Training and Scheduling in the Non-Coherent MIMO Uplink

A Thesis

Presented in Partial Fulfillment of the Requirements for
the Degree Master of Science in the
Graduate School of The Ohio State University

By

Sugumar Murugesan, B.E.

* * * * *

The Ohio State University

2006

Master's Examination Committee:

Philip Schniter, Adviser

Elif Uysal-Biyikoglu

Approved by

____

Adviser
Graduate Program in
Electrical Engineering

ABSTRACT

Channel state information (CSI) is important for achieving large rates in MIMO channels. However, in time-varying MIMO channels, there is a trade-off between the time (or energy) spent gathering CSI and the remaining time in which to transmit data before the channel loses coherence. This trade-off is accentuated in the MIMO multiple-access channel (MAC) as the number of users, thus the number of channel vectors to be estimated, increases. Furthermore, the problem is inherently coupled with multiuser scheduling. In this paper, 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. Our 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 MIMO MAC, 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.

Dedicated to my dear father late Mr. K. R. Murugesan.

Without his words of inspiration, I would not have reached this milestone of my life.

ACKNOWLEDGMENTS

I am indebted to my family for their continued encouragement and for enduring the pain of separation from me throughout my Master's studies.

My heartfelt thanks to my adviser Prof. Philip Schniter. His patience, understanding and mentorship made this thesis a reality. He taught me how to tackle even the most complicated problems by looking into the big picture. Special thanks to my co-adviser Prof. Elif Uysal-Biyikoglu for being highly supportive in completing my thesis. It was a great experience learning Queueing theory from her.

My lab-mates Arun, Praveen, Sib, Aditi, Arul, Kambiz, Lifeng, Iris, Sung-Jun, Young-Han made me feel at home away from home during my stay at IPS. These guys are a lot of fun to work with!

This material is based upon work supported by the National Science Foundation under Grant No. 0237037.

VITA

June 4, 1983	Born - Sivagangai, India
2004	B.E. Electronics and Communication Engineering
2004-2005	University Fellow, The Ohio State University
2005-present	Graduate Research/Teaching Associate, OSU.

PUBLICATIONS

Journal Publications

S. Murugesan, E. Uysal-Biyikoglu and P. Schniter "Optimization of Training and Scheduling in the Non-Coherent MIMO Multiple-Access Channel," submitted June 2006 to *IEEE Journal on Selected Areas in Communication.*

Conference Publications

S. Murugesan, E. Uysal-Biyikoglu and P. Schniter "Scaling-law optimal training and scheduling in the MIMO uplink," to appear in *Proc. Allerton Conf. on Communication, Control, and Computing,* (Monticello, IL), Oct. 2006.

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in:

Digital Signal Processing, Communication Theory	Prof. Philip Schniter
Information Theory	Prof. Hesham El Gamal
Queueing Theory	Prof. Elif Uysal-Biyikoglu

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

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 trade-off between how much time (or energy) is used to learn channel coefficients and how much time remains in which to transmit data. This trade-off 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.

1.2 Contributions

In this thesis, 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 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 MIMO 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 $L_{\text{opt}}\Big(1 - \frac{L_{\text{opt}}}{T}\Big) \log_2(\rho_{\text{avg}}) + \text{O}(1)$ 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*, $L_{\text{opt}}\Big(1 - \frac{L_{\text{opt}}}{T}\Big)$, 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. 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 MIMO MAC where the per-user throughput remains a constant for $n \leq M$ and drops with n only for $n > M$. We identify the availability of CSI at the BS to be behind this behavior in coherent channels as against the non-coherent MIMO MAC.

1.3 Organization of this Thesis

The rest of this thesis is organized as follows.

Chapter 2 reviews fundamental ideas and sheds light on the existing literature.

Chapter 3 describes the problem setup more precisely.

Chapter 4 includes the derivation of the sum capacity lower bound, to be used as the main performance metric.

Chapter 5 discusses the optimal design of various parameters involved in our scheme.

Chapter 6 contains an asymptotic analysis of the proposed scheme.

Chapter 7 includes conclusions and insights on extending our work.

CHAPTER 2

BACKGROUND

2.1 MIMO systems

Multiple-input multiple-output technology that employ multiple antennas at both the transmitter and the receiver was first invented by Jack Winters at the Bell Laboratories in 1984. Since then, a large amount of research [1,7] etc., has been performed in using this technique to improve transmission reliability for a given data rate or increase data rate for a given transmission reliability. The former gain is called the diversity gain obtained by transmitting same information over multiple paths. These multiple paths are made possible by the rich scattering environment that offers independent point to point links between the multiple antennas at the transmitter side to those at the receiver side. The latter gain is known as the multiplexing gain achieved by multiplexing independent data over the available multiple paths.

2.2 Multiuser communication

In applications such as wireless local area networks (WLANs), satellite-based networks and cellular networks two possible communication links exists: (a) Uplink -

a common receiver (base station, BS) is accessed by multiple users through a common medium. This type of channel is also referred to as Multiple Access channel, MAC, [8]. (b) Downlink (also known as Broadcast channel, BC) - a single transmitter (BS) sends independent data to multiple receivers through a shared medium [9]. In both these cases, like in point to point links, multi-antenna transmitter/receiver structures are proved to be of tremendous value in increasing the spectral efficiency. For a MIMO MAC, the capacity region is a convex n-dimensional region that defines the individual rates achievable simultaneously by all the n users in the system [3, 10]. The corner points can be achieved by linear MMSE filtering followed by successive interference cancellation at the BS. Although the BC region is not known explicitly with closed form expressions, it has been proven that the BC rate region is the union of the MAC region for various combinations of individual powers that sum to the same total power constraint and that this region is achievable with Costa precoding [11–15]. These results on MIMO multiuser channels apply straightforwardly to fading channels by considering the channel in each fading state as a parallel non-interfering multiuser channel.

2.3 Opportunistic communication

Apart from the multiplexing and diversity gains seen by the use of multiple antennas in fading channels, multiuser systems offer a type of gain called multiuser diversity gain or scheduling gain. The diversity or richness in the links of users that are geographically far apart renders the channel of some users better than others during a fading state. When the BS exploits this diversity, i.e., when it schedules communication with users whose instantaneous channel quality is good, a considerable increase in the achievable sum rate can be realized. This opportunistic scheduling was first introduced in [16] for single antenna multiuser systems and extended to MIMO multiuser systems in [17]. These works clearly projected fading as a beneficial phenomenon that acts as a source of randomness in the channel coefficients, hence, leading to multiuser diversity gains. In fact, researchers have considered promoting fading in these channels using transmit beamforming [18]. For a detailed treatment on opportunistic communication the reader may refer to [19] and the references therein.

Throughout the discussion thus far, we have assumed the availability of CSI at the BS. CSI at BS is instrumental in realizing the MIMO and multiuser gains effectively. But if the channel fading states change rapidly, it is impractical to trace the channel accurately, forcing us to deploy a suitable non-coherent communication technique. These techniques can be broadly classified as (a) those that explicitly estimate the channel coefficients (b) those that equalize the channel without estimating it explicitly. Despite the richness of this problem, non-coherent MIMO multiuser systems is a topic that has enjoyed relatively less attention than its single user counterpart [4, 6]. Even more, when it comes to multiuser scheduling in non-coherent MIMO MAC, there is no literature available to the best of our knowledge. We attempt to fill this gap by considering a cross layer design of pilot based training (Physical layer) and user scheduling (MAC layer) in this thesis.

CHAPTER 3

PROBLEM SETUP

3.1 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.

3.2 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 $X_{\tau,c} \in \mathbb{C}^{T_\tau \times M}$:

$$X_{\tau,c} = \sqrt{\rho_\tau} S_{\tau,c} H_{\tau,c} + V_{\tau,c}, \tag{3.1}$$

where $H_{\tau,c}$ is the channel matrix made up of i.i.d. $\mathcal{CN}(0,1)$ coefficients. $V_{\tau,c} \in \mathbb{C}^{T_\tau \times M}$ 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:

$$\hat{H}_{\tau,c} = \sqrt{\frac{1}{\rho_\tau}} \Big(\frac{I_L}{\rho_\tau} + S^*_{\tau,c} S_{\tau,c} \Big)^{-1} S^*_{\tau,c} X_{\tau,c}, \tag{3.2}$$

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

3.3 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 $X^i_{d,c} \in \mathbb{C}^{T_d \times M}$:

$$\begin{aligned} X^i_{d,c} &= \sqrt{\rho_d} S^i_{d,c} H^i_{d,c} + V^i_{d,c} \\ &= \sqrt{\rho_d} S^i_{d,c} \hat{H}^i_{d,c} + \underbrace{\sqrt{\rho_d} S^i_{d,c} \tilde{H}^i_{d,c} + V^i_{d,c}}_{\bar{V}^i_{d,c}}, \end{aligned} \tag{3.3}$$

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 $\mathrm{E}\,\mathrm{tr}[S^{i*}_{d,c} S^i_{d,c}] \leq K T_d$, 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 (3.3), $X^i_{d,c}$ 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) allocated to all coherence times, the total energy spent in any coherence time is $\rho_{\text{avg}} n T$ thus giving the relation $\rho_{\text{avg}} n T = \rho_d T_d K + \rho_\tau T_\tau L$. 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 chapter, 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).

CHAPTER 4

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 (3.3) for one symbol time given by

$$\boldsymbol{x}_d^i = \sqrt{\rho_d}\boldsymbol{s}_d^i \hat{H}_d^i + \bar{\boldsymbol{v}}_d^i. \tag{4.1}$$

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 (3.3) respectively. Let I^i be 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 A for its derivation):

$$C_{\text{LB}}(R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}) = \text{E} \inf_{p_{\bar{\boldsymbol{v}}_d^i}, \forall i} \sup_{p_{\boldsymbol{s}_d^i}, \forall i} \max_i \frac{T - T_\tau}{T} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}) \tag{4.2}$$

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 $R_{\boldsymbol{s}_d^i}$ and $R_{\bar{\boldsymbol{v}}_d^i}$. The reason for this will be clear as we proceed.

The signal correlation matrix is $R_{\boldsymbol{s}_d^i} = \text{E}[\boldsymbol{s}_d^{i*} \boldsymbol{s}_d^i]$. Since the users cannot cooperate and since we do not perform power control across space or time (other than multiuser scheduling), $R_{\boldsymbol{s}_d^i} = I_K, \forall i$ (where I_K is the $K \times K$ identity matrix). The correlation matrix of the zero-mean noise, $\bar{\boldsymbol{v}}_d^i$, is given by

$$\begin{aligned} R_{\bar{\boldsymbol{v}}_d^i} &= \mathrm{E}(\sqrt{\rho_d}\boldsymbol{s}_d^i \tilde{H}_d^i + \boldsymbol{v}_d^i)^*(\sqrt{\rho_d}\boldsymbol{s}_d^i \tilde{H}_d^i + \boldsymbol{v}_d^i) \\ &= \rho_d \, \mathrm{E}[\tilde{H}_d^{i*} \tilde{H}_d^i] + I_M. \end{aligned} \tag{4.3}$$

For brevity we will refer to this capacity lower bound as C_{LB} hereafter and provide an explicit expression for it in the following lemma.

Lemma 1.

$$C_{\mathrm{LB}} = \frac{T - T_\tau}{T} \, \mathrm{E} \max_i \log\det(I_M + \rho_d R_{\bar{\boldsymbol{v}}_d^i}^{-1} \hat{H}_d^{i*} \hat{H}_d^i) \tag{4.4}$$

Proof. Consider the 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. Following the proof of Theorem 1 in [4], under the constraint $\mathrm{E}[\boldsymbol{s}^*\boldsymbol{v}] = 0_{K\times M}$, for any $R_{\boldsymbol{s}}$, $R_{\boldsymbol{v}}$,

$$\inf_{p_{\boldsymbol{v}}} \sup_{p_{\boldsymbol{s}}} I(\boldsymbol{s};\boldsymbol{x}|\breve{H}) = \log\det(I_M + \rho R_{\boldsymbol{v}}^{-1} \breve{H}^* R_{\boldsymbol{s}} \breve{H}) \tag{4.5}$$

with Gaussian distributed signal and noise (throughout this paper, logarithms are to the base 2, unless mentioned otherwise.) Now, for our channel in (4.1), the following holds:

$$\begin{aligned} \mathrm{E}[\boldsymbol{s}_d^{i*}\bar{\boldsymbol{v}}_d^i | X_\tau, S_\tau] &= \sqrt{\rho_d}\, \mathrm{E}[\boldsymbol{s}_d^{i*}\boldsymbol{s}_d^i | X_\tau, S_\tau] \, \mathrm{E}[\tilde{H}_d^i | X_\tau, S_\tau] \\ &= 0_{K\times M}. \end{aligned} \tag{4.6}$$

Hence, using Theorem 1 of [4],

$$(\mathcal{CN}(0, R_{\bar{\boldsymbol{v}}_d^i}), \mathcal{CN}(0, I_K)) = \arg \inf_{p_{\bar{\boldsymbol{v}}_d^i}(.)} \sup_{p_{\boldsymbol{s}_d^i}(.)} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, I_K). \tag{4.7}$$

But note that

$$I^i(\mathcal{CN}(0, R_{\bar{\boldsymbol{v}}_d^i}), \mathcal{CN}(0, I_K)) \quad = \quad \log\det(I_M + \rho_d R_{\bar{\boldsymbol{v}}_d^i}^{-1} \hat{H}_d^{i*} \hat{H}_d^i). \tag{4.8}$$

Define $I_{\text{lb}}^i = \frac{T-T_\tau}{T} I^i(\mathcal{CN}(0, R_{\bar{\boldsymbol{v}}_d^i}), \mathcal{CN}(0, I_K))$ for brevity. Due to symmetry, (4.8) holds for any subset i, and so we obtain

$$C_{\text{LB}} \quad = \quad \mathrm{E} \max_i I_{\text{lb}}^i \tag{4.9}$$

□

We show in Appendix A 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.

$$C_{\text{LB}} \leq R_{\text{worst}}^{\max} \leq C_{\text{sum}} \tag{4.10}$$

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 chapter.

CHAPTER 5

PARAMETER DESIGN

Within the training based scheme described in Section 3, the following three are design choices: training sequence S_τ, training power ρ_τ, and training period T_τ.

In light of the analysis in the preceding chapter, it is tempting to choose these parameters to maximize C_{LB}. However, from (4.9), 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 (4.9), $C_{\mathrm{LB}} = \mathrm{E}\max_i I_{\mathrm{lb}}^i \geq \mathrm{E}\, I_{\mathrm{lb}}^q$, for any fixed q. In the rest of this chapter, $\mathrm{E}\, I_{\mathrm{lb}}^q$, 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 $\sigma^2_{\tilde{H}_d^i} := \frac{1}{MK}\,\mathrm{E}\,\mathrm{tr}[\tilde{H}_d^{i*}\tilde{H}_d^i], \forall i$.

5.1 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 $I_{\rm lb}^q$ 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

$$\begin{aligned} \sigma^2_{\bar{\boldsymbol{v}}_d^q} &:= \frac{1}{M}\,\mathrm{tr}[R_{\bar{\boldsymbol{v}}_d^q}] \\ &= 1 + K\rho_d \sigma^2_{\tilde{H}_d^q}. \end{aligned} \tag{5.1}$$

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,

$$\mathrm{E}\, I_{\rm lb}^q = \frac{T - T_\tau}{T}\,\mathrm{E} \log\det(I_M + \rho_{\rm eff}^q \ddot{R}^{-1}_{\bar{\boldsymbol{v}}_d^q} \ddot{H}_d^{q*} \ddot{H}_d^q), \tag{5.2}$$

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

$$\begin{aligned} \rho_{\rm eff}^q &= \frac{\rho_d \sigma^2_{\hat{H}_d^q}}{1 + K\rho_d \sigma^2_{\tilde{H}_d^q}} \\ &= \frac{1}{K}\Big[\frac{1 + K\rho_d}{1 + K\rho_d \sigma^2_{\tilde{H}_d^q}} - 1\Big] \end{aligned} \tag{5.3}$$

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 (5.2) through $\rho_{\rm eff}^q$, we choose to maximize $\rho_{\rm eff}^q$ by minimizing $\sigma^2_{\tilde{H}_d^q}$. Let,

$$\begin{aligned} \sigma^2_{\tilde{H}_\tau} &:= \frac{1}{ML}\,\mathrm{E}\,\mathrm{tr}[\tilde{H}_\tau^* \tilde{H}_\tau] \\ &= \frac{1}{ML}\sum_{a=1}^{L}\sum_{b=1}^{M} \mathrm{var}[\tilde{H}_\tau]_{a,b} \end{aligned} \tag{5.4}$$

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. $Q = \binom{L}{K}$, 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

$$\sum_{i=1}^{Q} \sigma^2_{\tilde{H}_d^i} MK = \sum_{a=1}^{L} \sum_{b=1}^{M} \text{var}[\tilde{H}_\tau]_{a,b} \frac{QK}{L} \tag{5.5}$$

and (5.4) becomes,

$$\begin{aligned} \sigma^2_{\tilde{H}_\tau} &= \frac{1}{Q} \sum_{i=1}^{Q} \sigma^2_{\tilde{H}_d^i} \\ &= \sigma^2_{\tilde{H}_d^q}, \end{aligned} \tag{5.6}$$

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 B):

$$S_\tau^* S_\tau = T_\tau I_L. \tag{5.7}$$

Observe that we need $T_\tau \geq L$ to achieve (5.7). 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 (see Appendix B) that,

$$\begin{aligned} R_{\tilde{H}_d^q} &:= \mathrm{E}\left[(\text{vec}\, \tilde{H}_d^q)(\text{vec}\, \tilde{H}_d^q)^*\right] \\ &= \frac{1}{1 + \rho_\tau T_\tau} I_{KM} \\ \sigma^2_{\tilde{H}_d^q} &= \frac{1}{1 + \rho_\tau T_\tau} \end{aligned} \tag{5.8}$$

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 (4.3) and (5.1),

$$\begin{aligned} \ddot{R}_{\bar{\boldsymbol{v}}_d^q} &= \frac{1}{\sigma^2_{\bar{\boldsymbol{v}}_d^q}} \left[\frac{\rho_d K}{1+\rho_\tau T_\tau} I_M + I_M \right] \\ &= I_M. \end{aligned} \tag{5.9}$$

$$\rho_{\mathrm{eff}}^q = \frac{\rho_d \rho_\tau T_\tau}{1+\rho_\tau T_\tau + K\rho_d}. \tag{5.10}$$

Note that (5.8) 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 chapter on S_τ. In fact, all the equations from (5.6) through (5.10) apply equally well to any subset i leading to $\mathrm{E}\, I_{\mathrm{lb}}^i = \mathrm{E}\, I_{\mathrm{lb}}^j, \forall i, j$. Defining $I_{\mathrm{lb}} := I_{\mathrm{lb}}^q$, the objective function can be rewritten as the following where $\rho_{\mathrm{eff}} = \rho_{\mathrm{eff}}^i, \forall i$:

$$\mathrm{E}\, I_{\mathrm{lb}} = \frac{T - T_\tau}{T} \mathrm{E} \log\det(I_M + \rho_{\mathrm{eff}} \ddot{H}_d^{i*} \ddot{H}_d^i) \tag{5.11}$$

5.2 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 $\rho_{\mathrm{avg}} nT = \rho_\tau T_\tau L + \rho_d T_d K$, where, recall that ρ_{avg} is the average power constraint of each user. Let $\rho_d T_d K = \alpha \rho_{\mathrm{avg}} nT$ for some $\alpha \in (0, 1]$. Then

$$\rho_{\mathrm{eff}} = \frac{(\rho_{\mathrm{avg}} nT)^2}{T_d K} \frac{\alpha(1-\alpha)}{L + \rho_{\mathrm{avg}} nT - \alpha\rho_{\mathrm{avg}} nT(1 - \frac{L}{T_d})} \tag{5.12}$$

The value of α that maximizes ρ_{eff} is derived in Appendix C to be the following:

$$\alpha_{\text{opt}} = \begin{cases} \frac{1}{2} & T_d = L \\ \gamma - \sqrt{\gamma(\gamma-1)} & T_d > L \\ \gamma + \sqrt{\gamma(\gamma-1)} & T_d < L \end{cases} \quad \text{where} \quad \gamma = \frac{L + \rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1 - \frac{L}{T_d}]} \tag{5.13}$$

The intuition behind (5.13) will be apparent after we discuss the design of the training period.

5.3 Training Period, T_τ

We now derive the training period T_τ that maximizes $\mathrm{E}\, I_{\text{lb}}$. It can be proven (see Appendix D) 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 (5.7)), 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 (5.13), 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 which we summarize as follows.

5.4 Summary

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 $\rho_d T_d K = \alpha \rho_{\text{avg}} n T$ and the total energy spent on training is $\rho_\tau T_\tau L = (1-\alpha)\rho_{\text{avg}} n T$, where $\alpha = \alpha_{\text{opt}}$ is given in (5.13).

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 $i^{\max} = \arg\max_i I^i_{\text{lb}}$ (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 $\frac{T}{n(T-L)} C_{\text{LB}}$ and interleaves its codewords across the coherence intervals in which it transmits data. Using the designed parameters, we update C_{LB} and ρ_{eff} as follows,

$$C_{\text{LB}} = \frac{T-L}{T} \operatorname{E} \max_i \log\det(I_M + \rho_{\text{eff}} \ddot{H}^{i*}_d \ddot{H}^i_d) \tag{5.14}$$

with ρ_{eff} in (5.12) rewritten as,

$$\rho_{\text{eff}} = \begin{cases} \frac{(\rho_{\text{avg}} n)^2}{K(1+2\rho_{\text{avg}} n)} & T = 2L \\ \frac{\rho_{\text{avg}} n T}{K(T-2L)} (\sqrt{\gamma} - \sqrt{\gamma - 1})^2 & T > 2L \\ \frac{\rho_{\text{avg}} n T}{K(2L-T)} (\sqrt{-\gamma} - \sqrt{1-\gamma})^2 & T < 2L \end{cases} \quad \text{where} \quad \gamma = \frac{L + \rho_{\text{avg}} n T}{\rho_{\text{avg}} n T} \frac{T-L}{T-2L} \tag{5.15}$$

We now illustrate the proposed scheme with a numerical example in the following section.

5.5 Illustration

System parameters:

$$\begin{aligned}
\text{number of users, } n &= 5 \\
\text{number of BS antennas, } M &= 3 \\
\text{average power constraint, } \rho_{\text{avg}} &= 3 \quad \text{dB} \\
\text{coherence interval, } T &= 10 \quad \text{channel uses}
\end{aligned}$$

Design parameters: Since we have not designed L and K yet (this will be addressed in the next chapter), we set L=4 and K=3 arbitrarily without compromising the objective of this example.

$$\begin{aligned}
\text{training period, } T_\tau &= L = 4 \\
\text{data period, } T_d &= T - T_\tau = 6 \\
\text{training sequence matrix, } S_\tau &= \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \\
\text{power allocation, } \alpha_{\text{opt}} &= 0.5481
\end{aligned}$$

With these system and design parameters, we have $\gamma = 3.1203$, effective SNR, $\rho_{\text{eff}} = 1.6011$ and number of subsets $Q = 4$. Also,

$$\begin{aligned}
\text{total power spent in training, } \rho_\tau L &= 11.2696 \\
\text{total power spent in data, } \rho_d K &= 9.1141
\end{aligned}$$

thus giving more power to training than to data since more time is spent on data than on training.

For a single realization, the channel estimate matrix (*normalized* by its standard deviation) of the $L = 4$ trained users is obtained by generating $\mathcal{CN}(0,1)$ entries as follows,

$$\underbrace{\hat{H}_{\mathcal{T}}}_{normalized} = \begin{pmatrix} 0.5524 - 0.6697i & -0.8399 + 1.0412i & 0.2315 - 0.7979i \\ 0.4023 - 0.2648i & -1.5573 + 0.0394i & 0.1655 - 0.9541i \\ -0.5810 - 0.8385i & 0.6974 - 0.8608i & 0.0152 - 0.1846i \\ -0.1878 - 0.7466i & -0.3667 - 0.0292i & -0.7099 + 0.6742i \end{pmatrix}$$

where the j^{th} row corresponds to the channel signature vector (estimate) of the user indexed by j.

Now, we proceed with the calculation of the mutual information associated with each subset $i \in \{1, 2, 3, 4\}$ as follows.

Subset 1:

$$\text{contributing user indices} = \{1, 2, 3\}$$

$$\ddot{H}_d^1 = \begin{pmatrix} 0.5524 - 0.6697i & -0.8399 + 1.0412i & 0.2315 - 0.7979i \\ 0.4023 - 0.2648i & -1.5573 + 0.0394i & 0.1655 - 0.9541i \\ -0.5810 - 0.8385i & 0.6974 - 0.8608i & 0.0152 - 0.1846i \end{pmatrix}$$

$$\text{mutual information, } I_{\text{lb}}^1 = 3.4979$$

Subset 2:

$$\text{contributing user indices} = \{1, 2, 4\}$$

$$\ddot{H}_d^2 = \begin{pmatrix} 0.5524 - 0.6697i & -0.8399 + 1.0412i & 0.2315 - 0.7979i \\ 0.4023 - 0.2648i & -1.5573 + 0.0394i & 0.1655 - 0.9541i \\ -0.1878 - 0.7466i & -0.3667 - 0.0292i & -0.7099 + 0.6742i \end{pmatrix}$$

$$\text{mutual information, } I_{\text{lb}}^2 = 3.5590$$

Subset 3:

$$\text{contributing user indices} = \{1, 3, 4\}$$

$$\ddot{H}_d^3 = \begin{pmatrix} 0.5524 - 0.6697i & -0.8399 + 1.0412i & 0.2315 - 0.7979i \\ -0.5810 - 0.8385i & 0.6974 - 0.8608i & 0.0152 - 0.1846i \\ -0.1878 - 0.7466i & -0.3667 - 0.0292i & -0.7099 + 0.6742i \end{pmatrix}$$

$$\text{mutual information, } I_{\text{lb}}^3 = 3.8238$$

Subset 4:

$$\text{contributing user indices} = \{2, 3, 4\}$$

$$\ddot{H}_d^4 = \begin{pmatrix} 0.4023 - 0.2648i & -1.5573 + 0.0394i & 0.1655 - 0.9541i \\ -0.5810 - 0.8385i & 0.6974 - 0.8608i & 0.0152 - 0.1846i \\ -0.1878 - 0.7466i & -0.3667 - 0.0292i & -0.7099 + 0.6742i \end{pmatrix}$$

$$\text{mutual information, } I_{\text{lb}}^4 = 3.8702$$

Comparing the mutual information of all the four subsets, we conclude the user subset {2,3,4} is associated with the largest mutual information. Hence these users are allowed to transmit data during the data transmission phase. This completes our numerical example.

This chapter addressed the design of various parameters involved in the proposed scheme. However, an important question is left unanswered: What are the optimum numbers of users to be trained (L) and allowed to transmit data (K)? We explore this in the following chapter.

CHAPTER 6

ASYMPTOTIC ANALYSIS

In this chapter, we address the design of L and K in regimes where various parameters such as the SNR (ρ_{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,*

$$C_{\text{LB}} = \frac{T-L}{T}\min(K,M)\log(\rho_{\text{avg}}) + O(1) \quad \textit{as} \quad \rho_{\text{avg}} \to \infty \tag{6.1}$$

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

Proof. From (5.14), with $i_{\max}$ indexing the C_{LB}-maximizing subset,

$$\begin{aligned} C_{\text{LB}} &= \frac{T-L}{T}\,\mathrm{E}\log\det(I_M + \rho_{\text{eff}}\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}) \\ &= \frac{T-L}{T}\,\mathrm{E}\log\prod_{j=1}^{\min(K,M)}(1+\rho_{\text{eff}}\lambda_j^{i_{\max}}) \end{aligned} \tag{6.2}$$

where, $\lambda_j^{i_{\max}}$ corresponds to the j^{th} non-zero eigenvalue of $\ddot{H}_d^{i_{\max}*}\ddot{H}_d^{i_{\max}}$.

As $\rho_{\text{avg}} \to \infty$, from (5.15),

$$\frac{\rho_{\text{eff}}}{\rho_{\text{avg}}} = \frac{n}{K(\sqrt{1-\frac{L}{T}}+\sqrt{\frac{L}{T}})^2} \quad \text{since} \quad \gamma = \frac{T-L}{T-2L} \tag{6.3}$$

and $\frac{\log(1+\rho_{\text{eff}}\lambda_j^{i_{\max}})}{\log(\rho_{\text{eff}}\lambda_j^{i_{\max}})} = 1$. Thus

$$\begin{aligned} C_{\text{LB}} &= \frac{T-L}{T}\min(K,M)\log(\rho_{\text{avg}}) + \underbrace{\frac{T-L}{T}\min(K,M)\log\Big(\frac{n}{K(\sqrt{1-\frac{L}{T}}+\sqrt{\frac{L}{T}})^2}\Big)}_{A} \\ &+ \underbrace{\frac{T-L}{T}\mathrm{E}\sum_{j=1}^{\min(K,M)}\log\lambda_j^{i_{\max}}}_{B} + \quad O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty \end{aligned} \tag{6.4}$$

We have already shown that the optimal value of L is equal to T_τ, where $0 \leq T_\tau \leq T-1$ (where the second inequality is due to the necessity of reserving at least one symbol time for data transmission.) As $L \leq n$ by definition, we have $L \leq \min(n, T-1)$. Note that the term B can be upper bounded as,

$$\mathrm{E}\sum_{j=1}^{\min(K,M)}\log\lambda_j^{i_{\max}} \leq \mathrm{E}\max_i \sum_{j=1}^{\min(K,M)}\log\lambda_j^i. \tag{6.5}$$

The upper bound contains a sum of the λ_j^i's, which is finite with probability 1 (for any subset i), since K, L are upper-bounded by n and M, T and n are fixed. Thus B is bounded and does not increase with ρ_{avg}. Similarly, A is also bounded as ρ_{avg} increases. Putting these together,

$$C_{\text{LB}} = \frac{T-L}{T}\min(K,M)\log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty. \tag{6.6}$$

The prelog factor $\frac{T-L}{T}\min(K,M)$ is decreasing in L for a given K, and increasing in K up to $K = M$ for a given L. But $K \leq L$. Hence this prelog factor is maximized when $L = K \leq M$. Thus, under the constraint $L \leq \min(n, M, T-1)$, the prelog factor is given by $\left(\frac{T-L}{T}\right)L$ which is concave and quadratic in L with the maximum at $L = \frac{T}{2}$. Therefore L has to be as close as possible to $\frac{T}{2}$ giving $L_{\text{opt}} = \min(n, M, T-1, \frac{T}{2})$. Since $T \geq 2$, and since $\frac{T}{2}$ might not be an integer, we have $L_{\text{opt}} = K_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Thus the optimized prelog factor is given by $\left(\frac{T-L_{\text{opt}}}{T}\right)L_{\text{opt}}$. □

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

$$C = \frac{T-n^*}{T} n^* \log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty \tag{6.7}$$

with $n^* = \min(n, M, \lfloor \frac{T}{2} \rfloor)$ thus giving C_{LB} the same prelog factor as C. This is illustrated in Fig. 6.1. 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 MIMO channel where 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, multiple access MIMO channel, from Theorem 1 and [6],*

$$C_{\text{sum}} = \frac{T-n^*}{T} n^* \log(\rho_{\text{avg}}) + O(1) \quad \textit{as} \quad \rho_{\text{avg}} \to \infty \tag{6.8}$$

giving the non-coherent multiple access MIMO 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_{sum}.

Now we proceed to analyze how the gain from exploiting multiuser diversity behaves as SNR grows. If $C_{\text{LB}}(L, K)$ indicates the lower bound in (5.14), then the

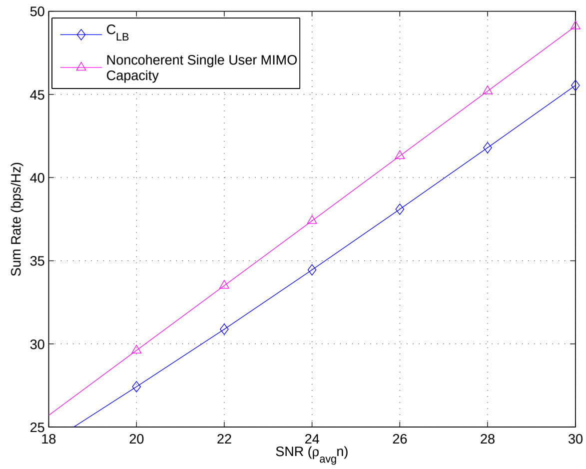

Figure 6.1: Comparison of C_{LB} with the non-coherent single user MIMO capacity when $T = 30$, $n = M = 8$.

baseline case (i.e., no multiuser scheduling) occurs with $L = K$ as,

$$C_{\text{LB}}(L, L) = \frac{T - L}{T} \, \mathrm{E} \log\det(I_M + \rho_{\text{eff}} \ddot{H}_d^{i*} \ddot{H}_d^i) \tag{6.9}$$

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

$$C_{\text{LB}}(L, L) = \Big(\frac{T - L_{\text{opt}}}{T}\Big) L_{\text{opt}} \log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty \tag{6.10}$$

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

$$\lim_{\rho_{\text{avg}} \to \infty} \frac{\max_{L,K} C_{\text{LB}}(L, K)}{\max_L C_{\text{LB}}(L, L)} = 1 \tag{6.11}$$

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,*

$$C_{\text{LB}} = \min(n, M)\log(\rho_{\text{avg}}) + O(1) \quad \textit{as} \quad \rho_{\text{avg}} \to \infty, T \to \infty \tag{6.12}$$

and $L_{\text{opt}} = K_{\text{opt}} = n$.

Proof. As ρ_{avg}, $T \to \infty$, it is easy to verify that terms A and B in (6.4) are bounded w.r.t T using a similar analysis used for ρ_{avg}. Thus $C_{\text{LB}} = \frac{T-L}{T}\min(K, M)\log(\rho_{\text{avg}}) + O(1) = \min(K, M)\log(\rho_{\text{avg}}) + O(1)$ as $\rho_{\text{avg}} \to \infty$, $T \to \infty$. Now the prelog factor is maximized when $K = n$ and, since $L \geq K$, we have $L = K = L_{\text{opt}} = n$ and thus $C_{\text{LB}} = \min(n, M)\log(\rho_{\text{avg}}) + O(1)$ as $\rho_{\text{avg}} \to \infty,\ T \to \infty$. □

Note that (6.12) has the same prelog factor as that of the capacity expression of the coherent multiuser uplink [6, 10], 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. 6.2. In fact, using an argument similar to that of Corollary 1, we have the quite intuitive result that, as $T \to \infty$ and $\rho_{\text{avg}} \to \infty$, the non-coherent sum capacity increases at the same rate as coherent capacity.

$$C_{\text{sum}} = \min(n, M)\log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty, T \to \infty. \tag{6.13}$$

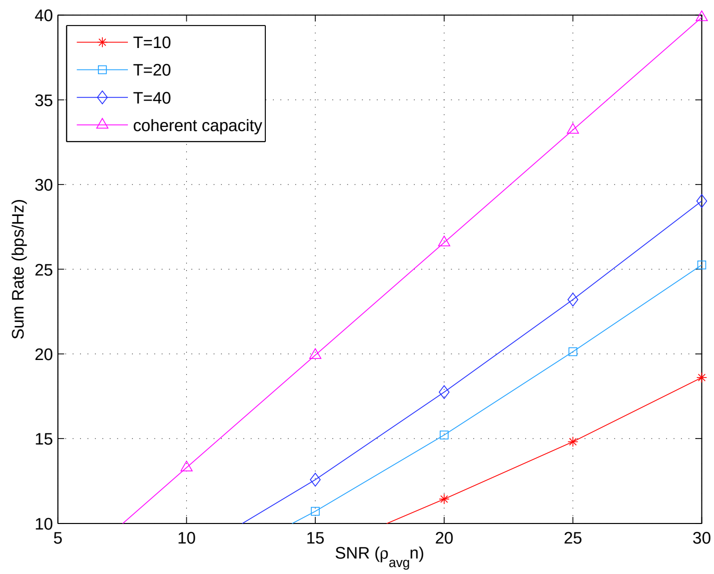

Figure 6.2: Illustration to show how the slope of sum rate achieved by the training-based scheme approaches that of coherent channel capacity when $n = M = 4$.

Theorem 3. *With T, M, ρ_{avg} fixed,*

$$C_{\text{LB}} \;=\; \frac{T-L}{T}\min(K,M)\log(n)+O(1) \quad \textit{as} \quad n\to\infty \tag{6.14}$$

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 (5.15), as $n \to \infty$,

$$\frac{\rho_{\text{eff}}}{n} = \frac{\rho_{\text{avg}}}{K(\sqrt{1-\frac{L}{T}}+\sqrt{\frac{L}{T}})^2} \quad \text{since} \quad \gamma = \frac{T-L}{T-2L} \tag{6.15}$$

Thus from (6.2),

$$
\begin{aligned}
C_{\text{LB}} &= \frac{T-L}{T}\min(K,M)\log(n) + \underbrace{\frac{T-L}{T}\min(K,M)\log\left(\frac{\rho_{\text{avg}}}{K(\sqrt{1-\frac{L}{T}}+\sqrt{\frac{L}{T}})^2}\right)}_{A} \\
&\quad + \underbrace{\frac{T-L}{T}\,\mathrm{E}\sum_{j=1}^{\min(K,M)}\log\lambda_j^{i_{\max}}}_{B} \\
&= \frac{T-L}{T}\min(K,M)\log(n) + O(1) \quad \text{as} \quad n\to\infty \qquad (6.16)
\end{aligned}
$$

Here A is bounded w.r.t n since $K \le L \le \min(n, T-1)$ and B is bounded using a similar argument as in the $\rho_{\text{avg}} \to \infty$ case. 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 prelog factor equal to the degrees of freedom of the non-coherent uplink obtained in Corollary 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. 6.3. 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 importance of this phenomenon in multiple access channels deserves a detailed analysis. In fact, the increase in the power of the active users with increasing number of users in the system may mislead us to the following conclusion: In every coherence interval, the users that are not active share their power with those that are active, thus suggesting cooperation among supposedly independent users. A careful analysis will disprove this conception. With the number of users in the system increasing, every user is allowed to access the channel less frequently. Specifically, any user accesses the channel for a fraction, $\frac{L}{n}$, of the coherence intervals. Thanks to the average power constraint for each user, the power that is not spent when an user doesn't transmit data is saved and used when it becomes active. Thus, the more users are present in the system, the less frequently an user accesses the channel and more power is saved to be used when it transmits data. Therefore, the increase in power of any active user and hence the rise in sum rate with n is due to sharing of power within a user across time and not due to sharing of power across space, i.e., across users.

Also, 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.

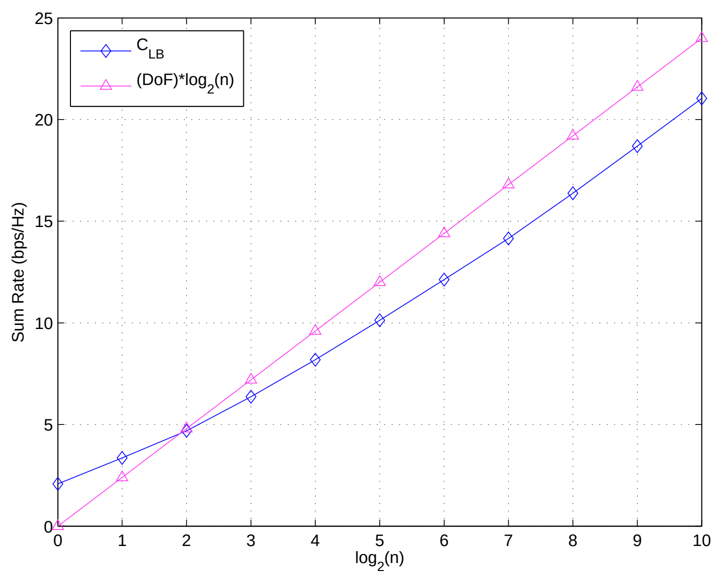

Figure 6.3: C_{LB} increases by the channel's degrees of freedom (DoF) every time the number of users in the system doubles. $T = 10$, $M = 4$, $\rho_{\text{avg}} = 3$ dB used.

Theorem 4. *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 (6.1), $\frac{C_{\text{LB}}}{n}$ is given by

$$\frac{C_{\text{LB}}}{n} = \left(\frac{T-L}{T}\right)\left(\frac{\min(K,M)}{n}\right)\log(\rho_{\text{avg}}) + O(1) \quad \text{as} \quad \rho_{\text{avg}} \to \infty \tag{6.17}$$

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 the following two cases,

(i) $n \geq \min(M, \lfloor \frac{T}{2} \rfloor)$: Here $\frac{C_{\text{LB}}}{n} = \left(\frac{T-\min(M,\lfloor \frac{T}{2} \rfloor)}{T}\right)\left(\frac{\min(M,\lfloor \frac{T}{2} \rfloor)}{n}\right)\log(\rho_{\text{avg}}) + O(1)$, which is monotonically decreasing with n with the maximum given by,
$\left(\frac{T-\min(M,\lfloor \frac{T}{2} \rfloor)}{T}\right)\log(\rho_{\text{avg}}) + O(1)$ at $n = \min(M, \lfloor \frac{T}{2} \rfloor)$.
(ii) $1 \leq n < \min(M, \lfloor \frac{T}{2} \rfloor)$: Here $\frac{C_{\text{LB}}}{n} = \left(\frac{T-n}{T}\right)\log(\rho_{\text{avg}}) + O(1)$ which is also monotonically decreasing with n, with the maximum at $n = 1$ given by $\left(\frac{T-1}{T}\right)\log(\rho_{\text{avg}}) + O(1)$.

From the above two cases, combined with the fact that $\frac{C_{\text{LB}}}{n}\big|_{n=\min(M,\lfloor \frac{T}{2} \rfloor)-1} > \frac{C_{\text{LB}}}{n}\big|_{n=\min(M,\lfloor \frac{T}{2} \rfloor)}$, we prove that $\frac{C_{\text{LB}}}{n}$ is monotonically decreasing with n for all $n \geq 1$. Also, from Corollary 1, since the non-coherent sum capacity has the same prelog factor as C_{LB}, the per-user *capacity*, $\frac{C_{\text{sum}}}{n}$, also monotonically decreases with n. The per-user rate of our scheme, which is lower bounded by $\frac{C_{\text{LB}}}{n}$ and upper bounded by $\frac{C_{\text{sum}}}{n}$, also monotonically decreases with n. □

The result of Theorem 4 is illustrated in Fig. 6.4 where C_{LB} from (5.14) is used to calculate the per-user rate lower bound $\frac{C_{\text{LB}}}{n}$. It is instructive to compare this result with the coherent channel case. Here, as $\rho_{\text{avg}} \to \infty$, the per-user capacity is $\frac{\min(n,M)}{n}\log(\rho_{\text{avg}}) + O(1)$ [10], [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

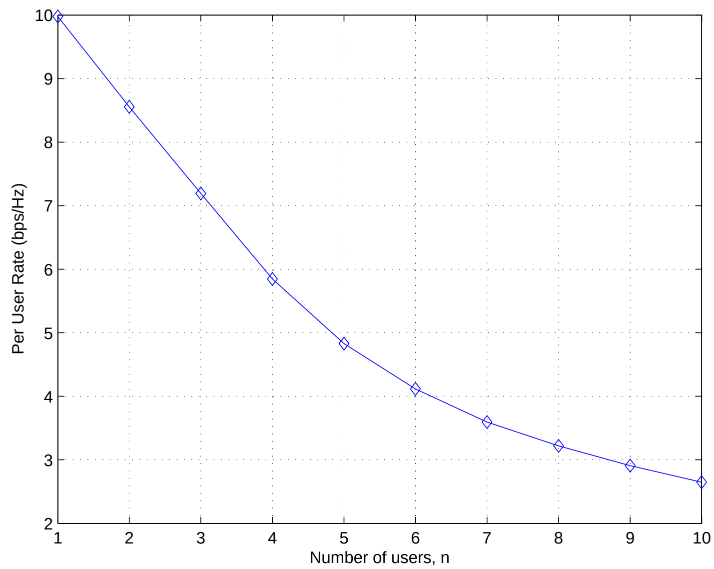

Figure 6.4: As the total number of users increases, $\frac{C_{\mathrm{LB}}}{n}$, the per-user rate (lower bound) drops monotonically. $T = 10$, $M = 4$, $\rho_{\mathrm{avg}} = 30$ dB was used.

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,*

$$C_{\mathrm{LB}} = \left(\frac{T-L}{T}\right) K \log(M) + O(1) \quad \textit{as} \quad M \to \infty \tag{6.18}$$

with the maximum at $L = K = L_{\mathrm{opt}} = \min(n, \lfloor \frac{T}{2} \rfloor)$

Proof. Recall from (4.9) and (5.9), with $T_\tau = T_{\tau,\mathrm{opt}} = L$, $I^i_{\mathrm{lb}} = \frac{T-L}{T} \log\det(I_M + \rho_{\mathrm{eff}} \ddot{H}^{i*}_d \ddot{H}^i_d)$ from which $C_{\mathrm{LB}} = \mathrm{E}\max_i I^i_{\mathrm{lb}}$. Since $\ddot{H}^i_d \in \mathbb{C}^{K \times M}$ is made up of i.i.d.

$\mathcal{CN}(0,1)$ elements (as noted before (5.9)), as $M \to \infty$, $\forall i$ I^i_{lb} converges (in distribution) to a Gaussian [20],

$$I^i_{\text{lb}} \stackrel{d}{=} \mathcal{N}\left(\left(\frac{T-L}{T}\right) K \log(1+\rho_{\text{eff}} M), \left(\frac{T-L}{T}\right)^2 \frac{K}{M} \log_2^2 e\right) \quad \text{as} \quad M \to \infty \quad (6.19)$$

Using $\log(1+\rho_{\text{eff}}M) \to \log(\rho_{\text{eff}}M)$ as $M \to \infty$, and defining $X_1, \ldots, X_Q$ as $\mathcal{CN}(0,1)$ random variables, for Q defined after (5.4), 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 (as L, K and hence Q are bounded), we have

$$C_{\text{LB}} = \left(\frac{T-L}{T}\right) K \log(M) + O(1) \quad \text{as} \quad M \to \infty \quad (6.20)$$

This is maximized when $L = K = L_{\text{opt}} = \min(n, \lfloor \frac{T}{2} \rfloor)$ giving a prelog factor which is the same as the available degrees of freedom of the non-coherent uplink channel. □

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. 6.5. Also note that as $M \to \infty$, from (6.19), the variance of the mutual information associated with any subset goes to zero (*channel hardening* [20]) and consequently the scheduling gain disappears (see Fig. 6.6). That is, as M grows, $\max_{L,K} C_{\text{LB}}(L,K) - \max_L C_{\text{LB}}(L,L)$ converges to zero, where $\max_L C_{\text{LB}}(L,L)$ corresponds to the case with no multiuser scheduling. 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 was suboptimal.

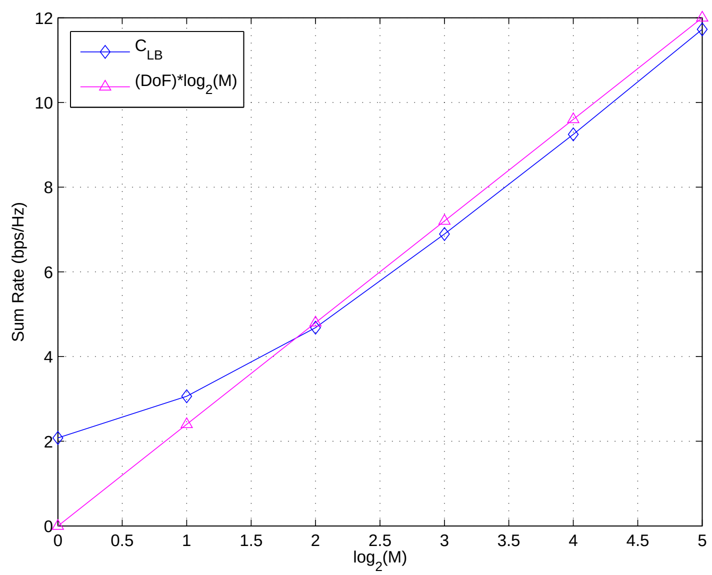

Figure 6.5: As $M \to \infty$, C_{LB} (in bits per channel use) increases by the channel's degrees of freedom (DoF) every time the number of receive antennas at the base station doubles. $T = 10$, $n = 4$, $\rho_{\mathrm{avg}} = 3$ dB.

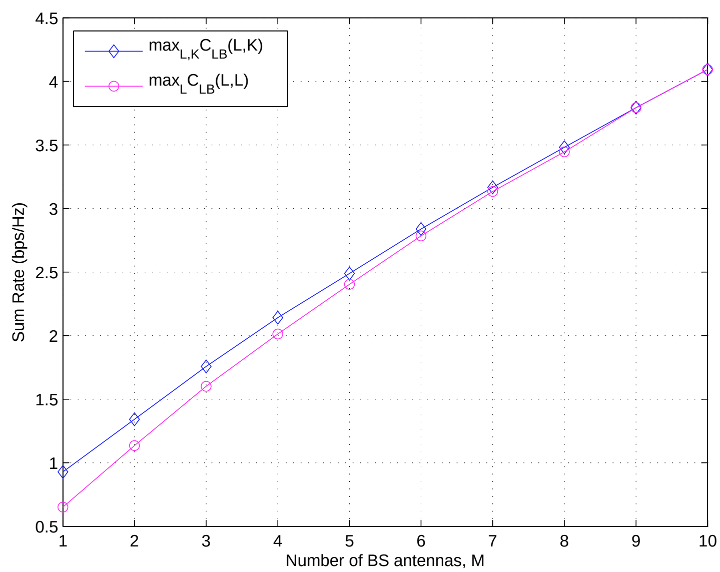

Figure 6.6: Comparison of $\max_{L,K} C_{\text{LB}}(L, K)$ with $\max_L C_{\text{LB}}(L, L)$, where there is no multiuser scheduling. The scheduling gain vanishes as M grows due to channel hardening effects. Here $n = 8$, $T = 50$, total power $\rho_{\text{avg}}n$=1.

CHAPTER 7

CONCLUSIONS

We designed a training based communication scheme for a non-coherent MIMO MAC wherein training and user selection are jointly optimized. We established that the non-coherent MIMO multiple-access channel has the same degrees of freedom as the non-coherent single user MIMO channel given by $L_{\text{opt}}\left(1 - \frac{L_{\text{opt}}}{T}\right)$, where $L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. Further, we proved that our training-based scheme has a prelog factor equal to the above degrees of freedom of the non-coherent MIMO MAC. 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 vanishes as SNR grows. The scheduling gain also vanishes as M grows 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 $L_{\text{opt}} = \min(n, M, \lfloor \frac{T}{2} \rfloor)$. 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 [5,21,22] that non-coherent channel capacity decays due to energy being spread over a wide bandwidth. More recently, Agarwal and Honig [23] 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 MIMO 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 A

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 (3.3), 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 (3.3) 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,

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

where the inequality comes from the fact that conditioning reduces entropy. Now, with C_{sum} indicating the sum capacity of the non-coherent channel, and $R_{\text{worst}}^{\max}$ the maximum sum rate achieved by the training-based scheme described in Section 3 under worst noise and best signal design conditions, we have

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

From now on, we constrain S_d^i to be independently distributed across time, so that we have,

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

Fixing the noise distribution corresponding to the inf of the above expression as $p_{\bar{V}_d^i}^1$, from the inequality proved in (A.1),

$$\begin{aligned} \mathrm{E} \inf_{p_{\bar{V}_d^i}, \forall i} \sup_{p_{S_d^i}, \forall i} \max_i \frac{T_d}{T} I_o^i(p_{\bar{V}_d^i}, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i}) &\geq \mathrm{E} \sup_{p_{S_d^i}, \forall i} \max_i \frac{T_d}{T} I^i(p_{\bar{V}_d^i}^1, p_{S_d^i}, R_{\bar{V}_d^i}, R_{S_d^i}) \\ &\geq \mathrm{E} \inf_{p_{\bar{\boldsymbol{v}}_d^i}, \forall i} \sup_{p_{\boldsymbol{s}_d^i}, \forall i} \max_i \frac{T_d}{T} I^i(p_{\bar{\boldsymbol{v}}_d^i}, p_{\boldsymbol{s}_d^i}, R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}) \\ &=: C_{\text{LB}}(R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i}). \end{aligned} \tag{A.2}$$

Thus we have $C_{\text{sum}} \geq R_{\text{worst}}^{\max} \geq C_{\text{LB}}(R_{\bar{\boldsymbol{v}}_d^i}, R_{\boldsymbol{s}_d^i})$.

APPENDIX B

TRAINING SEQUENCE DESIGN

Following the analysis in [4],

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

where vec(A) operator stacks all of the columns of A into one long column. From (5.4),

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

Thus,

$$\begin{aligned} \min \quad \sigma^2_{\tilde{H}_\tau} &= \min_{S_\tau, \mathrm{tr}\, S_\tau^* S_\tau \le L T_\tau} \frac{1}{L} \mathrm{tr}\left[\left(I_L + \rho_\tau S_\tau^* S_\tau\right)^{-1}\right] \\ &= \min_{\lambda_1, \ldots, \lambda_L, \sum_j \lambda_j \le L T_\tau} \frac{1}{L} \sum_{j=1}^{L} \frac{1}{1 + \rho_\tau \lambda_j} \end{aligned} \tag{B.3}$$

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, (B.1) and (B.2) yield

$$\begin{aligned} R_{\tilde{H}_\tau} &= \frac{1}{1+\rho_\tau T_\tau} I_{LM} \\ \sigma^2_{\tilde{H}_\tau} &= \frac{1}{1+\rho_\tau T_\tau} \end{aligned} \tag{B.4}$$

and

$$\begin{aligned} R_{\tilde{H}_d^q} &:= \mathrm{E}\left[(\mathrm{vec}\,\tilde{H}_d^q)(\mathrm{vec}\,\tilde{H}_d^q)^*\right] \\ &= \frac{1}{1+\rho_\tau T_\tau} I_{KM} \\ \sigma^2_{\tilde{H}_d^q} &= \frac{1}{1+\rho_\tau T_\tau} \end{aligned} \tag{B.5}$$

APPENDIX C

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 (5.12), considering the following three cases,

When $T_d = L$:

$$\rho_{\text{eff}} = \frac{(\rho_{\text{avg}} nT)^2 \alpha(1-\alpha)}{LK(L + \rho_{\text{avg}} nT)} \tag{C.1}$$

with $\alpha_{\text{opt}} = \frac{1}{2}$. Thus,

$$\rho_{\text{eff}} = \frac{(\rho_{\text{avg}} nT)^2}{4LK(L + \rho_{\text{avg}} nT)} \tag{C.2}$$

When $T_d > L$:

$$\begin{aligned} \rho_{\text{eff}} &= \frac{\rho_{\text{avg}} nT\alpha(1-\alpha)}{(T_d - L)K\Big[-\alpha + \frac{L+\rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1-\frac{L}{T_d}]}\Big]} \\ &= \frac{\rho_{\text{avg}} nT\alpha(1-\alpha)}{(T_d - L)K[-\alpha + \gamma]}, \end{aligned} \tag{C.3}$$

where $\gamma := \frac{L+\rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1-\frac{L}{T_d}]} > 1$. Since $\frac{d\rho_{\text{eff}}}{d\alpha}|_{\alpha=\alpha_{\text{opt}}} = 0$,

$$\alpha_{\text{opt}} = \gamma - \sqrt{\gamma(\gamma - 1)}. \tag{C.4}$$

Thus

$$\rho_{\text{eff}} = \frac{\rho_{\text{avg}} nT}{K(T_d - L)}(\sqrt{\gamma} - \sqrt{\gamma - 1})^2. \tag{C.5}$$

When $T_d < L$: In this case, with $\gamma = \frac{L+\rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1-\frac{L}{T_d}]} < 0$, following the previous steps,

$$\alpha_{\text{opt}} = \gamma + \sqrt{-\gamma(1-\gamma)} \tag{C.6}$$

and

$$\rho_{\text{eff}} = \frac{\rho_{\text{avg}} nT}{K(L - T_d)}(\sqrt{-\gamma} - \sqrt{1 - \gamma})^2. \tag{C.7}$$

In summary,

$$\alpha_{\text{opt}} = \begin{cases} \frac{1}{2} & T_d = L \\ \gamma - \sqrt{\gamma(\gamma-1)} & T_d > L \\ \gamma + \sqrt{\gamma(\gamma-1)} & T_d < L \end{cases} \tag{C.8}$$

and,

$$\rho_{\text{eff}} = \begin{cases} \frac{(\rho_{\text{avg}} nT)^2}{4LK(L+\rho_{\text{avg}} nT)} & T_d = L \\ \frac{\rho_{\text{avg}} nT}{K(T_d-L)}(\sqrt{\gamma} - \sqrt{\gamma-1})^2 & T_d > L \\ \frac{\rho_{\text{avg}} nT}{K(L-T_d)}(\sqrt{-\gamma} - \sqrt{1-\gamma})^2 & T_d < L \end{cases} \tag{C.9}$$

APPENDIX D

TRAINING PERIOD DESIGN

Under the condition $T_d \leq T - L$, 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 (5.11),

$$\begin{aligned}
\mathrm{E}\, I_{\mathrm{lb}} &= \frac{T - T_\tau}{T}\,\mathrm{E}\log\det(I_M + \rho_{\mathrm{eff}}\ddot{H}_d^{i*}\ddot{H}_d^i) \\
&= \frac{T_d}{T}\,\mathrm{E}\log\prod_{j=1}^{\min(K,M)}(1 + \rho_{\mathrm{eff}}\lambda_j^i) \\
&= \frac{T_d}{T}\min(K,M)\,\mathrm{E}\log(1 + \rho_{\mathrm{eff}}\lambda) \qquad \text{(D.1)}
\end{aligned}$$

where the expectation in the last equality is with respect to the non-zero eigenvalue λ given by,

$$\lambda = \lambda_j^i \quad \text{w.p.} \quad \frac{1}{\min(K,M)}, \forall j$$

Thus,

$$\begin{aligned}
\frac{d\,\mathrm{E}\, I_{\mathrm{lb}}}{dT_d} &= \frac{\min(K,M)}{T}\,\mathrm{E}\log(1 + \rho_{\mathrm{eff}}\lambda) \\
&\quad + \frac{\min(K,M)T_d}{T}\frac{d\rho_{\mathrm{eff}}}{dT_d}\,\mathrm{E}\left[\frac{\lambda}{1 + \rho_{\mathrm{eff}}\lambda}\right] \qquad \text{(D.2)}
\end{aligned}$$

When $T_d > L$:

From (C.9),

$$\frac{d\rho_{\text{eff}}}{dT_d} = \frac{\rho_{\text{eff}}}{(T_d - L)}\Big(\frac{L\sqrt{\gamma}}{T_d\sqrt{\gamma - 1}} - 1\Big) \tag{D.3}$$

with, $\gamma = \frac{L+\rho_{\text{avg}} nT}{\rho_{\text{avg}} nT[1-\frac{L}{T_d}]}$. Hence,

$$\frac{L}{T_d}\sqrt{\frac{\gamma}{\gamma - 1}} = \sqrt{\frac{L(L + \rho_{\text{avg}} nT)}{T_d(T_d + \rho_{\text{avg}} nT)}}. \tag{D.4}$$

Therefore,

$$\begin{aligned}\frac{d\,\mathrm{E}\, I_{\text{lb}}}{dT_d} &= \frac{\min(K, M)}{T}\,\mathrm{E}\Big[\log(1 + \rho_{\text{eff}}\lambda) \\ &\quad - \frac{\rho_{\text{eff}} T_d}{(T_d - L)}\Big(1 - \sqrt{\frac{L(L + \rho_{\text{avg}} nT)}{T_d(T_d + \rho_{\text{avg}} nT)}}\Big)\frac{\lambda}{1 + \rho_{\text{eff}}\lambda}\Big]\end{aligned} \tag{D.5}$$

Since $T_d > L$,

$$\frac{T_d}{T_d - L}\Big(1 - \sqrt{\frac{L(L + \rho_{\text{avg}} nT)}{T_d(T_d + \rho_{\text{avg}} nT)}}\Big) < 1. \tag{D.6}$$

Also, since $\lambda > 0$,

$$\log(1 + \rho_{\text{eff}}\lambda) - \frac{\rho_{\text{eff}}\lambda}{1 + \rho_{\text{eff}}\lambda} > 0. \tag{D.7}$$

Combining these, $\frac{d\,\mathrm{E}\, I_{\text{lb}}}{dT_d} > 0$ when $T_d > L$.

When $T_d < L$:

$$\frac{d\rho_{\text{eff}}}{dT_d} = \frac{\rho_{\text{eff}}}{(L - T_d)}\Big(1 - \sqrt{\frac{L(L + \rho_{\text{avg}} nT)}{T_d(T_d + \rho_{\text{avg}} nT)}}\Big) \tag{D.8}$$

and

$$\begin{aligned}\frac{d\,\mathrm{E}\, I_{\text{lb}}}{dT_d} &= \frac{\min(K, M)}{T}\,\mathrm{E}\Big[\log(1 + \rho_{\text{eff}}\lambda) \\ &\quad - \frac{\rho_{\text{eff}} T_d}{(T_d - L)}\Big(1 - \sqrt{\frac{L(L + \rho_{\text{avg}} nT)}{T_d(T_d + \rho_{\text{avg}} nT)}}\Big)\frac{\lambda}{1 + \rho_{\text{eff}}\lambda}\Big].\end{aligned} \tag{D.9}$$

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

Denoting ρ_{eff} by $\rho_{\text{eff}}(T_d)$ 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,

$$\lim_{m\to\infty} \rho_{\text{eff}}(t_m) \;=\; \rho_{\text{eff}}(L). \tag{D.10}$$

Since $\rho_{\text{eff}} \leq \rho_d, \forall T_d$ (from (5.10)),

$$\log(1+\rho_{\text{eff}}(t_m)\lambda)f_\lambda(\lambda) \leq \log(1+\rho_d\lambda)f_\lambda(\lambda) \tag{D.11}$$

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

$$\begin{aligned}\int_\lambda \log(1+\rho_d\lambda)f_\lambda(\lambda)d\lambda &= \frac{\mathrm{E}(\log\det(I_M + \rho_d \ddot{H}_d^{i*}\ddot{H}_d^{i}))}{\min(K,M)} \\ &\leq \frac{\log\det(I_M + \rho_d\,\mathrm{E}(\ddot{H}_d^{i*}\ddot{H}_d^{i}))}{\min(K,M)} \\ &< \infty,\end{aligned} \tag{D.12}$$

using the Dominated Convergence Theorem [24],

$$\begin{aligned}\lim_{m\to\infty} \mathrm{E}\log(1+\rho_{\text{eff}}(t_m)\lambda) &= \lim_{m\to\infty}\int_\lambda \log(1+\rho_{\text{eff}}(t_m)\lambda)f_\lambda(\lambda)d\lambda \\ &= \int_\lambda \log(1+\lim_{m\to\infty}\rho_{\text{eff}}(t_m)\lambda)f_\lambda(\lambda)d\lambda \\ &= \mathrm{E}\log(1+\rho_{\text{eff}}(L)\lambda)\end{aligned} \tag{D.13}$$

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

BIBLIOGRAPHY

[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 multiantenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.

[7] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, Aug. 1996.

[8] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. on Information Theory*, vol. 31, pp. 124–142, Mar. 1985.

[9] T. Cover, "Broadcast channels," *IEEE Trans. on Information Theory*, vol. 18(1), pp. 2–14, Jan. 1972.

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

[11] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, pp. 439–441, May. 1983.

[12] G. Caire and S. Shamai, "On the achievable throughput in multiple antenna gaussian broadcast channel," *IEEE Trans. on Information Theory*, vol. 49(7), pp. 1691–1706, July 2003.

[13] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," *IEEE Trans. on Information Theory*, vol. 49(10), pp. 2658–2668, October 2003.

[14] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna broadcast channel and uplink-downlink duality," *IEEE Trans. on Information Theory*, vol. 49(8), pp. 1912–1921, Aug. 2003.

[15] W. Yu and J. M. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 1875–1892, Sept. 2004.

[16] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," *Proc. IEEE Intern. Conf. on Communication*, pp. 331–335, June 1995.

[17] S. Borst and P. Whiting, "The use of diversity antennas in high-speed wireless systems: Capacity gains, fairness issues, multiuser scheduling," *Bell Labs. Tech. Memo.*, 2001.

[18] P. Viswanath, D. N. C. Tse, and R. Laroia, "Oppurtunistic beamforming using dumb antennas," *IEEE Trans. on Information Theory*, vol. 48, pp. 1277–94, June 2002.

[19] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems," *IEEE Network Magazine*, vol. 19, pp. 43–48, Sept.-Oct. 2005.

[20] 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.

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

[22] 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.

[23] 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.

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