{V}_d^i$ in (3), respectively. Let $p_{\boldsymbol{s}_d^i}$ and $p_{\bar{\boldsymbol{v}}_d^i}$ denote the PDFs of $\boldsymbol{s}_d^i$ and $\bar{\boldsymbol{v}}_d^i$, respectively, and let $R_{\boldsymbol{s}_d^i}$ and $R_{\bar{\boldsymbol{v}}_d^i}$ denote the correlation matrices of $\boldsymbol{s}_d^i$ and $\bar{\boldsymbol{v}}_d^i$, respectively. Since the users cannot cooperate, the transmitted signal is independent across users. Also, since we do not perform power control across space or time (other than multiuser scheduling), $R_{\boldsymbol{s}_d^i} = \mathrm{E}[\boldsymbol{s}_d^{i*}\boldsymbol{s}_d^i] = I_K$ for all i, where I_K is used to denote the $K \times K$ identity matrix. The correlation matrix of the zero-mean noise, $\bar{\boldsymbol{v}}_d^i$, is given by

$$\begin{aligned} R_{\bar{\boldsymbol{v}}_d^i} &= \mathrm{E}(\sqrt{\rho_d}\boldsymbol{s}_d^i\tilde{H}_d^i + \boldsymbol{v}_d^i)^*(\sqrt{\rho_d}\boldsymbol{s}_d^i\tilde{H}_d^i + \boldsymbol{v}_d^i) \\ &= \rho_d\,\mathrm{E}[\tilde{H}_d^{i*}\tilde{H}_d^i] + I_M. \end{aligned} \quad (5)$$

If I^i indicates the mutual information between $\boldsymbol{s}_d^i$ and $\boldsymbol{x}_d^i$ given $\hat{H}_d^i$, i.e., $I(\boldsymbol{s}_d^i;\boldsymbol{x}_d^i|\hat{H}_d^i)$, then the lower bound is given by (see Appendix I)

$$C_{\rm LB} = \mathrm{E}\inf_{p_{\bar{\boldsymbol{v}}_d^i},\forall i}\sup_{p_{\boldsymbol{s}_d^i},\forall i}\max_i \frac{T-T_\tau}{T} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}{=}I_K)$$

subject to the constraint that the transmitted signal is independent across users. The following lemma provides an explicit expression for $C_{\rm LB}$.

Lemma 1:

$$C_{\rm LB} = \frac{T-T_\tau}{T}\mathrm{E}\max_i \log\det(I_M + \rho_d R_{\bar{\boldsymbol{v}}_d^i}^{-1}\hat{H}_d^{i*}\hat{H}_d^i)$$

Proof: Consider a single-user MIMO channel $\boldsymbol{x} = \sqrt{\rho}\boldsymbol{s}\breve{H} + \boldsymbol{v}$, where $\boldsymbol{s} \in \mathbb{C}^{1\times K}$ is the zero mean transmitted signal with autocorrelation matrix $R_{\boldsymbol{s}}$, $\boldsymbol{v} \in \mathbb{C}^{1\times M}$ is zero mean additive noise with autocorrelation matrix $R_{\boldsymbol{v}}$, $\boldsymbol{x} \in \mathbb{C}^{1\times M}$ is the received signal and $\breve{H} \in \mathbb{C}^{K\times M}$ is a known channel matrix. From Theorem 1 in [4], under the constraint $\mathrm{E}[\boldsymbol{s}^*\boldsymbol{v}] = 0_{K\times M}$ and for any $R_{\boldsymbol{s}}$ and $R_{\boldsymbol{v}}$,

$$\inf_{p_{\boldsymbol{v}}}\sup_{p_{\boldsymbol{s}}} I(\boldsymbol{s};\boldsymbol{x}|\breve{H}) = \log\det(I_M + \rho R_{\boldsymbol{v}}^{-1}\breve{H}^* R_{\boldsymbol{s}}\breve{H}), \quad (6)$$

with[1] Gaussian-distributed signal and noise. We now make several observations.

- The result of Theorem 1 in [4], which identifies the Gaussian distribution for worst-case additive noise and best-matched signal, does not change when we restrict the transmitted signal to be independent across transmit antennas with $R_{\boldsymbol{s}} = I_K$. This can be verified by following the proof of Theorem 1 in [4].
- The sum-rate of a MAC with K independent users, each with one transmit antenna, and a BS with M receive antennas, is equal to the mutual information of a single-user $K \times M$ MIMO channel when the transmitted signal is independent across transmit antennas.

From these observations, along with the fact that, for our channel in (4),

$$\begin{aligned} \mathrm{E}[\boldsymbol{s}_d^{i*}\bar{\boldsymbol{v}}_d^i|X_\tau,S_\tau] &= \sqrt{\rho_d}\,\mathrm{E}[\boldsymbol{s}_d^{i*}\boldsymbol{s}_d^i|X_\tau,S_\tau]\,\mathrm{E}[\tilde{H}_d^i|X_\tau,S_\tau] \\ &= 0_{K\times M}, \end{aligned}$$

we conclude that

$$\begin{aligned} &(\mathcal{CN}(0,R_{\bar{\boldsymbol{v}}_d^i}),\mathcal{CN}(0,I_K)) \\ &\quad = \arg\inf_{p_{\bar{\boldsymbol{v}}_d^i}(.)}\sup_{p_{\boldsymbol{s}_d^i}(.)} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, I_K) \end{aligned}$$

with

$$\begin{aligned} &I^i(\mathcal{CN}(0,R_{\bar{\boldsymbol{v}}_d^i}),\mathcal{CN}(0,I_K)) \\ &\quad = \log\det(I_M + \rho_d R_{\bar{\boldsymbol{v}}_d^i}^{-1}\hat{H}_d^{i*}\hat{H}_d^i). \end{aligned} \quad (7)$$

We will henceforth use $I_{\rm lb}^i = \frac{T-T_\tau}{T} I^i(\mathcal{CN}(0,R_{\bar{\boldsymbol{v}}_d^i}),\mathcal{CN}(0,I_K))$ for brevity. Due to symmetry, (7) holds for any subset i, so that

$$C_{\rm LB} = \mathrm{E}\max_i I_{\rm lb}^i. \quad (8)$$

■

We show in Appendix I that $C_{\rm LB}$ is also a lower bound to the maximum sum-rate achievable with the two-phased training scheme described earlier, under worst noise and best signal design conditions. Denoting this rate by $R_{\rm worst}^{\max}$, we state this result as

$$C_{\rm LB} \le R_{\rm worst}^{\max} \le C_{\rm sum}. \quad (9)$$

Note that $C_{\rm LB}$ is influenced by the choice of training sequence, the energy shared between the training and the data transmission phases, and the duration of training. We consider the roles and selection of these parameters in the following section.

IV. Parameter Design

For the training-based scheme described in Section II, the designer must choose the training sequence S_τ, the training power ρ_τ, and the training period T_τ. In light of the preceding section, we would ideally choose these parameters to maximize $C_{\rm LB}$. However, from (8), it can be seen that the effect of these parameters on $C_{\rm LB}$ is highly convoluted. Hence, for tractability, we relax the objective function and limit consideration to a certain solution space. In particular, we do the following.

- Since, from (8), $C_{\rm LB} = \mathrm{E}\max_i I_{\rm lb}^i \ge \mathrm{E}\, I_{\rm lb}^q$ for any fixed q, we choose $\mathrm{E}\, I_{\rm lb}^q$, for a fixed q, to be the objective function for the design of S_τ, ρ_τ and T_τ.
- We restrict S_τ to the class of training sequences that render equal estimation-error variance across user subsets, i.e., $\sigma_{\tilde{H}_d^i}^2 = \sigma_{\tilde{H}_d^j}^2 \;\; \forall i,j$, where $\sigma_{\tilde{H}_d^i}^2 := \frac{1}{MK}\mathrm{E}\,\mathrm{tr}[\tilde{H}_d^{i*}\tilde{H}_d^i]$.

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

[1]Throughout this paper, logarithms are to the base 2, unless mentioned otherwise.

the noise-correlation and channel-estimate matrices in $I_{\rm lb}^q$. To do this, we define $\ddot{R}_{\bar{\boldsymbol v}_d^q} := \frac{1}{\sigma^2_{\bar{\boldsymbol v}_d^q}} R_{\bar{\boldsymbol v}_d^q}$, where

$$\sigma^2_{\bar{\boldsymbol v}_d^q} := \frac{1}{M}\operatorname{tr}[R_{\bar{\boldsymbol v}_d^q}] = 1 + K\rho_d \sigma^2_{\tilde{H}_d^q}. \tag{10}$$

Letting $\ddot{H}_d^q := \frac{1}{\sigma_{\hat{H}_d^q}} \hat{H}_d^q$ with $\sigma^2_{\hat{H}_d^i} := \frac{1}{MK}\operatorname{E}\operatorname{tr}[\hat{H}_d^{i*}\hat{H}_d^i]$, we have

$$\operatorname{E} I_{\rm lb}^q = \frac{T-T_\tau}{T}\operatorname{E}\log\det(I_M + \rho_{\rm eff}^q \ddot{R}^{-1}_{\bar{\boldsymbol v}_d^q}\ddot{H}_d^{q*}\ddot{H}_d^q), \tag{11}$$

where $\rho_{\rm eff}^q$ is the effective SNR for the subset q, given by

$$\rho_{\rm eff}^q = \frac{\rho_d \sigma^2_{\hat{H}_d^q}}{1 + K\rho_d\sigma^2_{\tilde{H}_d^q}} = \frac{1}{K}\Big[\frac{1+K\rho_d}{1+K\rho_d\sigma^2_{\tilde{H}_d^q}} - 1\Big], \tag{12}$$

since $\sigma^2_{\tilde{H}_d^q} + \sigma^2_{\hat{H}_d^q} = \sigma^2_{H_d^q} = 1$. As argued in [4], the training sequence primarily affects the objective function in (11) through $\rho_{\rm eff}^q$. Hence, we choose to maximize $\rho_{\rm eff}^q$ by minimizing $\sigma^2_{\tilde{H}_d^q}$. Let

$$\sigma^2_{\tilde{H}_\tau} := \frac{1}{ML}\operatorname{E}\operatorname{tr}[\tilde{H}_\tau^*\tilde{H}_\tau] = \frac{1}{ML}\sum_{a=1}^{L}\sum_{b=1}^{M}\operatorname{var}[\tilde{H}_\tau]_{a,b}, \tag{13}$$

where $\operatorname{var}[\tilde{H}_\tau]_{a,b}$ 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. $Q = \binom{L}{K}$, then $\sum_{i=1}^{Q}\sigma^2_{\tilde{H}_d^i}MK$ can be considered as the sum of QKM entries, each of which corresponds to the variance of one of the LM elements in the $\tilde{H}_\tau$ matrix. Since the subsets we form are symmetric with respect to all the users, and hence to the entries in $\tilde{H}_\tau$, each variance has $\frac{QK}{L}$ representations in this summation. Therefore, we have $\sum_{i=1}^{Q}\sigma^2_{\tilde{H}_d^i}MK = \sum_{a=1}^{L}\sum_{b=1}^{M}\operatorname{var}[\tilde{H}_\tau]_{a,b}\frac{QK}{L}$, and (13) 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 second equality arises from our assumption on S_τ ensuring $\sigma^2_{\tilde{H}_d^i} = \sigma^2_{\tilde{H}_d^j}$ for all i,j. Thus, we can minimize $\sigma^2_{\tilde{H}_d^q}$ by minimizing $\sigma^2_{\tilde{H}_\tau}$. As shown in Appendix II, the following condition on the training sequence is necessary and sufficient for minimizing $\sigma^2_{\tilde{H}_\tau}$.

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

We observe that (15) requires $T_\tau \geq L$, but note that this constraint is intuitive: during training, every transmission yields M equations, and, since there are LM unknowns, at least L training transmissions are needed. With $T_\tau \geq L$, we prove in Appendix II that

$$R_{\tilde{H}_d^q} := \operatorname{E}\left[(\operatorname{vec}\tilde{H}_d^q)(\operatorname{vec}\tilde{H}_d^q)^*\right] = \frac{1}{1+\rho_\tau T_\tau} I_{KM} \quad \text{and}$$
$$\sigma^2_{\tilde{H}_d^q} = \frac{1}{1+\rho_\tau T_\tau}. \tag{16}$$

Furthermore, since $R_{\tilde{H}_d^q} + R_{\hat{H}_d^q} = I_{KM}$, we find $R_{\hat{H}_d^q} := \operatorname{E}\left[(\operatorname{vec}\hat{H}_d^q)(\operatorname{vec}\hat{H}_d^q)^*\right] = \frac{\rho_\tau T_\tau}{1+\rho_\tau T_\tau} I_{KM}$. 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. Then, from (5) and (10),

$$\ddot{R}_{\bar{\boldsymbol v}_d^q} = \frac{1}{\sigma^2_{\bar{\boldsymbol v}_d^q}}\Big[\frac{\rho_d K}{1+\rho_\tau T_\tau} I_M + I_M\Big] = I_M \quad \text{and}$$
$$\rho_{\rm eff}^q = \frac{\rho_d \rho_\tau T_\tau}{1+\rho_\tau T_\tau + K\rho_d}. \tag{17}$$

Note that (16) applies to any q, and thus S_τ renders symmetry in estimation-error variance across user subsets. This is consistent with the assumption we made in the beginning of this section on S_τ. In fact, equations (14)–(17) apply equally well to any subset i, leading to $\operatorname{E} I_{\rm lb}^i = \operatorname{E} I_{\rm lb}^j, \forall i,j$. Defining $I_{\rm lb} := I_{\rm lb}^q$, the objective function can be rewritten as follows, where $\rho_{\rm eff} = \rho_{\rm eff}^i, \forall i$:

$$\operatorname{E} I_{\rm lb} = \frac{T-T_\tau}{T}\operatorname{E}\log\det(I_M + \rho_{\rm eff}\ddot{H}_d^{i*}\ddot{H}_d^i). \tag{18}$$

Power Allocation, α: The energy consumed by the active users in any coherence time can be decoupled into the energy used in the training phase and that used in the data-transmission phase. Hence, it is possible to maximize $\rho_{\rm eff}$ by appropriate power allocation between these two phases. In each coherence time, the total energy consumed by all the users can be written as $\rho_{\rm avg} nT = \rho_\tau T_\tau L + \rho_d T_d K$, where we recall that $\rho_{\rm avg}$ is the average power constraint of each user. Letting $\rho_d T_d K = \alpha\rho_{\rm avg} nT$ for some $\alpha \in (0,1]$,

$$\rho_{\rm eff} = \frac{(\rho_{\rm avg} nT)^2}{T_d K}\frac{\alpha(1-\alpha)}{L + \rho_{\rm avg}nT - \alpha\rho_{\rm avg}nT(1-\frac{L}{T_d})}. \tag{19}$$

In Appendix III, we show that the value of α maximizing $\rho_{\rm eff}$ is

$$\alpha_{\rm opt} = \begin{cases} \frac{1}{2} & T_d = L \\ \gamma - \sqrt{\gamma(\gamma-1)} & T_d > L \\ \gamma + \sqrt{\gamma(\gamma-1)} & T_d < L \end{cases}, \tag{20}$$

where $\gamma = \frac{L+\rho_{\rm avg}nT}{\rho_{\rm avg}nT[1-\frac{L}{T_d}]}$. The intuition behind (20) will become apparent after we discuss the design of the training period.

Training Period, T_τ: To find the training period T_τ that maximizes $\operatorname{E} I_{\rm lb}$, we first prove (in Appendix IV) that $\operatorname{E} I_{\rm lb}$ increases monotonically with T_d for $0 < T_d \leq T - L$. Combining this with the fact that $T_\tau \geq L$ (recalling the argument following (15)), we conclude that the value of T_τ which maximizes $\operatorname{E} I_{\rm lb}$ is $T_{\tau,\rm opt} = L$.

Using the result in (20) with $T_\tau = L$, it follows that $\rho_\tau L > \rho_{\rm avg} n > \rho_d K$ when $T_d > L$ and that $\rho_\tau L < \rho_{\rm avg} n < \rho_d K$ when $T_d < L$, thus giving the intuitive physical interpretation that, when more time is spent on data transmission than on training, less total power should be spent on data transmission than on training, and vice versa. A similar observation was made in [4].

At this point, we have optimized all design parameters with the exception of L and K. The following summarizes our findings.

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: Any S_τ such that $S_\tau^* S_\tau = T_\tau I_L$. Note that, if standard L-dimensional basis vectors are used as training sequences for the L users, then each participating user would get exactly one channel-use to train its channel.

Power Share: The total energy spent on data should be $\rho_d T_d K = \alpha \rho_{\rm avg} n T$ and that spent on training should be $\rho_\tau T_\tau L = (1-\alpha)\rho_{\rm avg} n T$, for $\alpha = \alpha_{\rm opt}$ given in (20).

User Selection Protocol:

- In each coherence time, L users are selected either randomly or by round-robin to train their channel.
- After the training is complete, the BS schedules the users in subset $i^{\max} = \arg\max_i I_{\rm lb}^i$ to transmit data. Note that user-scheduling can be accomplished with a low-rate feedback channel, since, for each trained user, one bit suffices to indicate whether that user should transmit data or not. Note also that, for scheduling, we have resumed the use of the original objective function $C_{\rm LB}$ from (8); the surrogate objective function, $\mathrm{E}\, I_{\rm lb}^q$ for fixed q, was used only to facilitate the designs of S_τ, ρ_τ and T_τ, which were otherwise intractable.

The user-symmetry of our protocol ensures that the achievable ergodic rate is the same for all users. When the coherence time is brief, achieving these ergodic rates may require interleaving the data symbols across coherence intervals. In this case, each user would maintain a codebook of rate $\frac{T}{n(T-L)} C_{\rm LB}$ and interleave its codewords across the coherence intervals for which it is scheduled to transmit data.

We can now rewrite the expressions for $C_{\rm LB}$ and $\rho_{\rm eff}$ using optimized design parameters:

$$C_{\rm LB} = \frac{T-L}{T}\, \mathrm{E} \max_i \log\det(I_M + \rho_{\rm eff} \ddot{H}_d^{i*} \ddot{H}_d^i) \quad (21)$$

$$\rho_{\rm eff} = \begin{cases} \frac{(\rho_{\rm avg} n)^2}{K(1+2\rho_{\rm avg} n)} & T = 2L \\ \frac{\rho_{\rm avg} n T}{K(T-2L)}(\sqrt{\gamma} - \sqrt{\gamma-1})^2 & T > 2L \\ \frac{\rho_{\rm avg} n T}{K(2L-T)}(\sqrt{-\gamma} - \sqrt{1-\gamma})^2 & T < 2L \end{cases}$$
$$\text{where} \quad \gamma = \frac{L + \rho_{\rm avg} n T}{\rho_{\rm avg} n T}\, \frac{T-L}{T-2L}. \quad (22)$$

As for protocol design, the question remains: How many users should be trained (L) and allowed to transmit data (K)? We seek answers to this question in the following two sections.

V. Asymptotic Analysis – Part I

In this section, we address the design of L and K in regimes where 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 SIMO multiple access channel and prove that our scheme is scaling-law optimal.

Theorem 1: With T, n, M fixed,

$$C_{\rm LB} = \frac{T-L}{T} \min(K,M) \log(\rho_{\rm avg}) + O(1) \quad \text{as } \rho_{\rm avg} \to \infty \quad (23)$$

and this rate of increase is maximized when $L = K = L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Thus, non-trivial scheduling (i.e., $K < L$) is sub-optimal at high SNR.

Proof: With $i_{\max} = \arg\max_i \log\det(I_M + \rho_{\rm eff} \ddot{H}_d^{i*} \ddot{H}_d^i)$, and using (21),

$$\begin{aligned} C_{\rm LB} &= \frac{T-L}{T}\, \mathrm{E} \log\det(I_M + \rho_{\rm eff} \ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}}) \\ &= \frac{T-L}{T}\, \mathrm{E} \log \prod_{j=1}^{\min(K,M)} (1 + \rho_{\rm eff} \lambda_j^{i_{\max}}), \end{aligned} \quad (24)$$

where $\lambda_j^{i_{\max}}$ denotes the j^{th} non-zero eigenvalue of $\ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}}$. As $\rho_{\rm avg} \to \infty$, from (22),

$$\frac{\rho_{\rm eff}}{\rho_{\rm avg}} = \frac{n}{K(\sqrt{1-\frac{L}{T}} + \sqrt{\frac{L}{T}})^2} \quad (25)$$

since $\gamma = \frac{T-L}{T-2L}$, and $\frac{\log(1+\rho_{\rm eff}\lambda_j^{i_{\max}})}{\log(\rho_{\rm eff}\lambda_j^{i_{\max}})} = 1$. Thus,

$$\begin{aligned} C_{\rm LB} &= \frac{T-L}{T} \min(K,M) \log(\rho_{\rm avg}) \\ &+ \underbrace{\frac{T-L}{T}\, \mathrm{E} \sum_{j=1}^{\min(K,M)} \log \lambda_j^{i_{\max}}}_{A} \\ &+ \underbrace{\frac{T-L}{T} \min(K,M) \log\Big(\frac{n}{K(\sqrt{1-\frac{L}{T}} + \sqrt{\frac{L}{T}})^2}\Big)}_{B} \\ &+ O(1) \quad \text{as} \quad \rho_{\rm avg} \to \infty \end{aligned} \quad (26)$$

Notice that A in (26) can be upper bounded as $\mathrm{E} \sum_{j=1}^{\min(K,M)} \log \lambda_j^{i_{\max}} \le \mathrm{E} \max_i \sum_{j=1}^{\min(K,M)} \log \lambda_j^i$. The upper bound is finite with probability 1, since K and L are upper-bounded by n and since M and n are fixed. Thus A does not grow unbounded with $\rho_{\rm avg}$. Similarly, B is also bounded as $\rho_{\rm avg}$ increases. Putting these facts together, we have

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

It then can be verified that the pre-log factor $\frac{T-L}{T} \min(K,M)$ is maximized when $L = K = L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$, which completes the proof. ∎

It is instructive to compare this result with that of the coherent SIMO MAC. While it is strictly sub-optimal to perform non-trivial scheduling at high SNR in the non-coherent SIMO MAC, it can be readily verified that, in the coherent SIMO MAC at high SNR, it is optimal to schedule *any* $K \ge \min(n, M)$ users to transmit data and that non-trivial scheduling does not offer any additional gain.

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

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

where $n^* = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Recall also that 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 MAC, where the users *cannot* cooperate. Thus we have the following corollary to Theorem 1.

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

$$C_{\text{sum}} = \frac{T - n^*}{T} n^* \log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty \quad (27)$$

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 pre-log factor as C_{sum}.

Now we proceed to analyze how the gain obtained from multiuser scheduling behaves as SNR grows. Using $C_{\text{LB}}(L, K)$ to denote the lower bound in (21) with explicit dependence on L and K, the baseline case (i.e., trivial scheduling) corresponds to $K = L$ and gives $C_{\text{LB}}(L, L) = \frac{T-L}{T} \operatorname{E} \log \det(I_M + \rho_{\text{eff}} \ddot{H}_d^{i*} \ddot{H}_d^i)$ with $i = 1$ since, in this case, there exists only one user subset (i.e., $Q = 1$). Following the proof of Theorem 1, we can see that

$$C_{\text{LB}}(L, L) = \Big(\frac{T - L_{\text{opt}}}{T}\Big) L_{\text{opt}} \log(\rho_{\text{avg}}) + O(1) \text{ as } \rho_{\text{avg}} \to \infty$$

with $L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Thus, we see that $\lim_{\rho_{\text{avg}} \to \infty} \frac{\max_{L,K} C_{\text{LB}}(L,K) - \max_L C_{\text{LB}}(L,L)}{\max_L C_{\text{LB}}(L,L)} = 0$ where we call $\frac{\max_{L,K} C_{\text{LB}}(L,K) - \max_L C_{\text{LB}}(L,L)}{\max_L C_{\text{LB}}(L,L)}$ the Fractional Scheduling Gain. To interpret this result, notice that, at high SNR, the power gain obtained through non-trivial scheduling (i.e., $K < L$) shows *inside* the log function. This gain cannot compensate for the loss in the pre-log factor that results from $K < L$. Thus, as SNR grows, non-trivial scheduling (i.e., choosing a strict subset of trained users for data transmission) becomes sub-optimal.

Theorem 2: With T, M, and ρ_{avg} fixed,

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

and C_{LB} is maximized when $L = K = L_{\text{opt}} = \min(M, \lfloor \frac{T}{2} \rfloor)$.

Proof: The proof follows that of Theorem 1. From (22), as $n \to \infty$,

$$\frac{\rho_{\text{eff}}}{n} = \frac{\rho_{\text{avg}}}{K(\sqrt{1 - \frac{L}{T}} + \sqrt{\frac{L}{T}})^2} \quad (28)$$

since $\gamma = \frac{T-L}{T-2L}$. Thus, from (24),

$$\begin{aligned} C_{\text{LB}} = & \frac{T - L}{T} \min(K, M) \log(n) \\ & + \underbrace{\frac{T - L}{T} \min(K, M) \log\Big(\frac{\rho_{\text{avg}}}{K(\sqrt{1 - \frac{L}{T}} + \sqrt{\frac{L}{T}})^2}\Big)}_{B} \\ & + \underbrace{\frac{T - L}{T} \operatorname{E} \sum_{j=1}^{\min(K,M)} \log \lambda_j^{i_{\max}}}_{A} + O(1) \quad \text{as} \quad n \to \infty \end{aligned}$$

Here, A is bounded using an argument similar to that from the $\rho_{\text{avg}} \to \infty$ case. Also, since $K \le L = T_\tau \le T - 1$ (where the last inequality is due to the necessity of reserving at least one symbol time for data transmission), B is bounded w.r.t n. Thus, as the number of users in the system grows, C_{LB} is maximized when $L = K = L_{\text{opt}} = \min(M, \lfloor \frac{T}{2} \rfloor)$, giving a pre-log factor equal to the degrees-of-freedom of the non-coherent uplink obtained in Corollary 1. ■

An interesting physical interpretation is, when n is large, 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 behavior is explained by the fact that each additional user increases the total system power, i.e., that expended by the $L = K = \min(M, \lfloor \frac{T}{2} \rfloor)$ active users, thereby increasing the effective average SNR. This is unlike the case of a downlink with a total-power constraint at the BS, where the total power does not increase with the number of users.

The increase in the sum-rate with n is not without cost, however, since the per-user throughput monotonically decreases with the number of users, n. The result is made precise in the following theorem.

Theorem 3: For fixed M and T, as $\rho_{\text{avg}} \to \infty$, $\frac{C_{\text{LB}}}{n}$ monotonically decreases with n. Similarly, the per-user capacity $\frac{C_{\text{sum}}}{n}$ also decreases with n. Since the per-user rate of our scheme is sandwiched between $\frac{C_{\text{LB}}}{n}$ and $\frac{C_{\text{sum}}}{n}$, it also decreases with n.

Proof: From (23), $\frac{C_{\text{LB}}}{n}$ is given by,

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

For any value of n, $\frac{C_{\text{LB}}}{n}$ is maximized when $L = K = \min(n, M, \lfloor \frac{T}{2} \rfloor)$ from Theorem 1. We now consider two cases. Case (i): $n \ge \min(M, \lfloor \frac{T}{2} \rfloor)$. Here, $\frac{C_{\text{LB}}}{n} = \Big(\frac{T - \min(M, \lfloor \frac{T}{2} \rfloor)}{T}\Big)\Big(\frac{\min(M, \lfloor \frac{T}{2} \rfloor)}{n}\Big) \log(\rho_{\text{avg}}) + O(1)$, which monotonically decreases with n and achieves the maximum of $\Big(\frac{T - \min(M, \lfloor \frac{T}{2} \rfloor)}{T}\Big) \log(\rho_{\text{avg}}) + O(1)$ at $n = \min(M, \lfloor \frac{T}{2} \rfloor)$. Case (ii): $1 \le n < \min(M, \lfloor \frac{T}{2} \rfloor)$. Here, $\frac{C_{\text{LB}}}{n} = \Big(\frac{T-n}{T}\Big) \log(\rho_{\text{avg}}) + O(1)$, which also monotonically decreases with n and achieves the maximum of $\Big(\frac{T-1}{T}\Big) \log(\rho_{\text{avg}}) + O(1)$ at $n = 1$.

Considering these two cases, it is evident that $\frac{C_{\text{LB}}}{n}$ monotonically decreases with n for all $n \ge 1$. Since, from Corollary 1, the non-coherent sum capacity has the same pre-log factor as C_{LB}, the per-user *capacity*, $\frac{C_{\text{sum}}}{n}$, also monotonically decreases with n. Hence, the per-user rate of our scheme, which is lower bounded by $\frac{C_{\text{LB}}}{n}$ and upper bounded by $\frac{C_{\text{sum}}}{n}$, must also monotonically decrease with n. ■

It is instructive to compare this result with the coherent channel case. There, as $\rho_{\text{avg}} \to \infty$, the per-user capacity is $\frac{\min(n,M)}{n} \log(\rho_{\text{avg}}) + O(1)$ [6], [9], which remains constant for $n \le M$ and decreases *only* when $n > M$. Thus, the cost incurred in learning the channel (which, in our scheme, results from the channel-uses lost in training) accounts for the monotonic decrease in non-coherent per-user capacity versus $n \ge 1$.

Theorem 4: With T, n, and ρ_{avg} fixed,

$$C_{\text{LB}} = \left(\frac{T-L}{T}\right) K \log(M) + O(1) \quad \text{as} \quad M \to \infty, \tag{29}$$

and the maximum occurs at $L = K = L_{\text{opt}} = \min(n, \lfloor \frac{T}{2} \rfloor)$.

Proof: Recall, from (8) and (17), and with $T_\tau = T_{\tau,\text{opt}} = L$, that $I_{\text{lb}}^i = \frac{T-L}{T} \log\det(I_M + \rho_{\text{eff}} \ddot{H}_d^{i*} \ddot{H}_d^i)$, from which $C_{\text{LB}} = \mathrm{E}\max_i I_{\text{lb}}^i$. Since $\ddot{H}_d^i \in \mathbb{C}^{K\times M}$ is made up of i.i.d. $\mathcal{CN}(0,1)$ elements (as was noted before (17)), I_{lb}^i converges (in distribution and for all i) to a Gaussian [10] as $M \to \infty$:

$$I_{\text{lb}}^i \overset{d}{=} \mathcal{N}\left(\left(\frac{T-L}{T}\right) K \log(1+\rho_{\text{eff}} M), \left(\frac{T-L}{T}\right)^2 \frac{K}{M} \log_2^2 e\right) \quad \text{as} \quad M \to \infty. \tag{30}$$

Using the fact that $\log(1+\rho_{\text{eff}}M) \to \log(\rho_{\text{eff}}M)$ as $M \to \infty$, and defining $X_1, \ldots, X_Q$ as $\mathcal{CN}(0,1)$ random variables (with Q defined after (13)), we have

$$C_{\text{LB}} = \left(\frac{T-L}{T}\right) K \log(\rho_{\text{eff}} M) + \left(\frac{T-L}{T}\right)\sqrt{\frac{K}{M}} (\log_2 e)\, \mathrm{E}(\max_{i=1,\ldots,Q} X_i) + O(1) \quad \text{as} \quad M \to \infty.$$

Since $\mathrm{E}(\max_{i=1,\ldots,Q} X_i)$ is bounded w.r.t M (because Q is bounded w.r.t M), we have

$$C_{\text{LB}} = \left(\frac{T-L}{T}\right) K \log(M) + O(1) \quad \text{as} \quad M \to \infty.$$

The limiting value of C_{LB} is maximized when $L = K = L_{\text{opt}} = \min(n, \lfloor \frac{T}{2} \rfloor)$, giving a pre-log factor that equals the available degrees-of-freedom of the non-coherent uplink channel. ■

Note that, every time the receiver antenna doubles, the sum-rate (in bits per channel-use) increases by the channel's degrees-of-freedom. Also, Theorem 4 suggests that the fractional scheduling gain vanishes as $M \to \infty$, i.e., $\lim_{M\to\infty} \frac{\max_{L,K} C_{\text{LB}}(L,K) - \max_L C_{\text{LB}}(L,L)}{\max_L C_{\text{LB}}(L,L)} = 0$. This behavior can be interpreted as follows: note from (30) that, as $M \to \infty$, the variance of the mutual information associated with any subset goes to zero (i.e., *channel hardening* [10]) and consequently the scheduling gain vanishes making it optimal to schedule all the users trained to transmit data. It is interesting to compare this result with the case when $\rho_{\text{avg}} \to \infty$ (Theorem 1). There the scheduling gain was still present with increasing SNR, but we found that exploiting it is sub-optimal.

VI. Asymptotic Analysis – Part II

We now proceed to analyze the low-SNR regime, i.e., $\rho_{\text{avg}} \to 0$. For this case, we address the design of L and K, the behavior of the scheduling gain, and the overall performance of our scheme.

From (19), we have

$$\begin{aligned}\rho_{\text{eff}} &= \frac{\rho_{\text{avg}}^2 (nT)^2 \alpha(1-\alpha)}{T_d K L} \frac{1}{1 + \frac{\rho_{\text{avg}} nT}{L}(1 - \alpha(1 - \frac{L}{T_d}))} \\ &= \frac{\rho_{\text{avg}}^2 (nT)^2 \alpha(1-\alpha)}{T_d K L} + \rho_{\text{avg}}^3 c_3 + \rho_{\text{avg}}^4 c_4 + \ldots \end{aligned} \tag{31}$$

where the second equation follows from the binomial expansion with c_3, c_4, etc. being functions of n, T, L and α. Thus, from (24), using $\log_e(1+x) = \sum_{k=1}^{\infty} \frac{-(-x)^k}{k}$,

$$\begin{aligned} C_{\text{LB}} &= \frac{T_d}{T} \mathrm{E} \sum_{j=1}^{\min(K,M)} \log(1 + \rho_{\text{eff}} \lambda_j^{i_{\max}}) \\ &= \frac{T_d}{T} \log e\, \mathrm{E} \sum_{j=1}^{\min(K,M)} \sum_{k=1}^{\infty} \frac{-(-\rho_{\text{eff}} \lambda_j^{i_{\max}})^k}{k} \\ &= \frac{T_d}{T} \log e \sum_{k=1}^{\infty} \frac{-(-\rho_{\text{eff}})^k}{k} \mathrm{E} \sum_{j=1}^{\min(K,M)} (\lambda_j^{i_{\max}})^k \\ &= \frac{T_d}{T} \log e \frac{\rho_{\text{avg}}^2 (nT)^2 \alpha(1-\alpha)}{T_d K L} \mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}}) \\ &\quad + o(\rho_{\text{avg}}^2) \quad \text{as} \quad \rho_{\text{avg}} \to 0 \end{aligned} \tag{32}$$

The last equation follows from the fact that $\forall k \geq 1$, $\mathrm{E} \sum_{j=1}^{\min(K,M)} (\lambda_j^{i_{\max}})^k \leq \mathrm{E} \max_i \sum_{j=1}^{\min(K,M)} (\lambda_j^i)^k$ which is finite with probability 1 (see proof of Theorem 1) and that the quantities c_3, c_4, etc. can be upper bounded likewise. We are interested in analyzing how fast C_{LB} approaches zero as $\rho_{\text{avg}} \to 0$. Note from (32) that the scheduling rule can be written as

$$i_{\max} = \arg\max_i \mathrm{tr}(\ddot{H}_d^{i*} \ddot{H}_d^i) \tag{33}$$

and the optimum value of $\alpha = \alpha_{\text{opt}} = \frac{1}{2}$. We update (32) as follows,

$$C_{\text{LB}} = \rho_{\text{avg}}^2 \frac{\log e\, n^2 T}{4KL} \mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}}) + o(\rho_{\text{avg}}^2) \text{ as } \rho_{\text{avg}} \to 0. \tag{34}$$

Since $\ddot{H}_d^i$ is made up of independent $\mathcal{CN}(0,1)$ entries (recalling Section IV), $\mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}})$ remains invariant to ρ_{avg} when L and K are fixed. Thus, the sum-rate of the proposed training-based scheme decays quadratically with vanishing SNR. This can be contrasted with the single-user $n \times M$ MIMO channel, where the coherent capacity decays linearly with vanishing SNR [7], [8]. Thus, the proposed scheme is potentially rate-of-decay sub-optimal in the low-SNR regime. In the sequel, we propose a minor modification that fixes the rate-of-decay problem. First, however, we analyze how L and K should be chosen to improve the scaling factor on the dominant SNR term in (34), i.e., how scheduling gain can be leveraged. The result is summarized in the following theorem.

Theorem 5: With T, n, and M fixed, $L_{\text{opt}} = K_{\text{opt}} = 1$ as $\rho_{\text{avg}} \to 0$ and

$$C_{\text{LB}} = \rho_{\text{avg}}^2 \frac{\log e\, n^2 TM}{4} + o(\rho_{\text{avg}}^2) \quad \text{as} \quad \rho_{\text{avg}} \to 0.$$

Proof: Let $\ddot{\boldsymbol{h}}^k \in \mathbb{C}^{1\times M}$ denote the channel estimate vector (normalized to unit variance) for user $k \in \{1, \ldots, L\}$. If the set $\mathcal{S}_i$ contains the indices of the K unique users associated with subset index i, then $\mathrm{tr}(\ddot{H}_d^{i*} \ddot{H}_d^i) = \sum_{k\in\mathcal{S}_i} \|\ddot{\boldsymbol{h}}^k\|^2$. Let the users be ordered by channel norm $\|\ddot{\boldsymbol{h}}^k\|^2$, where the index k_l corresponds to the user with $(L-l+1)^{th}$ largest channel norm (i.e., users k_L and k_1 have the largest and the smallest channel norms respectively). Since $\mathrm{tr}(\ddot{H}_d^{i_{\max}*} \ddot{H}_d^{i_{\max}}) =$

$\sum_{k\in\mathcal{S}_{i_{\max}}} \|\ddot{\boldsymbol{h}}^k\|^2 = \sum_{l=L-K+1}^{L} \|\ddot{\boldsymbol{h}}^{k_l}\|^2$, we have that

$$\mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}) \;=\; \sum_{l=L-K+1}^{L} \mu_{l:L}, \qquad (35)$$

where $\mu_{l:L} = \mathrm{E}\|\ddot{\boldsymbol{h}}^{k_l}\|^2$ denotes the first moment of the l^{th} order-statistic of the random variables $\{\|\ddot{\boldsymbol{h}}^k\|^2\}_{k=1}^L$. The dominant-term scaling factor from (34) becomes

$$\frac{\log e\; n^2T}{4KL}\,\mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}) = \frac{\log e\; n^2T}{4L}\,\frac{\sum_{l=L-K+1}^{L}\mu_{l:L}}{K}$$

Since $\mu_{a:L} \leq \mu_{b:L}$ for $a < b$, the above equation is maximized when $K = 1$ for any fixed values of n, T and L. Thus, among the L trained users, it is optimal to schedule only one for data transmission. We now proceed to find the optimal number of users to train when $K = 1$.

Using $\ddot{h}_i^k$ to indicate the i^{th} element of $\ddot{\boldsymbol{h}}^k$, we see that $\ddot{h}_i^k \sim \mathcal{CN}(0,1)$ and that $\ddot{h}_i^k$ is independent of $\ddot{h}_j^k$ for $j \neq i$. Thus, $|\ddot{h}_i^k|^2$ is exponentially distributed with unit mean and variance, so that $\mathrm{E}\|\ddot{\boldsymbol{h}}^k\|^2 = M$ and $\mathrm{var}(\|\ddot{\boldsymbol{h}}^k\|^2) = M$. With these findings, the scaling factor can be bounded as [11, p. 79]

$$\begin{aligned}\frac{\log e\; n^2T}{4L}M \;\leq\; & \frac{\log e\; n^2T}{4L}\,\mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}) \\ \leq\; & \frac{\log e\; n^2T}{4}\left(\frac{M}{L} + \sqrt{M}\Big(\frac{L-1}{L^2}\Big)^{\frac{1}{2}}\right) \qquad (36)\end{aligned}$$

It can be easily verified that both the upper and the lower bounds in (36) monotonically decrease with L and coincide at $L = 1$. Thus, at low SNR, it is optimal to train a single user (chosen randomly or by round-robin) and allow that same user to transmit data, i.e., to use time-division multiple access (TDMA). With $L_{\mathrm{opt}} = K_{\mathrm{opt}} = 1$, it is clear that $\mathrm{E}\,\mathrm{tr}(\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}) = \mathrm{E}\|\ddot{\boldsymbol{h}}^1\|^2 = M$, from which the theorem follows. ■

For fixed L, the rule $K_{\mathrm{opt}} = 1$ is intuitive because, given the power-limited nature of the system, we expect an advantage in allocating full power to the user with the best channel. Also, it is intuitive that C_{LB} increases linearly with the number of BS antennas, M, because received signal power increases linearly with the number of BS antennas. The intuition behind the rule $L_{\mathrm{opt}} = 1$ can be gleaned from the effective SNR expressions (12) and (16). As the number of trained users increases, the energy available to train each user decreases, thereby reducing the channel estimate quality, and hence the effective SNR. With $K_{\mathrm{opt}} = 1$, the reduction in effective SNR is not compensated by an increase in scheduling gain, though.

We now propose a simple "*flash*" version of our training-based scheme that significantly improves behavior in the low-SNR regime. The motivation for flash signaling comes from the observation (see, e.g., [8], [12]) that, in single-user non-coherent channels, flashy signaling is needed to achieve a linear decay in throughput as SNR approaches zero. In "flash signaling," we keep the transmitter silent for all but a small fraction of the coherence intervals, allowing the effective SNR to increase during signal flashes (for the same time-averaged SNR). More precisely, let $\delta \in (0, 1]$ indicate the fraction of the coherence intervals for which the transmitter is active. Conditioning on these *active* coherence intervals, the total average energy (for training or data transmission) is $\frac{\rho_{\mathrm{avg}}}{\delta}nT$ and the average per-user power constraint is $\rho'_{\mathrm{avg}} = \frac{\rho_{\mathrm{avg}}}{\delta}$. With flash signaling, the sum-rate lower bound (21) becomes

$$C_{\mathrm{LB}} \;=\; \delta\frac{T-L}{T}\,\mathrm{E}\max_i \log\det(I_M + \rho'_{\mathrm{eff}}\ddot{H}_d^{i*}\ddot{H}_d^i), \quad (37)$$

where ρ'_{eff} is defined as the value of ρ_{eff} in (22) when ρ_{avg} is replaced by ρ'_{avg}. Notice the pre-log factor δ in (37) that accounts for the fraction of active channel-uses.

Focusing on the low-SNR regime (where $\rho_{\mathrm{avg}} < 1$), we set $\delta = \rho_{\mathrm{avg}}$, in which case $\rho'_{\mathrm{avg}} = 1$ and

$$C_{\mathrm{LB}} \;=\; \rho_{\mathrm{avg}}\frac{T-L}{T}\,\mathrm{E}\max_i \log\det(I_M + \rho'_{\mathrm{eff}}\ddot{H}_d^{i*}\ddot{H}_d^i) \quad (38)$$

Since $\frac{T-L}{T}\,\mathrm{E}\max_i \log\det(I_M + \rho'_{\mathrm{eff}}\ddot{H}_d^{i*}\ddot{H}_d^i)$ is independent of ρ_{avg}, it is evident that C_{LB} decays linearly with ρ_{avg} as $\rho_{\mathrm{avg}} \to 0$. Since the coherent capacity of the single-user $n\times M$ MIMO channel also decays linearly with vanishing SNR, we conclude that the sum capacity of the non-coherent SIMO MAC decays linearly with vanishing SNR and that our flash-based modification is order-optimal. Noticing that the scaling factor on the ρ_{avg} term in (38) equals that of the sum-rate lower bound C_{LB} in normal (i.e., not-flashy) operation, when the per-user power constraint $\rho_{\mathrm{avg}} = 1$, we expect a non-negligible improvement in this scaling factor when scheduling gain is exploited, i.e., when non-trivial scheduling is performed. This claim is supported in Fig. 1 where the fractional scheduling gain, given by $\frac{\max_{L,K} C_{\mathrm{LB}}(L,K) - \max_L C_{\mathrm{LB}}(L,L)}{\max_L C_{\mathrm{LB}}(L,L)}$ is plotted against ρ_{avg} in the non-flashy mode of operation. Here, for each value of ρ_{avg}, the optimum L and K were obtained numerically for both $C_{\mathrm{LB}}(L,K)$ and $C_{\mathrm{LB}}(L,L)$ using the C_{LB} expression (21).

Finally, considering the coherent SIMO MAC at low SNR, it can be proven (though we omit the proof in the interest of space) that the sum capacity decays linearly with vanishing SNR and that the SNR scaling factor can be increased by non-trivial scheduling (i.e., scheduling based on actual channel realizations). Here, too, it can be shown that only the *single best user* should be scheduled for data transmission.

To recap, we have shown that, at low SNR, throughput can be increased by non-trivial scheduling in both the non-coherent channel (via the flash-modified training-based scheme) and the coherent channel. At high SNR, however, non-trivial scheduling was found to be not advantageous.

VII. CONCLUSION

We proposed a training-based communication scheme for the non-coherent SIMO MAC, whereby L (out of n) users are trained, out of which K are scheduled for data transmission. The scheduling is conducted in accordance with a mutual-information lower bound. We then optimized the scheme's design parameters, namely, the training sequence (S_τ), the training period (T_τ), the power allocation between training and data phases (α), the number of trained users (L), and the number of scheduled users (K). Finally, we analyzed the performance of our scheme in several asymptotic regimes, such as when the SNR is very high or very low, when the number of users is very large, or when the number of BS antennas is very large.

We summarize the high-SNR results first. In this regime, we found it optimal to train $L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$ users and schedule *all* trained users for data transmission; non-trivial scheduling (i.e., $K < L$) was found to be sub-optimal. Next, we established that the non-coherent SIMO MAC has the same degrees-of-freedom as the non-coherent single-user $n \times M$ MIMO channel: $L_{\rm opt}(1 - \frac{L_{\rm opt}}{T})$, where $L_{\rm opt} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Since the pre-log factor of our training-based scheme was found to coincide with these degrees-of-freedom, we can state that our scheme is scaling-law optimal. We also observed that doubling either the number of users (n) or BS antennas (M) has the same effect as a 3 dB increase in SNR; both increase the throughput (in bits/channel-use) by the channel's degrees-of-freedom. However, the per-user throughput of the non-coherent SIMO MAC was found to decrease monotonically with n for all $n \geq 1$, in contrast to the per-user throughput of the coherent SIMO MAC, which remains constant for $n \leq M$ and drops only for $n > M$.

We now summarize the low-SNR results. When flash signaling is used with our training-based scheme, we showed that the sum-rate decays linearly with SNR as SNR goes to zero. Since the coherent capacity of the single user $n \times M$ MIMO channel also decays linearly with vanishing SNR, we can state that the sum capacity of the non-coherent SIMO MAC decays linearly with vanishing SNR and that the flash version of our scheme is order-optimal at low SNR. (Without flash signaling, the sum-rate of our scheme decays quadratically with vanishing SNR.) Furthermore, we demonstrated that non-trivial scheduling improves the sum-rate in the low-SNR regime. Finally, we reasoned that the sum capacity of the coherent SIMO MAC benefits from non-trivial scheduling and, in particular, scheduling the single best user is optimal at low SNR. Thus, for both coherent and non-coherent SIMO MACs, we established that non-trivial scheduling is advantageous at low SNR, but not at high SNR.

We note that our model contains mathematical similarities to the problem of communication in non-coherent wideband channels. In particular, when partitioning a wideband channel into many narrowband subchannels, one might optimize the number of subchannels to learn and the number to transmit through. The sub-optimality of learning too many subchannels is well known. Medard et al. have shown [5], [13], [14] that non-coherent channel capacity decays due to energy being spread over a wide bandwidth. More recently, Agarwal and Honig [15] 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 MAC 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.

APPENDIX I
CAPACITY BOUNDS

Let $I_o^i(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i}) = \frac{1}{T_d} I(S_d^i; X_d^i | \hat{H}_d^i)$ be the mutual information (normalized per channel-use) between the input and the output of the channel in (3), where $p_{S_d^i}$ and $p_{\bar{V}_d^i}$ are the signal and noise distributions respectively, with autocorrelation matrices given by $R_{S_d^i}$ and $R_{\bar{V}_d^i}$. Let $I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}) = I(\boldsymbol{s}_d^i; \boldsymbol{x}_d^i | \hat{H}_d^i)$ be the mutual information of the channel $\boldsymbol{x}_d^i = \sqrt{\rho_d} \boldsymbol{s}_d^i \hat{H}_d^i + \bar{\boldsymbol{v}}_d^i$, where $\boldsymbol{x}_d^i$ and $\boldsymbol{s}_d^i$ correspond to the input and output of the channel in (3) considering one channel-use, $p_{\boldsymbol{s}_d^i}$ and $p_{\bar{\boldsymbol{v}}_d^i}$ are the distributions of the signal and noise for this channel with $R_{\boldsymbol{s}_d^i}$ and $R_{\bar{\boldsymbol{v}}_d^i}$ denoting their autocorrelation matrices, respectively. Let $S_d^{i,t}$ corresponds to the input symbol at t^{th} channel-use and let S_d^{i,t_1} and S_d^{i,t_2} be generated i.i.d. when $t_1 \neq t_2$. Then, for any signal and additive noise distribution, with $h(X)$ denoting the differential entropy of a random variable X, by the definition of mutual information,

$$
\begin{aligned}
I_o^i\left(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i}\right) &= \frac{1}{T_d}\left(h(S_d^i | \hat{H}_d^i) - h(S_d^i | X_d^i, \hat{H}_d^i)\right) \\
&= \frac{1}{T_d}\Big(h(S_d^i) - \sum_{t=T_\tau+1}^{T} h(S_d^{i,t} | S_d^{i,T_\tau+1} .. S_d^{i,t-1}, X_d^i, \hat{H}_d^i)\Big) \\
&\geq h(\boldsymbol{s}_d^i) - h(\boldsymbol{s}_d^i | \boldsymbol{x}_d^i, \hat{H}_d^i) \\
&= I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i})
\end{aligned} \tag{39}
$$

where the inequality comes from the fact that conditioning reduces entropy. Let $C_{\rm sum}$ indicate the sum capacity of the non-coherent channel. Let $R_{\rm worst}^{\max}$ be the maximum sum-rate achieved by the training-based scheme described in Section II under worst noise and best signal design conditions subject to the following constraints: transmitted signal S_d^i is independent across users and the correlation matrix $R_{S_d^i}$ is such that $R_{\boldsymbol{s}_d^i} = I_K$. Then we have

$$
\begin{aligned}
C_{\rm sum} &\geq R_{\rm worst}^{\max} \\
&:= \mathrm{E} \inf_{p_{\bar{V}_d^i}, \forall i} \sup_{p_{S_d^i}, \forall i} \max_i \frac{T_d}{T} I_o^i(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i})
\end{aligned}
$$

From now on, we additionally constrain S_d^i to be i.i.d across time, so that we have,

$$
R_{\rm worst}^{\max} \geq \mathrm{E} \inf_{p_{\bar{V}_d^i}, \forall i} \sup_{p_{S_d^i}, \forall i} \max_i \frac{T_d}{T} I_o^i(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i})
$$

Fix the noise distribution corresponding to the inf of the above expression as $p_{\bar{V}_d^i}^1$. Let $p_{\bar{\boldsymbol{v}}_d^i}^1$ be the marginal distribution of noise (i.e., considering one channel-use) corresponding to $p_{\bar{V}_d^i}^1$. Then, from the inequality proved in (39),

$$
\begin{aligned}
&\mathrm{E} \inf_{p_{\bar{V}_d^i}, \forall i} \sup_{p_{S_d^i}, \forall i} \max_i \frac{T_d}{T} I_o^i(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i}) \\
&\geq \mathrm{E} \sup_{p_{\boldsymbol{s}_d^i}, \forall i} \max_i \frac{T_d}{T} I^i(p_{\bar{\boldsymbol{v}}_d^i}^1, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i} = I_K) \\
&\geq \mathrm{E} \inf_{p_{\bar{\boldsymbol{v}}_d^i}, \forall i} \sup_{p_{\boldsymbol{s}_d^i}, \forall i} \max_i \frac{T_d}{T} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, I_K) \ =: \ C_{\rm LB}.
\end{aligned}
$$

Thus we have $C_{\rm sum} \geq R_{\rm worst}^{\max} \geq C_{\rm LB}$.

APPENDIX II
TRAINING SEQUENCE DESIGN

Following the analysis in [4],

$$\begin{aligned} R_{\tilde{H}_\tau} &:= \mathrm{E}\left[(\mathrm{vec}\,\tilde{H}_\tau)(\mathrm{vec}\,\tilde{H}_\tau)^*\right] \\ &= R_{H_\tau} - R_{H_\tau X_\tau} R_{X_\tau}^{-1} R_{X_\tau H_\tau} \\ &= \left(I_L + \rho_\tau S_\tau^* S_\tau\right)^{-1} \otimes I_M \end{aligned} \tag{40}$$

where $\mathrm{vec}(A)$ operator stacks all of the columns of A into one long column. From (13),

$$\sigma^2_{\tilde{H}_\tau} = \frac{1}{ML}\,\mathrm{tr}\,R_{\tilde{H}_\tau} = \frac{\mathrm{tr}\left[\left(I_L + \rho_\tau S_\tau^* S_\tau\right)^{-1}\right]}{L}. \tag{41}$$

Thus,

$$\begin{aligned} \min \quad \sigma^2_{\tilde{H}_\tau} &= \min_{S_\tau, \mathrm{tr}\, S_\tau^* S_\tau \le L T_\tau} \frac{1}{L}\,\mathrm{tr}\left[\left(I_L + \rho_\tau S_\tau^* S_\tau\right)^{-1}\right] \\ &= \min_{\lambda_1,\ldots,\lambda_L, \sum_j \lambda_j \le L T_\tau} \frac{1}{L}\sum_{j=1}^{L} \frac{1}{1+\rho_\tau \lambda_j} \end{aligned}$$

where, $\lambda_1, \ldots, \lambda_L$ are the eigenvalues of $S_\tau^* S_\tau$. The minimum is achieved when $\lambda_j = T_\tau \quad \forall j \in 1, \ldots, L$. Thus, we design the training sequence based on the following condition: $S_\tau^* S_\tau = T_\tau I_L$. With this condition, (40) and (41) yield $R_{\tilde{H}_\tau} = \frac{1}{1+\rho_\tau T_\tau} I_{LM}$, $\sigma^2_{\tilde{H}_\tau} = \frac{1}{1+\rho_\tau T_\tau}$ and

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

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

APPENDIX III
POWER SHARE DESIGN

In this proof and in the preliminary part of the next proof, we closely follow the analysis done in [4], where the authors have dealt with a similar design problem for a single-user MIMO channel with $L = K$. We design the optimum value of α that maximizes ρ_{eff} in (19), considering the following three cases,

When $T_d = L$: $\rho_{\mathrm{eff}} = \frac{(\rho_{\mathrm{avg}} nT)^2 \alpha(1-\alpha)}{LK(L+\rho_{\mathrm{avg}} nT)}$ with $\alpha_{\mathrm{opt}} = \frac{1}{2}$. Thus, $\rho_{\mathrm{eff}} = \frac{(\rho_{\mathrm{avg}} nT)^2}{4LK(L+\rho_{\mathrm{avg}} nT)}$.

When $T_d > L$: $\rho_{\mathrm{eff}} = \frac{\rho_{\mathrm{avg}} nT \alpha(1-\alpha)}{(T_d - L)K[-\alpha+\gamma]}$ where $\gamma := \frac{L+\rho_{\mathrm{avg}} nT}{\rho_{\mathrm{avg}} nT[1-\frac{L}{T_d}]} > 1$. Since $\frac{d\rho_{\mathrm{eff}}}{d\alpha}|_{\alpha=\alpha_{\mathrm{opt}}} = 0$, $\alpha_{\mathrm{opt}} = \gamma - \sqrt{\gamma(\gamma-1)}$. Thus $\rho_{\mathrm{eff}} = \frac{\rho_{\mathrm{avg}} nT}{K(T_d - L)}(\sqrt{\gamma} - \sqrt{\gamma - 1})^2$.

When $T_d < L$: In this case, with $\gamma = \frac{L+\rho_{\mathrm{avg}} nT}{\rho_{\mathrm{avg}} nT[1-\frac{L}{T_d}]} < 0$, following the previous steps, $\alpha_{\mathrm{opt}} = \gamma + \sqrt{-\gamma(1-\gamma)}$ and $\rho_{\mathrm{eff}} = \frac{\rho_{\mathrm{avg}} nT}{K(L-T_d)}(\sqrt{-\gamma} - \sqrt{1-\gamma})^2$.

In summary, α_{opt} is given by (20).

APPENDIX IV
TRAINING PERIOD DESIGN

Under the condition $T_d \le T - L$, and assuming for the moment, that T_d is a continuous variable, we now prove that $\mathrm{E}\, I_{\mathrm{lb}}$ monotonically increases with T_d, which then holds for discrete T_d too. If λ_j^i is the j^{th} non-zero eigenvalue of the matrix $\ddot{H}_d^{i*} \ddot{H}_d^i$, then, from (18),

$$\begin{aligned} \mathrm{E}\, I_{\mathrm{lb}} &= \frac{T_d}{T}\,\mathrm{E}\log \prod_{j=1}^{\min(K,M)} (1 + \rho_{\mathrm{eff}} \lambda_j^i) \\ &= \frac{T_d}{T} \min(K,M)\,\mathrm{E}\log(1+\rho_{\mathrm{eff}}\lambda) \end{aligned}$$

where the expectation in the last equality is with respect to the non-zero eigenvalue λ given by, $\lambda = \lambda_j^i$ with probability $\frac{1}{\min(K,M)}, \forall j$. Thus, $\frac{d(\mathrm{E}\, I_{\mathrm{lb}})}{dT_d} = \frac{\min(K,M)}{T}\,\mathrm{E}\log(1+\rho_{\mathrm{eff}}\lambda) + \frac{\min(K,M) T_d}{T}\frac{d\rho_{\mathrm{eff}}}{dT_d}\,\mathrm{E}\left[\frac{\lambda}{1+\rho_{\mathrm{eff}}\lambda}\right]$.

When $T_d > L$: Using ρ_{eff} from Appendix III, $\frac{d\rho_{\mathrm{eff}}}{dT_d} = \frac{\rho_{\mathrm{eff}}}{(T_d - L)}\left(\sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}} - 1\right)$. Therefore,

$$\begin{aligned} \frac{d(\mathrm{E}\, I_{\mathrm{lb}})}{dT_d} &= \frac{\min(K,M)}{T}\,\mathrm{E}\Big[\log(1+\rho_{\mathrm{eff}}\lambda) - \frac{\rho_{\mathrm{eff}} T_d}{(T_d - L)} \\ &\qquad \times \left(1 - \sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}}\right)\frac{\lambda}{1+\rho_{\mathrm{eff}}\lambda}\Big]. \end{aligned}$$

Since $T_d > L$, $\frac{T_d}{T_d - L}\left(1 - \sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}}\right) < 1$ and since $\lambda > 0$, $\log(1+\rho_{\mathrm{eff}}\lambda) - \frac{\rho_{\mathrm{eff}}\lambda}{1+\rho_{\mathrm{eff}}\lambda} > 0$. Combining these, $\frac{d(\mathrm{E}\, I_{\mathrm{lb}})}{dT_d} > 0$ when $T_d > L$.

When $T_d < L$: $\frac{d\rho_{\mathrm{eff}}}{dT_d} = \frac{\rho_{\mathrm{eff}}}{(L - T_d)}\left(1 - \sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}}\right)$ and

$$\begin{aligned} \frac{d(\mathrm{E}\, I_{\mathrm{lb}})}{dT_d} &= \frac{\min(K,M)}{T}\,\mathrm{E}\Big[\log(1+\rho_{\mathrm{eff}}\lambda) - \frac{\rho_{\mathrm{eff}} T_d}{(T_d - L)} \\ &\qquad \times \left(1 - \sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}}\right)\frac{\lambda}{1+\rho_{\mathrm{eff}}\lambda}\Big]. \end{aligned}$$

It can be proven that $\frac{T_d}{T_d - L}\left(1 - \sqrt{\frac{L(L+\rho_{\mathrm{avg}} nT)}{T_d(T_d+\rho_{\mathrm{avg}} nT)}}\right) < 1$ for $T_d < L$ too. Now, using the same argument as in the previous case, $\frac{d(\mathrm{E}\, I_{\mathrm{lb}})}{dT_d} > 0$ when $T_d < L$. We proceed to prove that $\mathrm{E}\, I_{\mathrm{lb}}$ is continuous at $T_d = L$, thus proving $\mathrm{E}\, I_{\mathrm{lb}}$ monotonically increases with T_d for all T_d.

Denoting ρ_{eff} by $\rho_{\mathrm{eff}}(T_d)$ and defining t_m as $\lim_{m\to\infty} t_m = L$, using L'Hospital's rule, for both $T_d < L$ and $T_d > L$ regions, we can prove, $\lim_{m\to\infty} \rho_{\mathrm{eff}}(t_m) = \rho_{\mathrm{eff}}(L)$. Since $\rho_{\mathrm{eff}} \le \rho_d, \forall T_d$ (from (17)),

$$\log(1+\rho_{\mathrm{eff}}(t_m)\lambda) f_\lambda(\lambda) \le \log(1+\rho_d\lambda) f_\lambda(\lambda)$$

where $f_\lambda(\lambda)$ is the p.d.f of λ. Also, since

$$\begin{aligned} \int_\lambda \log(1+\rho_d\lambda) f_\lambda(\lambda) d\lambda &= \frac{\mathrm{E}(\log\det(I_M + \rho_d \ddot{H}_d^{i*}\ddot{H}_d^i))}{\min(K,M)} \\ &\le \frac{\log\det(I_M + \rho_d\,\mathrm{E}(\ddot{H}_d^{i*}\ddot{H}_d^i))}{\min(K,M)} < \infty \end{aligned}$$

using the Dominated Convergence Theorem [16],

$$\begin{aligned} &\lim_{m\to\infty} \mathrm{E}\log(1+\rho_{\mathrm{eff}}(t_m)\lambda) \\ &= \lim_{m\to\infty}\int_\lambda \log(1+\rho_{\mathrm{eff}}(t_m)\lambda) f_\lambda(\lambda) d\lambda \\ &= \int_\lambda \log(1+\lim_{m\to\infty}\rho_{\mathrm{eff}}(t_m)\lambda) f_\lambda(\lambda) d\lambda \\ &= \mathrm{E}\log(1+\rho_{\mathrm{eff}}(L)\lambda) \end{aligned}$$

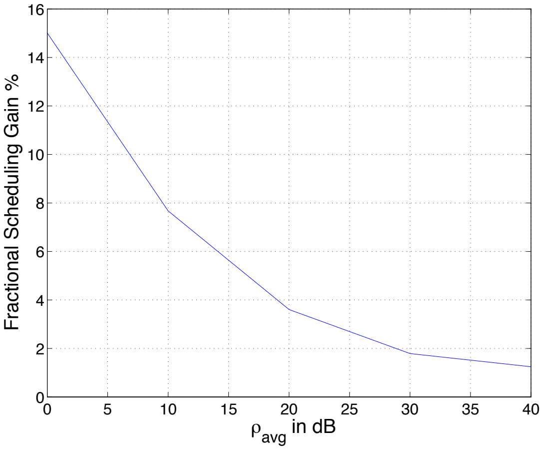


Fig. 1. Fractional scheduling gain $\frac{\max_{L,K} C_{\rm LB}(L,K) - \max_L C_{\rm LB}(L,L)}{\max_L C_{\rm LB}(L,L)}$ as a function of average per-user power $\rho_{\rm avg}$ (in dB). The optimum L and K were obtained numerically for each value of $\rho_{\rm avg}$ for both $C_{\rm LB}(L,K)$ and $C_{\rm LB}(L,L)$, using the $C_{\rm LB}$ expression (21) with $n = 10$, $T = 50$, and $M = 2$. A non-negligible scheduling gain can be observed when $\rho_{\rm avg} = 1$, supporting the claim made in Section VI.

Thus $\mathrm{E}\, I_{\rm lb} = \frac{T_d}{T} \min(K,M)\, \mathrm{E} \log(1 + \rho_{\rm eff}\lambda)$ is continuous at $T_d = L$. Hence, from our previous monotonicity results in $T_d < L$ and $T_d > L$ ranges, we conclude $\mathrm{E}\, I_{\rm lb}$ monotonically increases with T_d, for all T_d.

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 J. Select. Areas Commun.*, 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. Inform. 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. Inform. 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 multiantenna channel," *IEEE Trans. Inform. Theory,* vol. 48, pp. 359-383, Feb. 2002.

[7] A. Lozano, A. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inform. Theory,* vol. 49, pp. 2527-2544, Oct. 2003.

[8] C. Rao and B. Hassibi, "Analysis of Multiple-Antenna Wireless Links at Low SNR," *IEEE Trans. Inform. Theory,* vol. 50, pp. 2123-2130, Sep. 2004.

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

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

[11] H. A. David, *Order Statistics.* New York: John Wiley and Sons, Inc., 1981.

[12] L. Zheng, M. Medard, D. Tse, and C. Luo, "On Channel Coherence in the Low SNR Regime," in *Proc. Allerton Conf. on Communication, Control, and Computing,* Oct. 2003.

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

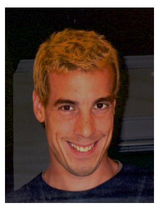
[14] M. Medard and D. N. C. Tse, "Spreading in block-fading channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers,* vol. 2, pp. 1598-1602, 2000.

[15] 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.

[16] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications.* Wiley-Interscience, 1999.

Sugumar Murugesan received the B.E. degree in Electronics and Communication Engineering from the College of Engineering, Anna University, India in 2004. He received the M.S. degree in Electrical and Computer Engineering from the Ohio State University (OSU), USA in 2006. He is a recipient of the OSU University Fellowship (2004-05). He is currently working towards the Ph.D. degree in Electrical and Computer Engineering at OSU. His current research interests include multiuser communication and queueing theory.

Elif Uysal-Biyikoglu received the B.S. degree from the Middle East Technical University (METU), Ankara, Turkey in 1997, the S.M. degree in Electrical Engineering and Computer Science from Massachusetts Institute of Technology (MIT) in 1999, and the Ph.D. degree in Electrical Engineering from Stanford University in 2003. From 2003 to 2005 she was at MIT as a Postdoctoral Lecturer of EECS and at the Research Laboratory for Electronics. From 2005 to 2006 she was an Assistant Professor of Electrical and Computer Engineering at the Ohio State University. She is currently an Adjunct Assistant Professor at OSU and Assistant Professor at METU. Dr. Uysal-Biyikoglu is a recipient of the Vinton Hayes Fellowship at MIT, the Stanford Graduate Fellowship, a 2006-2009 NSF research grant and the Turkish National Science Foundation Young Faculty Career Development Award. Her research interests are at the junction of communication and networking theories.



Philip Schniter received the B.S. and M.S. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively. From 1993 to 1996 he was employed by Tektronix Inc. in Beaverton, OR as a systems engineer. In 2000, he received the Ph.D. degree in Electrical Engineering from Cornell University in Ithaca, NY. Subsequently, he joined the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, OH, where he is now an Associate Professor. In 2003, he received the National Science Foundation CAREER Award, and he currently serves on the IEEE Signal Processing for Communications Technical Committee. Dr. Schniter's research interests include signal processing, communication theory, and wireless networks.