OPPORTUNISTIC COMMUNICATIONS WITH LIMITED FEEDBACK

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Aria Nosratinia, Chair

Dr. Naofal Al-Dhahir

Dr. John Fonseka

Dr. Hlaing Minn

I dedicate this dissertation to the memory of Tiberiu Constantinescu (1955-2005).

OPPORTUNISTIC COMMUNICATIONS WITH LIMITED FEEDBACK

by

SHAHAB SANAYEI, B.S.E.E, M.S.E.E

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2005

PREFACE

This dissertation was produced in accordance with guidelines which permit the inclusion as part of the dissertation the text of an original paper or papers submitted for publication. The dissertation must still conform to all other requirements explained in the "Guide for the Preparation of Master's Theses and Doctoral Dissertations at The University of Texas at Dallas." It must include a comprehensive abstract, a full introduction and literature review and a final overall conclusion. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported.

It is acceptable for this dissertation to include as chapters authentic copies of papers already published, provided these meet type size, margin and legibility requirements. In such cases, connecting texts which provide logical bridges between different manuscripts are mandatory. Where the student is not the sole author of a manuscript, the student is required to make an explicit statement in the introductory material to that manuscript describing the student's contribution to the work and acknowledging the contribution of the other author. The signatures of the Supervising Committee which precede all other material in the dissertation attest to the accuracy of this statement.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Aria Nosratinia, my advisor and committee chair, for supporting my research. Dr. Nosratinia's considerable effort in creating a worry-free environment in the *Multimedia and Communications Laboratory* has definitely been a key factor in academic excellence and success of myself and other graduate students. I would also like to thank the other members of my Ph.D. supervisory committee, Dr. Naofal Aldhahir, Dr. John Fonseka and Dr. Hlaing Minn for their time and effort in helping me improve the quality of this thesis.

Secondly, I would like to thank my family, my parents Sara and Rahman, my brothers Shahin, Shahriar and Shahram, and my sister Fereshteh for all of the love, support and encouragement they have given me during my studies.

I would also like to pay my tribute to Professor Tiberiu Constantinescu. During my studies in the University of Texas at Dallas, Dr. Constantinescu has been a great source of motivation, encouragement and knowledge for me and I have learned a lot in his classes.

I would like to express my gratitude towards Dr. Farid Khafizov and his family. They have been more than a host family for me and I owe them a lot. I have received a lot of encouragement and support from Farid and been inspired a lot by his ingenious way of looking into problems.

I would like to thank my friends Dr. Mohammad J. Borran and Dr. David J. Love for the fruitful discussions and collaborations I have had with them. I also like to thank all my former and current colleagues in the Multimedia and Communications Laboratory, for their friendship over the past several years.

I would like to thank Dr. Sook Kim. She has encouraged me a lot and has given me many useful tips on how to present my results to a professional audience. She is one of the most student-friendly people that I have seen during my graduate studies and I am very grateful for all the time, attention and useful advice that I received from her.

Last but definitely not least, I would like to thank Ms. Lila Foroutan for her cordial support, friendship and encouragement. She has been a great listener and often times the first one with whom I discussed my new results and papers.

August 2005

OPPORTUNISTIC COMMUNICATIONS WITH LIMITED FEEDBACK

Publication No. ____________________

Shahab Sanayei, Ph.D.
The University of Texas at Dallas, 2005

Supervising Professor: Aria Nosratinia

Opportunistic methods solve dynamic resource allocation problems by choosing only one (or a few) out of many allocation options at each point in time, thus solving the problem in sub-optimal, but efficient ways. This work explores several opportunistic methods in wireless communications and provides an analytic framework for such problems under limited feedback. An instance of opportunistic communication is antenna selection, a low-cost, low-complexity MIMO (multiple-input multiple-output) method that uses, at each point in time, the best of available antennas. We consider capacity-maximizing antenna selection algorithms, and explore their asymptotic performance at high- and low-SNR, as well as scaling laws with the number of antennas. We propose fast algorithms for joint transmit-receive antenna selection. We analyze the capacity and diversity of antenna selection algorithms under the so-called keyhole condition. In a multi-user network, opportunistic scheduling effectively captures multi-user diversity in the downlink, however, existing techniques require substantial channel state information, per user, at the base station. We propose a simple scheduling algorithm to achieve the multiuser diversity advantage with only minimal feedback requirement and prove its desirable properties in terms of the sum-rate capacity. Using this novel approach, we

propose a beamforming scheme with limited-rate feedback in multiuser MIMO systems and show its superior performance over the known methods. Finally we extend our scheduling method to multiuser OFDM networks where reduction of feedback overhead is a crucial issue.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Opportunistic techniques are greedy approaches to solve resource allocation problems in wireless networks. In these methods, resource allocation is based on the *best one or a few opportunities* among all possibilities. In this thesis, we investigate opportunistic methods for individual wireless links as well as multiuser networks.

1.1 Opportunistic Methods for the MIMO wireless Link

At the wireless link level, when there are multiple antennas at both transmit and receive side, *antenna selection* is a well-known opportunistic method to exploit the spatial diversity of MIMO at lower cost and complexity [1]. In antenna selection, from among all RF paths between the transmitter and the receiver, a few good ones are selected and communication is carried over the selected channel. The optimal solution to antennas selection problem is obtained by exhaustive search among all possibilities.

Eshaustive search has high complexity and might not be affordable at a hand-held device. One simple solution to the antenna selection is selecting antennas based on the norm of the channel vectors associated with the antennas. This is called *norm-based selection* and it has low computational complexity. Norm based selection may not be optimal especially in high signal-to-noise ratio (SNR). Nevertheless norm-based selection may be used because of its low computational complexity and known statistics [2, 3].

In an attempt to achieve near-optimal selection, Gorokhov [4] suggested a decremental selection algorithm where, starting from the full channel matrix, the rows of the channel matrix are discarded one by one so that at each step the capacity loss is minimized. Further work [5, 6] showed that an incremental algorithm (instead of a decremental one) leads to less complexity and almost the same capacity as optimal selection.

Our contribution is to explore the information theoretic limits of transmit antenna selection in the asymptote of large number of antennas for both high and low values of signal to noise ratio (SNR) [7, 8, 9]. We introduce the concept of *capacity gain* which can be used to evaluate the capacity of partially informed systems (such as transmit antenna selection) in the asymptote of high SNR [7]. We develop mathematical tools that can be utilized for asymptotic capacity analysis. Also we propose a fast algorithm for joint transmit-receive antenna selection whose complexity grows only linearly with the number of antennas and its performance is close to the optimal selection [10].

1.2 Opportunistic Methods for Multiuser Networks

In the downlink of a wireless network, where higher data rates are on demand, opportunistic scheduling is suggested as a practical way of exploiting the *multiuser diversity* [11]. In the opportunistic downlink scheduling, mobile users report the quality of their channel states to the base-station and the base-station chooses the user with *the best channel quality* for data transmission [12]. In a SISO system, it is known that choosing the user with the highest channel gain maximizes the sum-rate capacity [13]. However this scheme suffers from long delays incurred by the slow-fading environment; it may take a long time for a user to become the best user thus the fairness in this scheduling method is a problem. The method of *opportunistic beamforming with dumb antennas* proposed in [14] provides a solution to delay and fairness problems mentioned above. However, this method only exploits multiuser diversity gain and does not enjoy the advantage of array gain. The idea of phase randomization was further generalized for MIMO broadcast channel by Sharif and Hassibi [15].

In broadband networks there are abundant degrees of freedom in space, time and frequency. Thus the amount of channel state information for each user might be so large that it cannot be thoroughly fed back to the base-station. Furthermore error free transmission of the channel state information back to the base-station requires extra channel coding in the feedback channel.

These constraints motivate our research. We propose a scheduling algorithm for the flat fading case based on one bit quantization of channel gain of each user. In the proposed scheme, each mobile user compares his channel gain with the threshold value set by the base-station if it was above the threshold, the user reports a "1", otherwise it reports a "0" to the base station. The base-station from among all *eligible users* (those who are above the threshold) picks one at random and transmits to that user[1]. In our proposed scheme the base-station instead of collecting all the information and then making a decisions, shifts part of the decision-making process to the users, in other words our scheme can be considered as an instance of *distributed decision-making.*

When the quantization threshold is optimally set by the base-station, we prove that the sum-rate capacity growth of our scheduling scheme is the same as scheduling with full channel state information (CSI) [16]. Furthermore we show that the capacity gap between the two methods are insignificant when there exists a large number of users in the network.

We use this approach to devise a scheduling strategy for beamforming in a multiuser MIMO system. Despite having minimal feedback requirement, we show that our beamforming and scheduling scheme has superior performance over the known methods.

We also extend our method to broadband OFDM networks. We propose a dynamic sub-channel allocation for downlink transmission and show that with minimal

[1]The base-station can also perform round robin scheduling among all eligible users which has better fairness in small time scales.

feedback, this scheduling scheme leads to the same throughput scaling as scheduling with full CSI.

1.3 Outline of the Dissertation

In Chapter 2 we provide background material and theoretical results that will later be used for analysis.

In Chapter 3 we study antenna selection algorithms. We provide asymptotic capacity analysis for successive antenna selection algorithms. We also propose efficient low-complexity algorithms for joint transmit-receive antenna selection. We also study antenna selection over the keyhole channel.

Chapter 4 presents a novel method for exploiting multiuser diversity with limited feedback. We propose a simple scheduling scheme based on one-bit channel state information (CSI) feedback from each user. Furthermore we analyze this algorithm and show its desirable properties in terms of sum-rate capacity scaling.

In Chapter 5 we extend our approach for beamforming over a multiuser MIMO network. We prove the optimal capacity scaling of our algorithm and show that with minimal CSI feedback requirement, our method outperforms the opportunistic beamforming proposed in [14].

The objective of Chapter 6 is to show the effectiveness of our approach in reducing feedback overhead in a multiuser OFDM network. We propose a dynamic sub-channel allocation scheme for OFDM users that requires only one-bit feedback per cluster of sub-channels. We prove the optimal capacity scaling of our method. We show that our method can perform reasonably well even with correlated sub-channels.

Finally, Chapter 7 presents conclusions and discusses possible avenues for future work in this area.

CHAPTER 2
THEORETICAL BACKGROUND

This chapter is dedicated to the review of background material that is used throughout this thesis. In Section 2.1, we review the *channel capacity* formulation and the concept of *outage probability* as two important metrics that set the fundamental limits on the performance of communications systems.

In opportunistic resource allocation, we allocate resources based on *best opportunities* among all possibilities. When there is randomness in the system, the statistics of the best options determine the performance of the system. Therefore, *order statistics* are a natural tool for analysis and performance evaluation of opportunistic schemes. In Section 2.2, we review the basic theory of order statistics and present essential results that will be used in the proceeding chapters.

In many instances, asymptotic analysis provides insights into properties and trends that are not easily obtained otherwise. Furthermore, in many cases, the dimensionality of the problem is sufficiently large to justify asymptotic analysis. The behavior of order statistics in the asymptote of large samples is known to obey certain central limit laws. This topic is often referred to as *Extreme Value Theory* (EVT) and is treated in Section 2.3. We use these results for asymptotic analysis of the capacity of opportunistic systems.

2.1 Channel capacity and Outage Probability

Channel capacity was first introduced by Shannon in a landmark paper [17] as a quantity that sets limits on amount of data rate that can be reliably transmitted over a noisy channel. In the case where both the transmitter and the receiver on a wireless channel are equipped with multiple antennas (the MIMO channel), Telatar calculated the channel capacity [18]. The MIMO channel with M transmit and N receive antennas can be expressed with the following linear time invariant narrow band model

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n}$$

where the ij^{th} element of $\mathbf{H}$ represents the channel gain between transmit antenna i and receive antenna j. Assuming that the channel matrix $\mathbf{H}$ is known at the receiver, Telatar gives the following formula for the capacity of the MIMO channel [18]

$$C(\mathbf{H}, \rho) = \sup_{\mathrm{tr}(\mathbf{Q}) \leq \rho} \log_2 \det(\mathbf{I}_N + \mathbf{H}\mathbf{Q}\mathbf{H}^T) \tag{2.1}$$

where $\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{Q}$ is the input covariance matrix and ρ is the *signal-to-noise ratio* (SNR) at the receiver. When no channel state information (CSI) is available at the transmitter, it was shown in [18] that the best input covariance to maximize the mutual information is $\mathbf{Q} = \frac{\rho}{M}\mathbf{I}_M$. Hence the capacity formula for this case is

$$C(\mathbf{H}, \rho) = \log_2 \det(\mathbf{I}_N + \frac{\rho}{M}\mathbf{H}\mathbf{H}^T) \tag{2.2}$$

2.1.1 Ergodic Capacity

When the channel is *ergodic*, i.e., reliable communications is possible by using infinitely long codewords [19]. Assuming the receiver knows the channel coefficients, the maximum achievable rate can be calculated by taking the expected value of Equation 2.1

$$C_{ergodic}(\rho) = \mathbb{E}_{\mathbf{H}}[C(\rho, \mathbf{H})]$$

2.1.2 Outage Capacity and Outage Probability

When the transmitter encodes the code-word over one channel realization, there is always a non-zero probability that the channel is too poor for correct reception and decoding of the transmitted signal. In this case, reliable communication is not possible, hence the *Shannon capacity* is zero. For a given rate R, the probability of outage is defined as

$$P_{out} = \Pr[C(\rho, \mathbf{H}) < R]$$

The outage probability is an important performance metric in wireless systems as a lower bound on the frame-error rate for any coding scheme [19].

The *outage capacity* or ϵ-capacity denoted by C_ϵ is a rate below which, the outage probability does not exceed ϵ, i.e.

$$\Pr[C(\rho, \mathbf{H}) < C_\epsilon] = \epsilon$$

In the language of statistics, the outage capacity is in fact the ϵ-percent quantile value of the random variable $C(\rho, \mathbf{H})$.

2.2 Order Statistics

Suppose $\{X_n\}$ is a sequence of random variables, the corresponding order statistics are the X_i's arranged in decreasing order. We denote the smallest one by $X_{(1)}$ and the largest one by $X_{(n)}$.

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

If X_n's are i.i.d. random variables with the CDF $F(x)$ and pdf $f(x)$ then the distribution of the i^{th} order statistics is [20]

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} f(x)(F(x))^{i-1}(1-F(x))^{n-i} \tag{2.3}$$

and the CDF is

$$F_{(i)}(x) = \sum_{r=i}^{n} \binom{n}{r} (F(x))^r (1-F(x))^{n-r} \tag{2.4}$$

in particular the CDF of $X_{(1)}$ (the minimum) and $X_{(n)}$ (the maximum) is

$$F_{(1)}(x) = 1-(1-F(x))^n \quad , \quad F_{(n)}(x) = (F(x))^n \tag{2.5}$$

There is a rich literature on order statistics and their properties; a detailed treatment of the subject can be found in [20, 21].

In general $X_{(i)}$'s are jointly dependent random variables. However, when the parent distribution is the standard exponential distribution $F(x) = 1 - e^{-x}$, then the order statistics can be represented by a linear combination of *independent* exponential random variables.

Theorem 1 *Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics from the standard exponential family $F_{X_i}(x) = 1 - e^{-x}$, then*

$$X_{(i)} = \sum_{r=1}^{i} \frac{Z_r}{n - r + 1}$$

where Z_r's are i.i.d. random variables with exponential distribution $F_{Z_r}(z) = 1 - e^{-z}$.

Proof: See [20] (Section 4.6).

Theorem 1 says that the exponential order statistics form a *additive Markov chain*, as noted by Rényi [22]. Using Theorem 1 one can easily calculate the moments of order statistics for the exponential family, in particular the mean and the variance of the order statistics of the exponential family are

$$\mu_{(i)} = \mathbb{E}[X_{(i)}] = \sum_{r=1}^{i} \frac{1}{n - r + 1} \tag{2.6}$$

$$\sigma_{(i)}^2 = \mathrm{var}[X_{(i)}] = \sum_{r=1}^{i} \frac{1}{(n - r + 1)^2}. \tag{2.7}$$

For the general case, if F^{-1} denotes the inverse function of the parent CDF $F(x)$, then the i^{th} order statistics can be represented in the following way

$$X_{(i)} = F^{-1}\left(1 - \exp\left(-\sum_{r=1}^{i} \frac{Z_r}{(n - r + 1)}\right)\right) \tag{2.8}$$

where Z_i's are defined as in Theorem 1 This is known as the *Rényi Representation of order statistics* [22].

2.3 Extreme Value Theory (EVT)

It is well-known that the distribution of the sum of n i.i.d. random variables, normalized by the sample mean and variance, approaches the Gaussian distribution with zero mean and unit variance. This is known as the *Central Limit Theorem.* In this Section we show that maximum and minimum of n i.i.d. random variables also have a limit property as n goes to infinity. However, unlike the central limit theorem, neither the limit distribution nor the normalization constants are unique. The underlying theory is known as the *Extreme Value Theory (EVT)* and has been very well studies during the past decades [23, 20, 21]. In this Section we present some of the results from extreme value theory that will be used in later chapters.

Definition 1 *A CDF F is said to belong to the domain of maximal attraction of a non-degenerate cdf U if there exist sequences $\{a_n\}$ and $\{b_n > 0\}$ such that*

$$\lim_{n\to\infty} F^n(a_n + b_n x) = U(x) \tag{2.9}$$

at all continuity points of $U(x)$.

Theorem 2 *Let $X_{(n)}$ be the maximum of n i.i.d. random variables $\{X_i\}_{i=1}^n$, each with CDF F. If F belongs to the domain of maximal attraction of U, then:*

$$\frac{X_{(n)} - a_n}{b_n} \xrightarrow{d} W \tag{2.10}$$

where W is a random variable whose CDF, U, can only be one of the following three distributions:

$$U_1(x) = \exp(-e^{-x}) \quad -\infty < x < \infty\ ,$$
$$U_2(x) = \begin{cases} \exp(-x^\alpha) & x > 0, \alpha > 0 \\ 0 & x \le 0 \end{cases},$$
$$U_3(x) = \begin{cases} 1 & x > 0 \\ \exp(-(-x)^\alpha) & x \le 0, \alpha > 0 \end{cases},$$

these distributions are also known as Gumbel, Fréchet *and* Weilbull *distributions, respectively.*

The proof of the above theorem can be found in [23]. As mentioned earlier, Theorem 2 is analogous to the *central limit theorem* which was for the normalized sum of i.i.d. random variables. Although, unlike the central limit theorem, the normalization constants a_n and b_n are not necessarily unique. One way to find the constant a_n is given below [20]:

$$\begin{aligned} a_n &= F^{-1}(1-\frac{1}{n}) \\ b_n &= F^{-1}(1-\frac{1}{ne}) - F^{-1}(1-\frac{1}{n}) \end{aligned}$$

where F^{-1} is the inverse function of the parent CDF.

Example 1 *For the exponential family with CDF $F(x) = 1 - e^{-x}$, if we choose $a_n = \log n$ and $b_n = 1$ we have*

$$\Pr\left[\frac{X_{(n)} - a_n}{b_n} < x\right] = \left(1 - e^{-(x+\log n)}\right)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \longrightarrow e^{-e^{-x}}$$

as $n \to \infty$.

Example 2 *If X_n is distributed according to χ^2_{2p} with the CDF $F(x) = 1 - e_p(x) \cdot e^{-x}$, where $e_p(x) = \sum_{k=0}^{p-1} \frac{x^k}{k!}$. Then we can calculate a_n from the following equation:*

$$a_n = F^{-1}(1-\frac{1}{n})$$

but F^{-1} can not be explicitly calculated, thus we calculate a_n in the following recursive manner:

$$F(a_n) = 1 - \frac{1}{n} \quad \Rightarrow \quad e_p(a_n)e^{-a_n} = \frac{1}{n} \quad \Rightarrow \quad a_n - \log\left(e_p(a_n)\right) = \log n$$

hence $a_n = \log n + \alpha_n$ *where* $\alpha_n = o(\log n)$. *Therefore we have*

$$a_n - \log(e_p(a_n)) = \log n \quad \Rightarrow \quad \alpha_n = \log(e_p(\log n + \alpha_n)) \overset{\circ}{=} \log\left(\frac{(\log n)^{p-1}}{(p-1)!}\right)$$

hence $a_n = \log n + \log\left(\frac{(\log n)^{p-1}}{(p-1)!}\right) + \beta_n$. *It can be easily seen that* $\beta_n = o(1)$. *Similarly one can show that* $b_n = 1 + o(1)$ *therefore, the following choice of* a_n *and* b_n *leads to the desired result.*

$$a_n = \log n + \log\left(\frac{(\log n)^{p-1}}{(p-1)!}\right) \quad , \quad b_n = 1 \tag{2.11}$$

CHAPTER 3
OPPORTUNISTIC METHODS IN THE MIMO WIRELESS LINK

In a rich scattering environment, multiple antennas at transmit and receive sides can significantly improve the spectral efficiency. It has been known that the capacity of the system scales linearly with the minimum number of transmit and receive antennas [24, 18], and this has spurred a great flurry of research in recent years. However, this extended capacity is obtained using complex signal processing techniques at both ends. Furthermore, multiple antennas require multiple RF chains which are quite costly. Therefore cost and complexity are two major factors that may limit the use of multiple antennas in future communication systems.

On the other hand, not all antennas may be equally useful because in each realization of the channel, some of the antennas (either at the transmit side or the receive side) may be in deep fade. With increasing number of antennas, the probability increases that at least some of them are experiencing deep fading. Thus, a natural and practical solution offers itself: to choose a few good RF paths from among all RF paths from the transmitter to the receiver, namely, select a subset of the available antennas so that (a) the effective size of the channel matrix is reduced, thus the processing requirements are simplified, and (b) the number of RF chains, which is often the main driver of the unit cost, is also reduced [1].

At the receive side, antenna subset selection reduces the complexity. At the transmit side, antenna subset selection not only reduces the complexity, but also improves the capacity of the MIMO system [25, 26, 7, 8] at the cost of a minimal amount of feedback. Fast and efficient algorithms have been devised to determine the selected antenna subsets [4, 6, 5]. However, despite the practical importance of antenna selection, the information theoretic properties of the resulting channels remains a mostly unexplored territory. A few notable exceptions exist [27, 2], however, to date closed form expressions for capacity have not been available. We undertake the analysis of antenna selection using these algorithms.

Our motivation in taking this approach is two-fold. First, finding the optimal subset of antennas requires an exhaustive search, which is usually computationally unaffordable. Second, optimal antenna selection induces channel distributions that do not lead to a tractable formulation. Thankfully there exist antenna selection algorithms that deliver capacities almost indistinguishable from optimal selection [6, 7], which we use for our analysis.

In Section 3.3 we analyze the capacity of the antenna selection channel in the high-SNR and low-SNR regimes. In the high-SNR regime, we define the notion of *capacity gain* as the constant term in the high-SNR expansion of the capacity expression, and demonstrate that it is directly related to transmit-side channel state information. This concept was first introduced in [7, 8] and is closely related to a similar concept that was independently proposed by Lozano et al. [28]. We are able to draw conclusions about the behavior of the system in the asymptote of large number of transmit antennas, and draw comparisons between the antenna selection capacity and water-filling capacity.

Our results have interesting implications in the design and analysis of all MIMO systems (not just antenna selection). The waterfilling capacity (C_{wf}) has the same growth rate as the capacity of the uninformed transmitter (C) [29], but nevertheless $C_{wf} > C$ with non-vanishing difference at high SNR[1]. The difference is an *excess rate* that is due to channel state feedback. The excess rate is not limited to waterfilling; when partial CSI is available at the transmitter the excess rate is still there, but is smaller.

[1]We assume there are more transmit than receive antennas.

Examples of partial CSI include transmit antenna selection and channel covariance feedback. The growth of this excess rate with the number of transmit antennas is a measure of how effectively CSI is being used by a given method.

In Section 3.3.4 we employ a similar methodology for the low-SNR regime. In particular, we look at the power series expansion of the capacity around SNR=0, where the coefficient of the first-order term is called *channel gain*[2], a quantity that is related to the channel state information. Using this notion, we show that at low SNR, the optimal selection strategy at the transmitter is to select exactly one transmit antenna, regardless of other parameters. This is a new result that is reminiscent of, but distinct from, the well-known water-filling result at low SNR.

In Section 3.4, motivated by the symmetry inherent in the problem, we analyze the receive side antenna selection in a manner similar to transmit selection. Receive antenna selection always incurs a loss of receive power due to de-selected antennas,[3] thus, unlike transmit selection which may increase capacity (under conditions mentioned earlier), receive selection always incurs a capacity loss.

In Section 3.5, we propose fast algorithms for joint transmit-receive selection. We show that with a computational complexity that is only linear with respect to the number of available antennas, our proposed algorithm performs almost as well as the optimal antenna selection that has a higher complexity.

Finally in Section 3.6 we study antenna selection in keyhole channels. We show that, due to intrinsic redundancies in the keyhole channel, antenna selection can capture most of the capacity of the keyhole channel. Furthermore, we prove that the keyhole channel with antenna selection attains the same diversity order as the full system with

---

[2]Terminology due to Verdu [30]

[3]Receive antenna selection may also result in a loss of multiplexing gain, if the degrees of freedom of the channel are limited by the receiver.

no selection.

We use the following notation. $\mathbb{E}[\ ]$ refers to expected value of a random variable, I_N denotes the $N \times N$ identity matrix, $(x)^+ = \max\{x, 0\}$, and $\gamma \approx 0.57721566$ is the Euler-Mascheroni constant. We use $a_n \overset{\circ}{=} b_n$ to denote the asymptotic equivalence of a_n and b_n defined as: $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. We use the natural logarithm throughout this chapter so unless otherwise stated, the capacity unit is in Nats/Sec/Hz. The chi-square distribution with $2p$ degrees of freedom is shown by χ^2_{2p}, and the maximum of n independent χ^2_{2p} distributions is denoted by $\widetilde{\chi}^2_{2p,n}$. With an abuse of notation, we show random variables following the latter distribution also with $\widetilde{\chi}^2_{2p,n}$.

3.1 System Model

We assume a frequency non-selective (flat) linear time invariant fading channel between M transmit and N receive antennas. The signal model is:

$$\mathbf{y}(t) = \mathbf{H} \cdot \mathbf{x}(t) + \mathbf{n}(t) \tag{3.1}$$

where $\mathbf{y}(t)$ represents the $N \times 1$ received vector sampled at time t, and $\mathbf{x}(t)$ represents the $M \times 1$ vector transmitted by the antennas with power constrain $\mathbb{E}[\mathbf{x}^T\mathbf{x}] \leq \rho$, where ρ is the average SNR (per channel use), $\mathbf{n}(t)$ is the $N \times 1$ additive circularly symmetric complex Gaussian noise vector with zero mean and covariance matrix equal to $\mathbf{I}_N$ (the $N \times N$ identity matrix) and $\mathbf{H}$ is the $N \times M$ channel matrix, whose ij-th element is the scalar channel between the i-th receive and j-th transmit antenna. We assume that the elements of $\mathbf{H}$ are independent and have complex Gaussian distribution with zero mean and unit variance. We also assume that $\mathbf{H}$ is perfectly known at the receiver but it is not necessarily known at the transmitter. For antenna selection, we assume there is a rate-limited feedback channel from receiver to transmitter so that a subset of transmit antennas can be selected by the receiver and furthermore we assume that the feedback channel is without error or delay.

Input: $\rho, L, \mathbf{H}$

1. Let $\mathcal{C}_1 = \{\text{all columns of } \mathbf{H}\}$ and $\widetilde{\mathbf{P}}_1 = \mathbf{I}_N$.
2. choose $\widetilde{\mathbf{h}}_1 = \arg\max_{\mathbf{h}\in\mathcal{C}_1} \|\mathbf{h}\|_2$, $\widetilde{\mathbf{H}}_1 = \widetilde{\mathbf{h}}_1$.
3. for $i = 2 : L$
   (a) $\mathcal{C}_i = \mathcal{C}_{i-1} \setminus \{\widetilde{\mathbf{h}}_{i-1}\}$
   (b) $\widetilde{\mathbf{P}}_i = I - \widetilde{\mathbf{H}}_{i-1}(\frac{1}{\rho}\mathbf{I} + \widetilde{\mathbf{H}}_{i-1}^T\widetilde{\mathbf{H}}_{i-1})^{-1}\widetilde{\mathbf{H}}_{i-1}^T$
   (c) $\widetilde{\mathbf{h}}_i = \arg\max_{\mathbf{h}\in\mathcal{C}_i} \mathbf{h}^T\widetilde{\mathbf{P}}_i\mathbf{h}$
   (d) $\widetilde{\mathbf{H}}_i = [\widetilde{\mathbf{H}}_{i-1}\ \widetilde{\mathbf{h}}_i]$
4. $\widetilde{\mathbf{H}} = \widetilde{\mathbf{H}}_L$

Output: $\widetilde{\mathbf{H}} = \mathit{ISSA}(\rho, L, \mathbf{H})$

Figure 3.1. Incremental successive selection algorithm (ISSA)

3.2 Antenna Selection Algorithms

The complexity of MIMO transceiver algorithms increases exponentially with the number of transmit and/or receive antennas. This requires powerful signal processors at both sides which is costly especially at the mobile unit. Also in order to use all antennas at both sides, multiple RF chains are required and this requirement immensely increases the cost of the mobile unit. It is a well-known [1] that antenna selection provides a solution to the cost and complexity problem in MIMO systems. However, finding the optimal set of transmit and receive antennas requires an exhaustive search over all the possibilities, which has very high computational complexity. Successive selection algorithms [7, 8, 5, 6, 27] provide a solution to this problem, in particular the incremental successive selection algorithm (ISSA) [8, 5] seems very attractive when the number of selected antennas are much less than the total number of antennas.

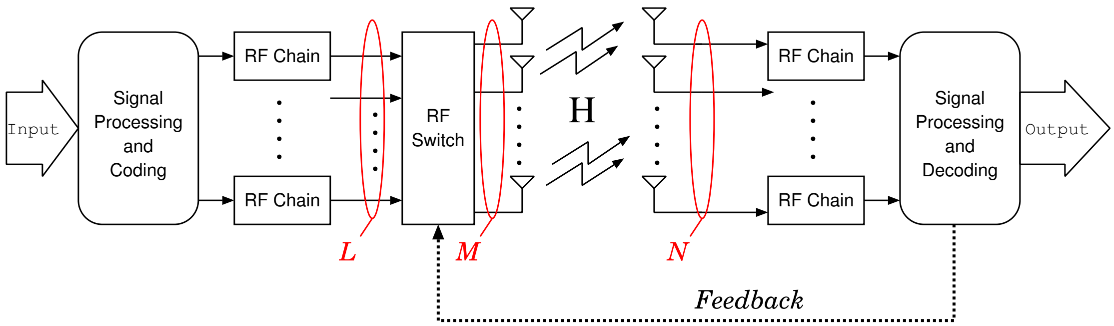

Figure 3.2. Diagram of a MIMO system with transmit antenna selection

3.2.1 Incremental Successive Selection Algorithm (ISSA)

The input of the algorithm consists of the channel realization $\mathbf{H}$, SNR per transmit antenna ρ, number of selected antennas L . The algorithm determines the selection, which are indices of the channel matrix. This is equivalent to knowing the selected channel sub-matrix $\widetilde{\mathbf{H}}$.

Note that the algorithm depicted in Figure 3.1 selects the columns of $\mathbf{H}$, in such a way that each step guarantees *the minimal loss in the capacity due to selection.* The complexity of this algorithm is $O(\max\{N_t, N_r\}N_tL)$ [5]. The same algorithm can be used for row selection (receive antenna selection) but instead of $\mathbf{H}$, $\mathbf{H}^T$ is input to the algorithm.

3.3 Transmit Antenna Selection

We consider a transmit antenna selection scheme where a subset of transmit antennas are used for transmission with equal power (Figure 3.2). For this purpose, we assume there is a rate-limited feedback channel from receiver to transmitter so that a subset of transmit antennas can be selected by the receiver and furthermore we assume that the feedback channel is without error or delay.

Optimal transmit antenna selection via exhaustive search among all $\binom{M}{L}$ combinations has complexity $O(M^L)$, which is impractical for large number of transmit antennas. One may reduce this complexity by employing a successive selection scheme, i.e., a greedy algorithm that at each step maximizes the capacity of the selected subchannel. A very similar methodology was mentioned in [4] for receive antenna selection. In this work, starting from the original channel matrix, the algorithm removes antennas one after another in a way that the capacity loss is minimized. Gharavi-Alkhansari and Greshman [5] showed that an incremental successive selection leads to less computational complexity. Simulation results show that this successive selection captures almost all the capacity of optimal antenna selection in a wide range of SNRs. Therefore, we adopt the latter algorithm for our analysis.

The input of the algorithm consists of ρ (the given SNR), L (the desired number of transmit antennas to be selected) and $\mathbf{H}$ (the original channel matrix). The output of the algorithm is $\widetilde{\mathbf{H}}$ (the channel matrix associated with selected transmit antennas). This algorithm, shown in Figure 3.1, is the basis for the following analysis.

3.3.1 A Framework for High SNR Analysis

Using the Sherman-Morisson formula for determinants [31], for the selected channel $\widetilde{\mathbf{H}}$ we have:

$$\det(\mathbf{I}_N + \frac{\rho}{L}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T) = \prod_{i=1}^{L}(1 + \frac{\rho}{L}\widetilde{\mathbf{h}}_i^T\widetilde{\mathbf{P}}_i\widetilde{\mathbf{h}}_i) \tag{3.2}$$

As $\rho \to \infty$, $\widetilde{\mathbf{P}}_i \to \mathbf{P}_i$, where, $\mathbf{P}_i = \mathbf{I}_N - \widetilde{\mathbf{H}}_{i-1}(\widetilde{\mathbf{H}}_{i-1}^T\widetilde{\mathbf{H}}_{i-1})^{-1}\widetilde{\mathbf{H}}_{i-1}$ is a projection matrix of rank $N - i + 1$. When M is also large, at each selection step, the distribution of the remaining channel vectors can still be well approximated by a circularly symmetric Gaussian distribution. Our simulations verify that for large M the Gaussianity assumption provides a good approximation for the actual distribution of the remaining columns.[4] Using this assumption we can approximate the statistics of the right side of (3.2). We know that for an uncorrelated complex Gaussian vector $\mathbf{x}$ and a projection matrix $\mathbf{P}$, $\mathbf{x}^T\mathbf{P}\mathbf{x}$ has χ^2 distribution with $rank(\mathbf{P})$ degrees of freedom. Hence for large ρ and large M, we have:

$$\det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T) \sim \prod_{i=1}^{L} \widetilde{\chi}^2_{2(N-i+1),M-i+1} \tag{3.3}$$

where $\widetilde{\chi}^2_{2p,n}$ stands for a random variable which is the maximum of n independent χ^2_{2p} random variables. The pdf of this random variable can be computed in closed form [20]:

$$f_{\widetilde{\chi}^2_{2p,n}}(x) = \frac{nx^{p-1}}{(p-1)!} e^{-x}(1 - e^{-x}e_p(x))^{n-1} \tag{3.4}$$

where $e_p(x) = \sum_{k=0}^{p-1} \frac{x^k}{k!}$.

3.3.2 Capacity gain of MIMO systems

We introduce here the concept of capacity gain as a measure of effectiveness of channel state information at the transmitter. We start with the capacity expression for a general MIMO system. Under the flat fading assumption, given a general channel matrix, the ergodic capacity of the MIMO channel is calculated as follows [18]:

$$C = \mathbb{E}[\max_{tr(\mathbf{Q})\leq\rho} \log(\det(\mathbf{I}_N + \mathbf{H}\mathbf{Q}\mathbf{H}^T))] \tag{3.5}$$

where $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ is the covariance of the transmitted vector $\mathbf{x}$.

We consider three different cases: First, uninformed transmission in which CSI is perfectly known at receiver, but not at transmitter. Second, informed transmission in which CSI is perfectly known both at transmitter and receiver. Third, transmission using an optimal subset of transmit antennas selected by the receiver. The only information available at the transmitter is the indices of the selected transmit antennas. Figure 3.3 shows the ergodic capacity and the capacity gain for the above three cases.

[4] Deriving the exact multivariate distribution of channel gains subject to selection remains an open problem.

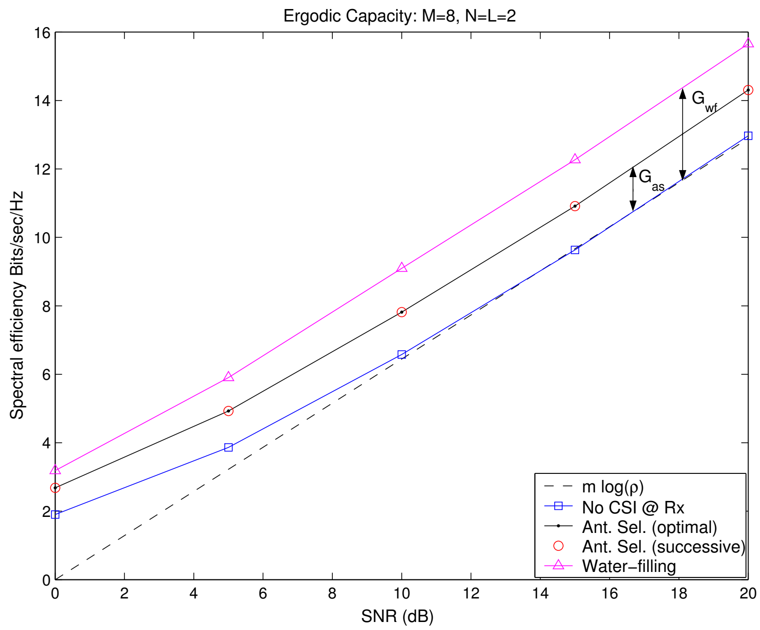

Figure 3.3. Capacity of transmit antenna selection: M=8 N=L=2

Uninformed transmitter: As shown in [18], when CSI is available only at the receiver, the covariance matrix that maximizes the capacity is of the form $\mathbf{Q} = \frac{1}{M}\mathbf{I}_N$, hence, the ergodic capacity is:

$$C = \mathbb{E}[\log(\det(\mathbf{I}_N + \frac{\rho}{M}\mathbf{H}\mathbf{H}^T))] = \mathbb{E}\left[\sum_{i=1}^{m}\log(1+\frac{\rho}{M}\lambda_i)\right]$$

where where $\lambda_1, \ldots, \lambda_m$ are ordered nonzero eigenvalues of the Wishart matrix $\mathbf{H}\mathbf{H}^T$ [18], and

$$m = \text{rank}(\mathbf{H}) = \min\{M, N\}$$

is the degrees of freedom of the MIMO channel, assuming the channel is full-rank. As shown in [24], $C = m\log\rho + O(1)$. So the ergodic capacity grows linearly with m. In the asymptotic expansion of C there is a constant term that does not vanish as $\rho \to \infty$. Thus we define the capacity gain as follows:

$$G \triangleq \lim_{\rho\to\infty}(C - m\log\rho) \tag{3.6}$$

For uninformed transmission we have:

$$\begin{aligned} G &= \lim_{\rho\to\infty}\left(\mathbb{E}\left[\sum_{i=1}^{m}\log(1+\frac{\rho}{M}\lambda_i)\right]-m\log\rho\right) \\ &= \lim_{\rho\to\infty}\mathbb{E}\left[\sum_{i=1}^{m}\log(\frac{1}{\rho}+\frac{\lambda_i}{M})\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right]-m\log M \end{aligned} \tag{3.7}$$

In the last step, the exchange of expectation and limit is allowed by the *monotone convergence theorem.*

Informed transmitter (*water-filling capacity*)**:** In this case the channel state information is available at the transmitter. The water-filling capacity of the MIMO channel is [18]:

$$C_{wf} = \mathbb{E}\left[\sum_{i=1}^{m}(\log(\mu\lambda_i))^{+}\right] \tag{3.8}$$

where μ should satisfy $\rho = \sum_{i=1}^{m}(\mu-\lambda_i^{-1})^{+}$. In large SNR scenario, all the eigen modes of the channel are used by the beamformer, hence $\mu = \frac{\rho+\sum_{i=1}^{m}\lambda_i^{-1}}{m}$ and the water-filling capacity is equal to:

$$\begin{aligned} C_{wf} &= \mathbb{E}\left[\sum_{i=1}^{m}(\log(\mu\lambda_i))\right] \\ &= m\mathbb{E}\left[\log(\rho+\sum_{i=1}^{m}\lambda_i^{-1})\right]+\mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right]-m\log m \end{aligned} \tag{3.9}$$

For large ρ we have: $C_{wf} \approx m\log\rho$. In other words, availability of CSI at transmitter side has no impact on the logarithmic growth rate of the ergodic capacity, because the growth rate only depends on the rank of the channel matrix. Now we similarly calculate the capacity gain for informed transmission:

$$G_{wf} = \mathbb{E}\left[\sum_{i=1}^{m}\log\lambda_i\right]-m\log m \tag{3.10}$$

$$\Delta G = G_{wf}-G = m\log(M/m) \tag{3.11}$$

In the asymptote of large SNR, this is *the maximum amount of excess rate obtainable by providing channel state information at the transmitter.* We note that if $M \leq N$ then $\Delta G = 0$, thus channel state information at the transmitter cannot provide any excess rate, asymptotically. This result agrees with one's intuition that beamforming is effective only when the number of transmit antennas is large. In the sequel, we only consider the interesting case of $M > N$. In particular, we are interested to understand the behavior of the capacity gain when $M \gg N$. In these cases, Equation (3.11) suggests that *at high SNR, the capacity gain can be used as an information-theoretic metric to evaluate any method that uses channel state information at the transmitter.*

Antenna selection: In the high SNR regime, one is interested in the case $L \geq N$, to maintain the degrees of freedom of the channel and prevent excessive rate loss. Suppose we have selected L ($L \geq N$) out of M transmit antennas ($M \gg N$). Then, if the selected channel is $\widetilde{\mathbf{H}}$, the capacity gain is:

$$\begin{aligned} \widetilde{G} &= \lim_{\rho \to \infty} \mathbb{E}\left[\log\left(\det(\mathbf{I}_N + \frac{\rho}{L}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T)\right)\right] - N \log \rho \\ &= \lim_{\rho \to \infty} \mathbb{E}\left[\log\left(\det(\frac{1}{\rho}\mathbf{I}_N + \frac{1}{L}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T)\right)\right] \\ &= \mathbb{E}\left[\log\left(\det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T)\right)\right] - N \log L \end{aligned} \tag{3.12}$$

3.3.3 Asymptotic Behavior of Capacity Gain

We now explore the behavior of capacity gain, for large M, in the case of informed, uninformed, and antenna selection transmitter.

Uninformed transmitter: in the case $M > N$, Equation (3.7) can be rewritten as:

$$G = \mathbb{E}\left[\log \det(\mathbf{H}\mathbf{H}^T)\right] - N \log M \tag{3.13}$$

It is known [24] that $\det(\mathbf{H}\mathbf{H}^T) \sim \prod_{i=1}^{N} \chi^2_{2(M-i+1)}$ therefore [32]:

$$G = \sum_{i=1}^{N} \left(\psi(M-i+1) - \log M\right) \tag{3.14}$$

where $\psi(n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k}$ is the *di-gamma function.* We have [24]:

$$\lim_{M\to\infty} G = 0 \tag{3.15}$$

Informed transmitter: Using Equations (3.11) and (3.15) for large M we have:

$$G_{wf} \mathrel{\mathring{=}} N\log(\frac{M}{N}) \mathrel{\mathring{=}} N\log M \tag{3.16}$$

Antenna selection: Using the results of Section 3.3.1, we can evaluate the capacity gain for antenna selection. Using (3.12):

$$\begin{aligned} \widetilde{G} &= \mathbb{E}\left[\log\det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T)\right] - N\log L \\ &\mathrel{\mathring{=}} \sum_{i=1}^{L} \mathbb{E}[\log(\tilde{\chi}^2_{2(N-i+1),M-i+1})] - N\log L \end{aligned} \tag{3.17}$$

Equation (3.17) suggests that for large M, selecting more than N antennas does not provide any further gain. In the previous section we argued that L cannot be less than N, hence for large M the optimal value for L is N. Henceforth we assume $L = N$. To evaluate the asymptotic behavior of $\widetilde{G}$, we only need to evaluate $\mathbb{E}[\log X]$, where $X \sim \tilde{\chi}^2_{p,n}$.

Using the developments provided in Appendix A, we first prove the following result which is useful for our asymptotic analysis:

Theorem 3 *Let $\{X_i\}_{i=1}^n$ be sequence of positive random variables with finite mean μ_n , finite variance σ_n, such that $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$, we also assume that for all n the logarithmic moment of X_n exits then*

$$\frac{\log(\mathbb{E}[X_n])}{\mathbb{E}[\log X_n]} \longrightarrow 1$$

Proof: From Theorem 10 for $\rho = \frac{1}{\epsilon} > 0$ we have

$$\lim_{n\to\infty} \frac{\mathbb{E}[\log(\epsilon + X_n)]}{\log(\epsilon + \mu_n)} = 1 \tag{3.18}$$

since $\log x$ is continuous on $(0, \infty)$, we have

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{\log(\epsilon + \mu_n)}{\log \mu_n} = 1 \tag{3.19}$$

Also from the *monotone convergence theorem* we can conclude that

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{\mathbb{E}[\log(\epsilon + X_n)]}{\mathbb{E}[\log X_n]} = 1 \tag{3.20}$$

from (3.18), (3.19) and (3.20) we can draw the desired conclusion.

■

Theorem 3 suggests that in order to asymptotically evaluate the logarithmic moment of a random variable satisfying the conditions of this theorem, it is sufficient to only calculate the logarithm of its mean value. In Section 2.3 we showed that the extreme value of a sequence of random variables has a limit behavior in distribution. Theorem 2 established this result in exact terms. For $\widetilde{\chi}^2_{2p,n}$ in Chapter 2.3, Example 2 we calculated the constants a_n and b_n

$$a_n = \log n + \log\left(\frac{(\log n)^{p-1}}{(p-1)!}\right) \quad , \quad b_n = 1$$

using these values we can show that

$$\frac{X_{(n)} - a_n}{b_n} \xrightarrow{d} W$$

where W is distributed according to Gumbel distribution (c.f. Theorem 2) with CDF $\Pr[W < x] = \exp(-e^{-x})$.

It is known that when the limiting distribution is of the first kind (Gumbel distribution) then Theorem 2 can also be used to evaluate the moments of $X_{(n)}$ (in fact for this case, convergence in distribution implies convergence in moments [23]). In particular we have

$$\mathbb{E}\left[\frac{X_{(n)} - a_n}{b_n}\right] \longrightarrow \mathbb{E}[W] = \gamma \tag{3.21}$$

$$\mathbb{E}\left[\left(\frac{X_{(n)} - a_n}{b_n}\right)^2\right] \longrightarrow \mathbb{E}[W^2] = \frac{\pi^2}{6} \tag{3.22}$$

where $\gamma = \lim_{n\to\infty} \log n - \sum_{k=1}^{n} \frac{1}{k} \approx 0.577$ is the Euler-Mascheroni constant. thus we have

$$\lim_{n\to\infty} \mathbb{E}[X_{(n)}] = \lim_{n\to\infty} a_n = \infty$$

and

$$\lim_{n\to\infty} \frac{\mathrm{var}(X_{(n)})}{\mathbb{E}[X_{(n)}]} = \lim_{n\to\infty} \frac{\frac{\pi}{\sqrt{6}} b_n}{a_n + \gamma b_n} = 0.$$

Hence the conditions of Theorem 3 is satisfied. Using this theorem the asymptotic growth of the logarithmic moment of the extreme order statistics of chi-square random variables is given by

$$\begin{aligned}
\mathbb{E}[\log(X_{(n)})] &\overset{\circ}{=} \log(\mathbb{E}[X_{(n)}]) \\
&\overset{\circ}{=} \log\left(\log n + \log(\log(\frac{n^{p-1}}{(p-1)!})) + \gamma\right) \\
&\overset{\circ}{=} \log(\log n)
\end{aligned} \tag{3.23}$$

Now we can evaluate the behavior of $\widetilde{G}$ in (3.17):

$$\begin{aligned}
\widetilde{G} &\overset{\circ}{=} \sum_{i=1}^{N} \mathbb{E}[\log(\tilde{\chi}^2_{2(N-i+1),M-i+1})] - N\log N \\
&\overset{\circ}{=} \sum_{i=1}^{N} \log\left(\frac{\log(M-i+1) + \log(\log(\frac{(M-i+1)^{N-i}}{(N-i)!})) + \gamma}{N}\right).
\end{aligned} \tag{3.24}$$

Hence

$$\widetilde{G} \overset{\circ}{=} N\log(\log M)$$

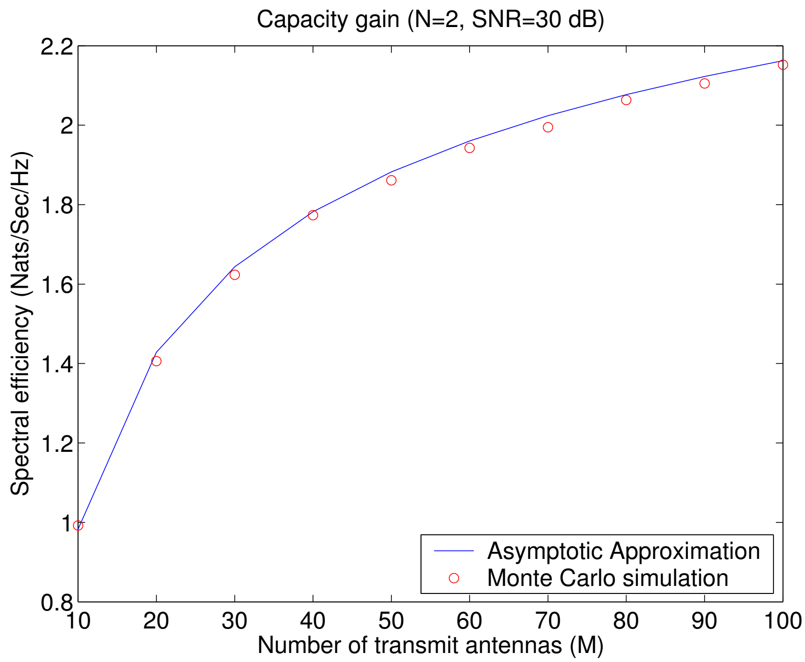

Figure 3.4. Capacity gain of antenna selection (N=2 and SNR=30 dB)

Thus the capacity gain for transmit antenna selection behaves like $O(\log(\log M))$. Also, for large number of transmit antennas Equation 3.24 can be used as an approximate formula for ergodic capacity of transmit antenna selection at high SNR, in fact

$$\widetilde{C} \approx N \log \rho + \widetilde{G} \tag{3.25}$$

Figure 3.4 compares our result with computer simulations. We run the simulation for SNR = 30 dB and $N = 2$. For each point on the plot, the capacity gain is calculated by averaging over 5000 different channel realizations and compared to the results obtained from equation (3.24). Simulations match our analysis very well, thus the asymptotic formula proves to be a useful tool for the evaluating of the capacity of transmit antenna selection.

3.3.4 Low SNR Analysis

For low SNR case, we use the concept of *channel gain*, introduced by Verdu [30], which is essentially the slope of the linear term in the Taylor expansion of the ergodic capacity, namely,

$$\Gamma \triangleq \left.\frac{\partial C(\rho)}{\partial \rho}\right|_{\rho=0} \tag{3.26}$$

since $C = \Gamma\ \rho + O(\rho^2)$, Γ can be considered as an information theoretic metric for evaluating the spectral efficiency at low SNR [30].

Uninformed transmitter: In this case the channel gain [30] is:

$$\Gamma = \mathbb{E}\left[\frac{\|\mathbf{H}\|_F^2}{M}\right] = N \tag{3.27}$$

We notice that in this case the *channel gain* does not depend on M, so increasing the number of transmit antenna will not affect the capacity.

Informed transmitter: When CSI is fully provided at the transmitter, at low SNR, the beamformer only uses the eigen mode of the channel associated with the largest eigen value of $\mathbf{H}\mathbf{H}^T$, hence the channel gain is [30]:

$$\Gamma_{wf} = \mathbb{E}[\lambda_{\max}(\mathbf{H}\mathbf{H}^T)] \tag{3.28}$$

It was first shown in [33] that $\lambda_{\max} \mathrel{\mathring{=}} (\sqrt{M} + \sqrt{N})^2$ when M or N are large. Thus for the case $M \gg N$, we have $\Gamma_{wf} \mathrel{\mathring{=}} M$.

Antenna selection: Suppose we are selecting L transmit antennas with equal power splitting among them. In low SNR scenario, the channel gain is:

$$\widetilde{\Gamma} = \mathbb{E}\left[\frac{\|\widetilde{\mathbf{H}}\|_F^2}{L}\right] = \frac{\mathbb{E}[\sum_{i=1}^{L} \|\widetilde{\mathbf{h}}_i\|_2^2]}{L}$$

where $\widetilde{\mathbf{H}}$ is the selected channel, and $\widetilde{\mathbf{h}}_1, \ldots, \widetilde{\mathbf{h}}_L$ are the L columns of $\widetilde{\mathbf{H}}$. Thus antenna selection in low SNR leads to selecting L antennas with highest norm. This is also consistent with the successive antenna selection algorithm presented in Figure 3.1 for $\rho \approx 0$. Moreover, the channel gain is maximized when $L = 1$, because

$$\frac{\mathbb{E}[\sum_{i=1}^{L} \|\widetilde{\mathbf{h}}_i\|_2^2]}{L} \leq \max_i\{\|\widetilde{\mathbf{h}}_i\|_2^2\}$$

This suggests that the optimal transmit antenna selection strategy, in the low SNR regime, is to select only one transmit antenna whose channel vector is of the highest norm. In other words $\widetilde{\mathbf{H}} = \mathbf{h}_j$ where $\|\mathbf{h}_j\| = \max\{\|\mathbf{h}_1\|, \ldots, \|\mathbf{h}_M\|\}$. For channel gain of antenna selection in the low-SNR regime, we need to evaluate $\mathbb{E}[\|\widetilde{\mathbf{H}}\|_F^2]$. The random variable $\|\widetilde{\mathbf{H}}\|_F^2$ is distributed according to $\widetilde{X}_{2N,M}$ defined in Section 3.3.1. For $M \gg N$, using Theorem 2 we have:

$$\widetilde{\Gamma}_{opt} \mathrel{\dot{=}} \log M + \log(\log(\frac{M^{N-1}}{(N-1)!})) + \gamma \mathrel{\dot{=}} \log M \tag{3.29}$$

3.4 Receive Antenna Selection

Up to this point, the emphasis has been on the transmitter side. However, it is not difficult to see that due to the reciprocity of electromagnetic propagation, the problem is highly symmetric, therefore the framework developed thus far also applies to receive antenna selection.

In particular, consider the following setup: A MIMO system with M transmit and N receive antennas, such that $N \gg M$. We wish to choose L out of N receive antennas in a way to maximize the retained capacity. The algorithm depicted in Figure 3.1 will perform the antenna selection, with the notable difference that we now must select *rows* and not columns of $\mathbf{H}$. As before, we call the selected channel $\widetilde{\mathbf{H}}$. Assuming receive antenna selection with $L = M$ and no CSI at transmitter, the capacity of the system is:

$$C = \log\det(\mathbf{I}_L + \frac{\rho}{M}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T)$$

Similarly to the previous case, we can write:

$$
\begin{aligned}
\log\det(I + \frac{\rho}{M}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^T) &\approx \log\prod_{i=1}^{M}(1 + \frac{\rho}{M}\widetilde{\chi}^2_{2(M-i+1),N-i+1}) \\
&\approx M\log\rho - M\log M + \sum_{i=1}^{M}\log\widetilde{\chi}^2_{2(M-i+1),N-i+1} \\
&= M\log\rho + \hat{G}
\end{aligned}
$$

where $\widetilde{\chi}^2_{2(M-i+1),N-i+1}$ is as earlier defined, the maximum of $N-i+1$ chi-square random variables with $M-i+1$ degrees of freedom.

Recall that in transmit selection the SNR scales by the factor ρ/L in Equation (3.2), i.e., the fewer the selected antennas, the more power can be sent through each antenna. The result was that the capacity actually increases through transmit antenna selection,[5] an increase that was characterized by $\widetilde{G}$.

In receive selection, selecting fewer antennas will result in smaller receive power, but antennas that are selected enjoy better channel distributions than the original MIMO channel. However, the loss of power cannot be made up by the improvement in the channel distributions. Therefore the capacity of the receive antenna selection is less than the capacity of the full-scale system. Unlike transmit selection, no additional capacity is obtained by selecting down to the best antennas, a result that is not surprising because information is being lost by receive selection. The above results are demonstrated on a 2×8 system in Figure 3.5.

In the low-SNR regime, we once again use the concept of channel gain, which for the receive antenna selection leads to:

$$
\widetilde{\Gamma} = \mathbb{E}\left[\frac{\|\widetilde{\mathbf{H}}\|_F^2}{M}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{L}\|\widetilde{\mathbf{h}}_i\|_2^2}{M}\right]
$$

[5]Assuming the multiplexing gain of the system is unaffected by selection, which we ensured via our transmit-selection assumption of $N \geq L \geq M$

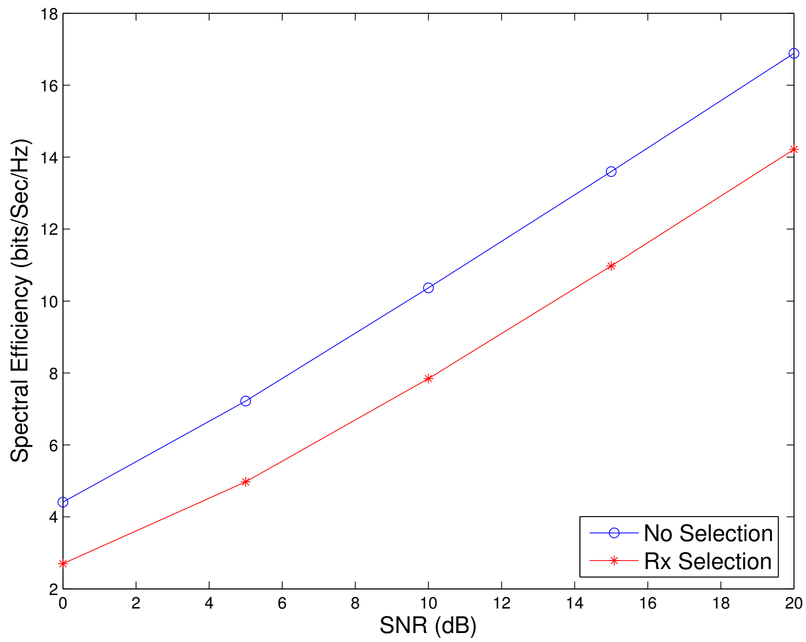

Figure 3.5. Receive antenna selection in a 2×8 system

It is evident that, unlike the transmit-selection case, there is no penalty for selecting more antennas, in fact the more antennas selected, the higher the capacity. Calculating this value using Theorem 2 we have

$$\widetilde{\Gamma} \doteq \frac{L}{M} \log N \tag{3.30}$$

3.5 Joint Transmit-Receive Selection

In this section, we consider the problem of joint transmit and receive selection in MIMO channels. We propose and study two methods. We first study a natural extension of greedy selection algorithms that have been studied in [6, 5] for receive side and in [7, 8] for transmit side. We first select the receive antennas via a greedy algorithm, and then transmit antennas selection is performed on the sub-channel resulting from receive antenna selection. The cost function of this algorithm is capacity, therefore it is assumed that signaling and outer coding will be designed to take advantage of the selected an-

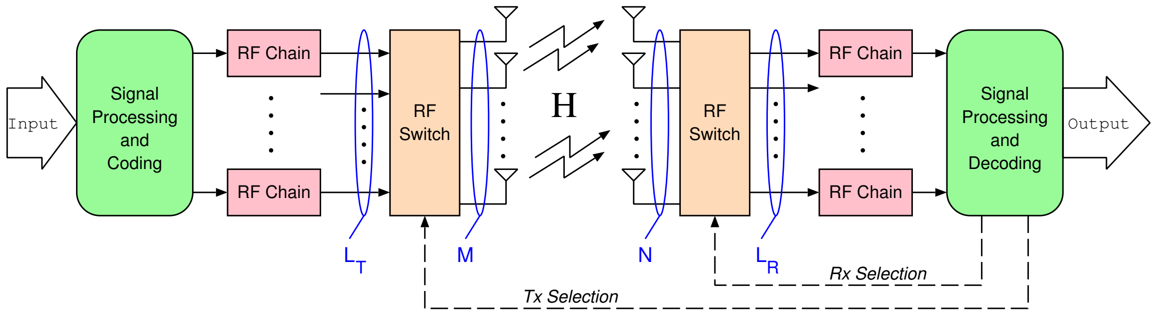

Figure 3.6. Joint transmit/receive antenna selection

tennas.[6] This algorithm has quadratic complexity in the number of antennas.

Motivated to reduce the computational complexity further, we also investigate a second approach. First, receive selection is done via a simple norm-based ranking, and then the usual incremental successive selection is performed on the reduced channel gain matrix to select the transmit antennas. Since successive selection is used only once, this algorithm has only linear computational complexity (in the number of antennas) with a performance that is very close to the previous algorithm.

Optimal joint transmit-receive selection involves an exhaustive search among $\binom{M}{L_t}\binom{N}{L_r}$ possible subsets where L_t and L_r are the number selected antennas at transmit and receive side, respectively. Hence the complexity of optimal joint selection is $O(M^{L_t}N^{L_r})$ which is too much for practical applications. Therefore we would like to devise low-complexity sub-optimal subset selection algorithms.

In the following, we compare two sub-optimal algorithms for jointly selecting the transmit and receive antennas. We assume that $L_r = L_t = L$. The rationale behind this assumption is as follows. To have the maximum spatial multiplexing gain for a given L_r, we should have $L_t \geq L_r$. Let $\mathbf{h}_1$ be an arbitrary column of $\widetilde{\mathbf{H}}$ then $\widetilde{\mathbf{H}} = [\widetilde{\mathbf{H}}_1\ \mathbf{h}_1]$.

[6] Capacity achieving coding and signaling may be non-trivial, since the selected channel is no longer Gaussian.

For high SNR we can approximate the capacity as follows [7, 8]:

$$\begin{aligned} C(\widetilde{\mathbf{H}}) &\approx L_r \log_2(\rho/L_t) + \log_2(\det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^\dagger))] \\ &= L_r \log_2(\rho/L_t) + \log_2(\det(\widetilde{\mathbf{H}}_1\widetilde{\mathbf{H}}_1^T + \mathbf{h}_1\mathbf{h}_1^T))] \\ &= L_r \log_2 \rho - L_r \log_2 L_t \\ &\quad + \log_2(\det(\widetilde{\mathbf{H}}_1\widetilde{\mathbf{H}}_1^T)) \\ &\quad + \log_2(1 + \mathbf{h}_1^T(\widetilde{\mathbf{H}}_1\widetilde{\mathbf{H}}_1^T)^{-1}\mathbf{h}_1) \end{aligned} \tag{3.31}$$

On the other hand

$$C(\widetilde{\mathbf{H}}_1) \approx L_r \log_2(\frac{\rho}{L_t - 1}) + \mathbb{E}[\log_2(\det(\widetilde{\mathbf{H}}_1\widetilde{\mathbf{H}}_1^T))]$$

thus we have

$$\begin{aligned} \Delta C &\approx C(\widetilde{\mathbf{H}}) - C(\widetilde{\mathbf{H}}_1) \\ &\approx -L_r \log_2(\frac{L_t}{L_t - 1}) \\ &+ \log_2(1 + \mathbf{h}_1^T(\widetilde{\mathbf{H}}_1\widetilde{\mathbf{H}}_1^T)^{-1}\mathbf{h}_1) \end{aligned} \tag{3.32}$$

The negative term in Equation (3.32) corresponds to the increase in power per antenna due to antenna selection, and the positive term corresponds to the fact that there are more paths for signal transmission when we use L_t antennas instead of $L_t - 1$. But when there are enough degrees of freedom (large L_t), then $\mathbf{h}_1$ can be chosen such that the logarithmic gain is dominated by the negative term in Equation (3.32). This means that there is something to gain by judiciously discarding one of the columns of $\widetilde{\mathbf{H}}$. By an inductive argument we deduce that $L_r = L_t$ is a near optimal assumption. Moreover, this assumption is convenient for many conventional space-time signaling schemes such as orthogonal space-time codes, linear receivers and VBLAST.

Input: $\rho, L_r, L_t, \mathbf{H}$

1. $\mathbf{H}_1 = \text{ISSA}(\rho/L_t, L_t, \mathbf{H})$
2. $\mathbf{H}_2 = \text{ISSA}(\rho/L_t, L_r, \mathbf{H}_1{}^T)$

Output: $\widetilde{\mathbf{H}} = \mathbf{H}_2{}^T$

Figure 3.7. Separable Tx-Rx antenna selection algorithm

3.5.1 Separable Transmit/Receive Successive Selection

An intuitive way for joint transmit/receive selection is through separating the transmit and receive selection problems. In this approach we use the ISSA algorithm introduced in Section 3.2.1 for first selecting the best L_t transmit antennas, and then selecting the best L_r receive antennas. Recall that $\text{ISSA}(\rho, L, \mathbf{H})$ is the algorithm defined in Section 3.2 which takes SNR per antenna, number of selected antennas, and the channel realization, and returns the selected channel sub-matrix. We use this algorithm in a two-step process (Figure 3.7).

Note that the order of transmit and receive selection can be interchanged, but when $L_r = L_t = L$ and $M \approx N \gg L$, this has little impact on the average performance of the system. The complexity of this algorithm is $O(\max\{M^2, N^2\}L)$.

3.5.2 Successive Joint Transmit/Receive Selection

The previous algorithm has a computational complexity that grows as a quadratic function of the number of antennas. In this section we explore options for reducing this computational complexity. In particular we propose the following sub-optimal algorithm whose complexity is linear in the maximum number of transmit and receive antennas.

We first select the maximum modulus element in the channel gain realization, $\mathbf{H}$, call it element h_{i_1,j_1}. Then order all elements in column j_1 according to their mod-

---

Input: $\rho, L, \mathbf{H}$

1. Find $(i_1, j_1) = \arg\max_{(i,j)} |h_{ij}|$
2. Find $(i_2, i_3, ..., i_N)$ such that:
   $|h_{i_2,j_1}| \geq |h_{i_3,j_1}| \geq \cdots \geq |h_{i_N,j_1}|$.
3. Form the $L \times M$ matrix, $\mathbf{G} = \begin{bmatrix} \mathbf{H}(i_1,:) \\ \mathbf{H}(i_2,:) \\ \mathbf{H}(i_L,:) \end{bmatrix}$
4. $\widetilde{\mathbf{H}} = \text{ISSA}(\rho/L, L, \mathbf{G})$

Output: $\widetilde{\mathbf{H}}$

Figure 3.8. Successive joint transmit-receive selection

ulus, and pick the top L elements in that column. The row indices of these elements are denoted $i_1, i_2, \ldots, i_L$. We form a sub-matrix using these rows, effectively selecting L receive antennas. In this sub-matrix, run the ISSA algorithm to pick the selected transmit antennas. The flow of the algorithm is depicted in Figure 3.5.2.

The key idea of the new algorithm and its difference with the previous one can be explained as follows: The previous algorithm attempted to make best (greedy) choices both in rows as well as columns of $\mathbf{H}$, thus the previous algorithm was condemned to a quadratic complexity. The new algorithm tries to make easier (yet attractive) choices in one dimension in a way that reduces computational complexity order, and yet does not have a significant impact on performance. The computational complexity of this algorithm is $O(\max\{N, M\}L^2)$ which is only linear in $\max\{N, M\}$.

Note that in the second step, one could sort along the i_1 row, and form $\mathbf{G}$ by removing unwanted columns. But the two approaches lead to the same average performance.

3.5.3 Simulation Results

For comparing the performance of the proposed algorithms, we assume a system with $M = N = 8$ and $L_r = L_t = 2$.

To compare the two algorithms, we performed a Monte-Carlo simulation using 2000 channel realizations. Figure 3.9 depicts the ergodic capacity of selected channel for optimal selection and the two sub-optimal algorithms. Algorithm I performs slightly better than Algorithm II, but its computational complexity is quadratic whereas that of Algorithm I is only linear in N. It is also observed that the selected 2×2 system obtained from any of these algorithms has a much higher capacity than a 2×2 Rayleigh fading channel with no selection.

Figure 3.10 shows the outage probability of the optimal selection and the two sub-optimal algorithms. In low SNR Algorithm I and II perform almost the same and they are only slightly worse than the optimal selection in. But as the SNR increases the difference between optimal and sub-optimal selection becomes more noticeable, although the two sub-optimal algorithms still perform very close to each other.

3.6 Antenna Selection in Keyhole Channels

The performance of MIMO channel may severely be degraded due to a condition known as *keyhole* or *pinhole* effect [34, 35]. Under keyhole condition the capacity scaling of the MIMO channel (with respect to SNR, the signal-to-noise ratio) is no better than a SISO channel. Even when the antennas are all uncorrelated, the keyhole condition may exist [34]. In such a case the channel loses its spatial degrees of freedom and the channel matrix becomes rank deficient.

When the MIMO channel is in the keyhole condition, MIMO techniques do not lead to higher capacities. In such a case, antenna selection at both ends can be employed

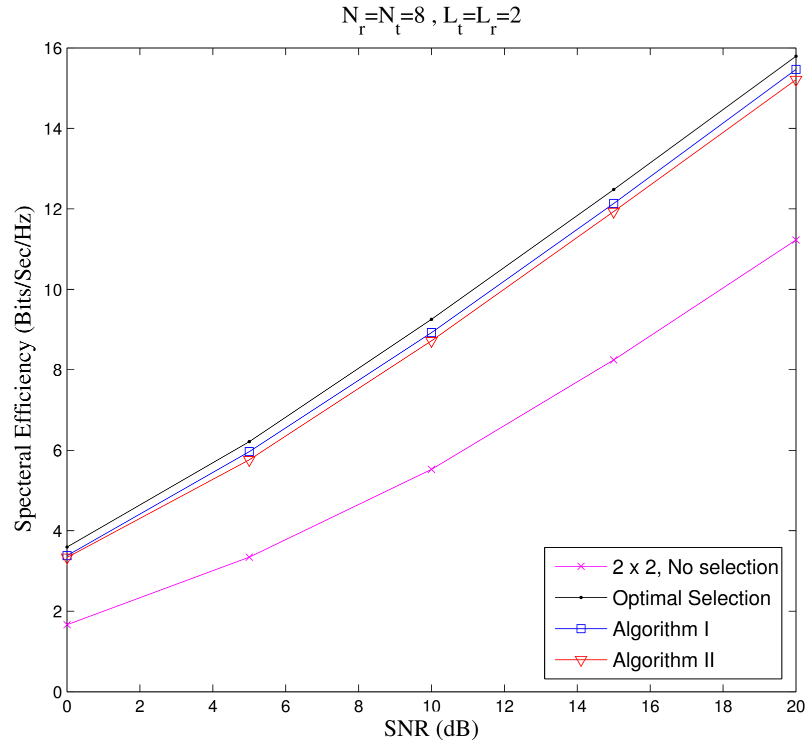

Figure 3.9. Ergodic capacity for joint Tx-Rx selection

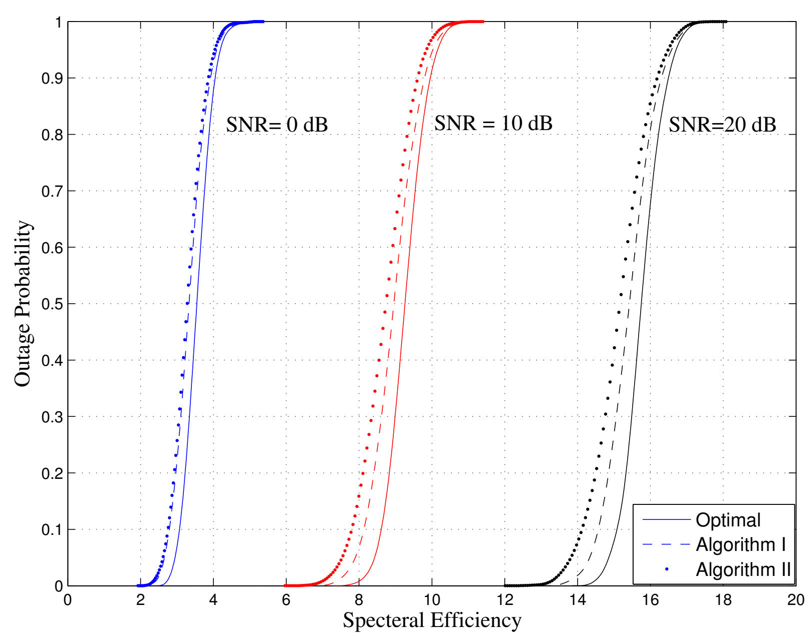

Figure 3.10. Outage probability for joint Tx-Rx selection

to capture the intrinsic redundancy of the keyhole channel. This is also attractive from a practical point of view because of the reduction in cost and complexity of data transmission [1]. Although for transmit antenna selection a low rate feedback from the receiver to the transmit is required. In this section we study antenna selection in keyhole channels. We wish to find a subset of antennas, both at the transmitter and the receiver, such that the selected sub-channel captures most of the capacity of the original keyhole channel.

In the keyhole channel, the channel matrix $\mathbf{H}$ has rank 1 thus it has only one nonzero eigenvalue. A general model for the keyhole channel is

$$\mathbf{H} = \mathbf{h}_r \mathbf{h}_t^T \tag{3.33}$$

where $\mathbf{h}_t \in \mathbb{C}^M$ and $\mathbf{h}_r \in \mathbb{C}^N$ are column vectors and each consisting of i.i.d. elements distributed as $\mathcal{CN}(0,1)$. Furthermore $\mathbf{h}_r$ and $\mathbf{h}_t$ are assumed to be independent. The received signal to noise is denoted by ρ. We also assume a low-rate but reliable feedback path from the receiver to the transmitter is available. Assuming equal power splitting among antennas, the capacity of the keyhole channel is:

$$C = \log_2 \det(I_N + \frac{\rho}{M}\mathbf{H}^T\mathbf{H}) = \log_2(1 + \frac{\rho}{M}\|\mathbf{h}_t\|_2^2 \ \|\mathbf{h}_r\|_2^2) \tag{3.34}$$

thus the instantaneous SNR, normalized by the received SNR can be defined as

$$\gamma \triangleq \frac{\|\mathbf{h}_t\|_2^2}{M}\|\mathbf{h}_r\|_2^2 \tag{3.35}$$

and the capacity (conditioned on $\mathbf{H}$) can be obtained as

$$C = \log_2(1 + \rho\gamma) \tag{3.36}$$

In order to maximize the capacity it is only sufficient to maximize γ. We notice that random variables $\|\mathbf{h}_t\|_2^2$ and $\|\mathbf{h}_r\|_2^2$ are distributed as χ^2_{2M} and χ^2_{2N} respectively. Thus the average normalized SNR for the keyhole channel is

$$\mathbb{E}[\gamma] = N \tag{3.37}$$

An important consequence of Equation (3.35) is that joint transmit and receive antenna selection in the keyhole channel *decouples* into MISO and SIMO antenna selection problems, i.e., the selected (sub)channel can be represented as $\widetilde{\mathbf{H}}_{L\times K} = \widetilde{\mathbf{h}}_r\widetilde{\mathbf{h}}_t^T$, where $\widetilde{\mathbf{h}}_t \in \mathbb{C}^K$ and $\widetilde{\mathbf{h}}_r \in \mathbb{C}^L$ are the channel vectors of selected antennas at the transmit and receive side. The normalized instantaneous SNR for the selected channel is

$$\widetilde{\gamma} = \frac{\|\widetilde{\mathbf{h}}_t\|^2}{K}\|\widetilde{\mathbf{h}}_r\|^2 \tag{3.38}$$

The average per-element energy of a vector is less than the energy of the largest vector element, that is:

$$\frac{\|\widetilde{\mathbf{h}}_t\|^2}{K} \leq \max_i |\mathbf{h}_t(i)|^2 \tag{3.39}$$

Therefore the best strategy is to select only one antenna at the transmitter; the one with the highest channel gain.[7] Transmit antenna selection is possible with a small feedback of $\lceil \log M \rceil$ bits. For $K = 1$ and $L = N$ it can be easily shown that

$$\mathbb{E}[\widetilde{\gamma}] = N\sum_{j=1}^{M}\frac{1}{j} \tag{3.40}$$

The above harmonic sum, when compared with Eq. (3.37), characterizes the gain of transmit antenna selection over the full channel without CSI.

We now proceed to receive-side selection. The diversity obtained by selecting L out of N antennas is a known as *generalized selection diversity* [36] and has been extensively studied in the literature. Using results from [36] we can calculate the average normalized SNR for antenna selection with $K = 1$ and arbitrary L:

$$\begin{aligned}\mathbb{E}[\widetilde{\gamma}] &= \sum_{k=1}^{M}\frac{1}{k}\left(L + L\sum_{j=L+1}^{N}\frac{1}{j}\right) &(3.41)\\ &= g(L) &(3.42)\end{aligned}$$

[7]With complete CSI at transmitter it is possible to do even better via beamforming. However that requires significant feedback rate as well as a more complex transmitter.

Table 3.1. Number of selected receive antennas to match a baseline system. In all cases only one transmit antenna is selected.

# of Rx Ant. / # of Tx Ant.	2	3	4	5	6	7	8	9	10
2	1	1	1	2	2	2	2	3	3
3	1	1	1	1	2	2	2	2	2
4	1	1	1	1	1	1	1	2	2
5	1	1	1	1	1	1	1	1	2
6	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1

As mentioned earlier, we wish to find the minimum number of selected receive antennas such that the equivalent SNR is no less than $E[\gamma]$. Considering that $g(L)$ above is a monotonic function, the number of selected antennas will be

$$\hat{L} = \begin{array}{c} \min \;\; g^{-1}(\tilde{\gamma}) \\ \text{subject to } \tilde{\gamma} \geq E[\gamma] \end{array}$$

Table 3.1 shows the calculated values of $\hat{L}$ for various systems. It is interesting to observe that across a large group of systems, a small number of selected antennas is sufficient to match the capacity of a baseline system with no transmit CSI, but with full hardware at both sides. Antenna selection requires a small amount of feedback, but has considerably smaller hardware requirements.

Figure 3.11 shows the ergodic capacity of the keyhole channel for 1)$M = N = 5$ and 2) $M = 5, N = 10$ with and without antenna selection. The Monte-Carlo simulation is performed over 5000 independent channel realizations. The simulation results show that at the cost of a few bits of feedback, a keyhole channel can be reduced to a low order SIMO channel without a considerable loss in capacity.

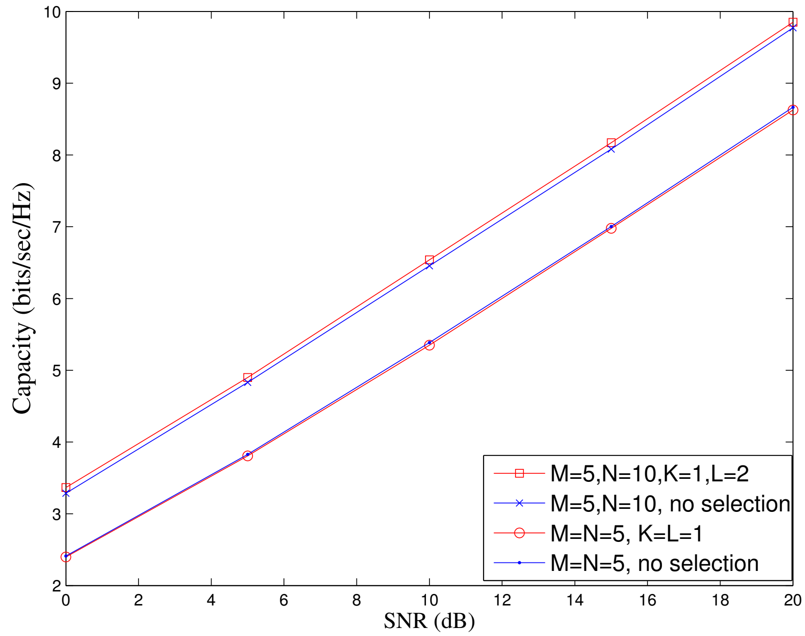

Figure 3.11. Ergodic capacity of the keyhole channel vs. SNR

3.6.1 Diversity Order

It has already been stated [37] and proved [38] that the diversity order of the keyhole channel is $\min(M, N)$. In this section we prove that antenna selection has no impact on the diversity order of the keyhole channel. The outage probability for the keyhole channel is:

$$\begin{aligned} P_{out} &= Pr[\mathcal{I} < R] = Pr[\log(1+\rho\gamma) < R] \\ &= Pr\left[\gamma < \frac{2^R - 1}{\rho}\right] = F_\gamma\left(\frac{2^R-1}{\rho}\right) \end{aligned} \tag{3.43}$$

where $F_\gamma(\cdot)$ is the CDF of the random variable γ. The diversity order is defined as:

$$d = \lim_{\rho\to\infty} \frac{-\log P_{out}}{\log \rho} \tag{3.44}$$

therefore Equation (3.43) suggests that the diversity order is equal to the exponent of the lowest order term in the asymptotic expansion of the $F_\gamma(a)$ for small enough a where $a \triangleq \frac{2^R-1}{\rho}$, i.e. $F_{\tilde{\gamma}}(a) = a^d + o(a^d)$.

Theorem 4 *The diversity order of the keyhole channel with antenna selection is* $\min(M,N)$*. In particular, the outage probability with antenna selection has the following asymptotic behavior:*

$$P_{out}(a) \overset{\circ}{=} \begin{cases} a^N & N < M \\ a^M & N > M \\ a^N \log(\frac{1}{a}) & N = M \end{cases}$$

In order to prove Theorem 4 we need the following lemma.

Lemma 1 *In the asymptote of small* a *we have*

$$\int_0^\infty \frac{e^{-(x+\frac{a}{x})}}{x^{\nu+1}} dx \overset{\circ}{=} \begin{cases} \frac{\Gamma(\nu)}{a^\nu} & \nu > 0 \\ -\log a & \nu = 0 \\ \Gamma(-\nu) & \nu < 0 \end{cases} \tag{3.45}$$

Proof: We have:

$$\int_0^\infty \frac{e^{-(x+\frac{a}{x})}}{x^{\nu+1}} dx = \frac{2}{a^{\frac{\nu}{2}}} K_\nu(2\sqrt{a}) \tag{3.46}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind [39]. For small z, the following asymptotic formulae hold ([39] page 375, Equation 9.6.6., 9.6.8 and 9.6.9.):

$$K_\nu(z) \overset{\circ}{=} \begin{cases} \frac{1}{2}\Gamma(\nu)(\frac{z}{2})^{-\nu} & \nu > 0 \\ -\log \frac{z}{2} & \nu = 0 \\ \frac{1}{2}\Gamma(-\nu)(\frac{z}{2})^{\nu} & \nu < 0 \end{cases} \tag{3.47}$$

Now let $z = 2\sqrt{a}$, combining Equation (3.46) and Equation (3.47) we arrive at Equation (3.45).

■

Proof of Theorem 4: We have

$$\max_i \{|\mathbf{h}_r(i)|^2\} \leq \|\widetilde{\mathbf{h}}_r\|_2^2 \leq \|\mathbf{h}_r\|_2^2 \tag{3.48}$$

hence

$$\log_2(1+\rho\hat{\gamma}) \leq \log_2(1+\rho\widetilde{\gamma}) \leq \log_2(1+\rho\gamma) \tag{3.49}$$

where

$$\begin{aligned}\hat{\gamma} &= \max_i\{|\mathbf{h}_t(i)|^2\}\max_i\{|\mathbf{h}_r(i)|^2\} \\ \widetilde{\gamma} &= \max_i\{|\mathbf{h}_t(i)|^2\}\|\widetilde{\mathbf{h}}_r\|_2^2 \\ \gamma &= \|\mathbf{h}_t\|_2^2\|\mathbf{h}_r\|_2^2\end{aligned}$$

This means that for all $a \geq 0$.

$$F_{\widehat{\gamma}}(a) \leq F_{\widetilde{\gamma}}(a) \leq F_{\gamma}(a) \tag{3.50}$$

It was shown in [38] that the diversity order of the keyhole channel (that corresponds to γ) is $\min(M, N)$. We only need to prove similar results for the lower bound $\widehat{\gamma}$. Note that $\widehat{\gamma}$ corresponds to antenna selection with $K = L = 1$. In other words, it suffices to prove that for the keyhole channel, selecting the best antennas at both the transmit and the receive sides provides full diversity. $\widehat{\gamma}$ is the product of two independent random variables $\widehat{\gamma}_1 = \max_i\{|\mathbf{h}_t(i)|^2\}$ and $\widehat{\gamma}_2 = \max_i\{|\mathbf{h}_r(i)|^2\}$ each of which is the extreme value of exponentially distributed independent random variables. Hence the pdf of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ (denoted by $f_1(\cdot)$ and $f_2(\cdot)$, respectively) are given by

$$f_1(x) = Me^{-x}(1-e^{-x})^{M-1} \tag{3.51}$$

$$f_2(x) = Ne^{-x}(1-e^{-x})^{N-1} \tag{3.52}$$

The pdf of $\widehat{\gamma}$ can be calculated as follows:

$$f_{\widehat{\gamma}}(a) = \int_0^\infty f_1(x)f_2(\frac{a}{x})\frac{dx}{x} \tag{3.53}$$

$$= \int_0^\infty f_2(x)f_1(\frac{a}{x})\frac{dx}{x} \tag{3.54}$$

Case I, M > N : In this case we use Equation (3.53)

$$f_{\widehat{\gamma}}(a) = MN\int_0^\infty e^{-(x+\frac{a}{x})}(1-e^{-x})^{M-1}(1-e^{-\frac{a}{x}})^{N-1}\frac{dx}{x} \tag{3.55}$$

Using the Taylor series expansion, we expand the exponentials:

$$\begin{aligned}
& (1-e^{-x})^{M-1}(1-e^{-\frac{a}{x}})^{N-1} = \\
& \left(x-\frac{x^2}{2!}+\cdots\right)^{M-1}\left(\frac{a}{x}-\frac{a^2}{2!x^2}+\cdots\right)^{N-1} \\
= \; & a^{N-1}x^{M-N}\left(1+\sum_{j\geq 0,k\leq j} c_{jk}a^j x^k\right)
\end{aligned} \tag{3.56}$$

Because the right hand side of Equation (3.56) is less than one, the *dominated convergence theorem* allows us to calculate Equation (3.55) via term by term integration of the left hand side of Equation (3.56). Hence

$$f_{\widehat{\gamma}}(a) = I_1(a) + I_2(a) \tag{3.57}$$

where

$$I_1(a) = MN\ a^{N-1}\int_0^\infty \frac{e^{-(x+\frac{a}{x})}}{x^{N-M+1}}dx \tag{3.58}$$

and,

$$I_2(a) = MN\sum_{j\geq 0,k\leq j} c_{jk}a^{N+j-1}\int_0^\infty \frac{e^{-(x+\frac{a}{x})}}{x^{N-M-k+1}}dx \tag{3.59}$$

From Lemma 1, $I_1(a) \stackrel{\circ}{=} a^{N-1}$; also each term in the sum in Equation 3.59 is $O(a^{N+p-1})$ for some $p \in \{0,1,2,\cdots\}$. Since $I_2(a)$ converges, we conclude that $I_2(a) \stackrel{\circ}{=} a^{N-1}$ thus $f_{\widehat{\gamma}}(a) \stackrel{\circ}{=} a^{N-1}$. On the other hand, using the L'Hospital's rule we know that in the asymptote of small a

$$F_{\widehat{\gamma}}(a) \stackrel{\circ}{=} a f_{\widehat{\gamma}}(a) \tag{3.60}$$

Thus when $M > N$, we have $F_{\widehat{\gamma}}(a) \stackrel{\circ}{=} a^N$, i.e. the diversity order is $N = \min(M,N)$.

Case II, N > M : In this case we use Equation (3.54). Considering the symmetry of the equation with respect to M and N, the problem is reduced to Case I with M and N interchanged, thus the diversity order again is $M = \min(M,N)$.

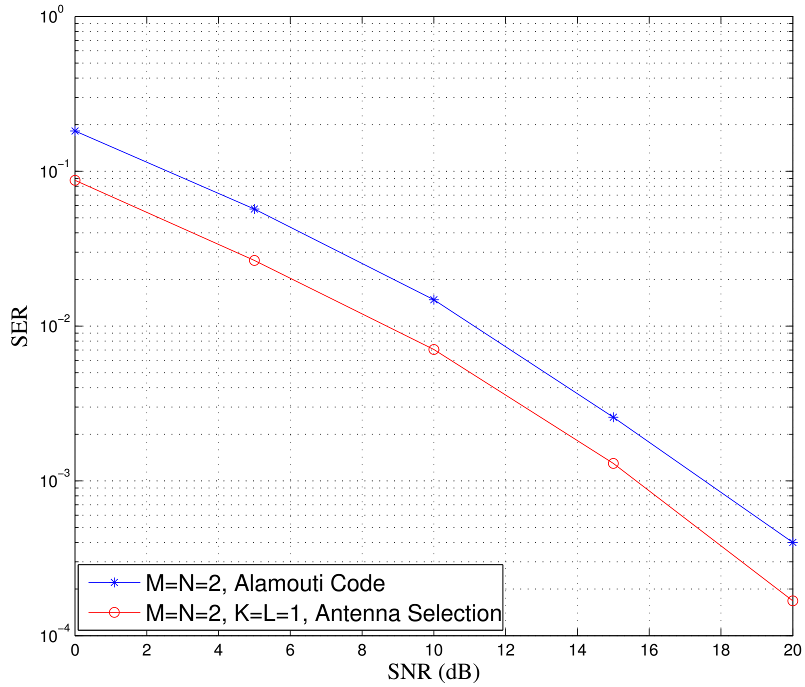

Figure 3.12. Symbol error rate (SER) vs. SNR for BPSK

Case III, N = M : Using Lemma 1 and an argument similar to Case I, we have $f_{\widehat{\gamma}}(a) \overset{\circ}{=} a^{N-1}\log(\frac{1}{a})$. Thus $F_{\widehat{\gamma}}(a) \overset{\circ}{=} a^{N-1}\log(\frac{1}{a})$.

$$F_{\widehat{\gamma}}(a) \overset{\circ}{=} \begin{cases} a^N & N < M \\ a^M & N > M \\ a^N \log(\frac{1}{a}) & N = M \end{cases}$$

But this is exactly the same as the asymptotic behavior of $F_{\gamma}(\cdot)$ for a general keyhole channel [38]. In order words $F_{\widehat{\gamma}}(a) \overset{\circ}{=} F_{\gamma}(a)$. Thus Equation (3.50) necessitates that $F_{\widetilde{\gamma}}(a)$ also have the same asymptotic behavior and this proves the theorem.

Notice that when $M = N$, then the diversity behavior can not be described by an integer. For such a case, the keyhole channel behaves better than a channel of diversity order $N - 1$, but its performance is not as good as a channel with diversity N. This is exactly the same result obtained in [38].

To demonstrate the diversity order of the antenna selection channel, we refer the reader to Figure 3.12, which depicts the symbols error rate of an uncoded keyhole channel with 2 transmit and 2 receive antennas ($M = N = 2$). We show the performance of Alamouti code using 2×2 antennas, vs. antenna selection system that at each point in time selects down to a SISO system (1×1). Both have diversity two, but we also see the the logarithmic penalty predicted by [38] and Theroem 4. Antenna selection provides 2 dB gain over Alamouti. Antenna selection requires a feedback of $\log M = 1$ bits per fading state, to select the best transmit antenna. Since the fading states vary much slower than the symbol transmission rate, the equivalent feedback rate is small.

CHAPTER 4

OPPORTUNISTIC METHODS IN MULTIUSER NETWORKS

The usual forms of diversity in single-user channels include time, frequency, and space diversity. In a multi-user environment with multiple independent wireless links, it is highly probable that at any given point in time, at least one of those links has high quality. This advantage is called *multiuser diversity.* Obviously, multi-user diversity requires the base station to know the channel coefficients for all users, which is usually estimated at the mobiles and fed back to the base station.

The notion of multiuser diversity was raised by Knopp and Humblet [12] for the uplink, where they mentioned that the best strategy is to always transmit to the user with the best channel. Tse [40] provided similar result for the downlink. Bender *et al.* [11] examined practical aspects of downlink multi-user diversity in the context of IS-95 CDMA standard. Viswanath, Tse and Laroia [14] examined this problem for the downlink and presented a method of opportunistic beamforming via phase randomization. Hochwald, Marzetta and Tarokh [41] investigate the problem of scheduling and rate feedback in the case of MIMO channels. Sharif and Hassibi [15] generalized the opportunistic beamforming of [14] to the case where mobiles also have multiple antennas.

Some of these works consider the question of the required feedback rate, but to our knowledge, only [41] and [42] explicitly consider the question of quantifying the required feedback. However, they do not consider capacity growth rates, nor optimize the quantization to minimize capacity loss. In this chapter we present, effectively, a one-bit quantization strategy and the associated scheduling algorithm that guarantees optimal capacity growth rate. The idea of exploiting multiuser gain by limited feedback was first proposed in [15]. In [42, 43] the idea of thresholding for reducing the feedback load required to exploit multiuser diversity has been proposed, however, as will be discussed in Section 4.2, their scheduling scheme and the amount of information fed back to the base station are different from our scheme. In particular, our method guarantees optimal growth rate with number of users via one-bit fixed-rate feedback, while[42, 43] requires a variable-rate feedback of real-valued numbers and, to our understanding, it has not been proved to guarantee optimal growth rates.

4.1 System Model

We consider a multi-user cellular network with n users who receive data from the base station. We assume the block fading model for each user's channel. The channel state information of each user is assumed to be fully known to that user, and it is constant over a coherence interval of length T. We assume a SISO case in which all users and also the base station are each equipped with only one antenna. Under the block-fading assumption, we have the following model for received signal for each user:

$$y_i(t) = \sqrt{\rho_i} h_i s_i(t) + z_i(t) \tag{4.1}$$

In the above model, we assume that $s_i(t) \in \mathbb{C}^T$ is the vector of transmitted symbols of the i^{th} user at time t with power constraint $\mathbb{E}[\|s_i(t)\|^2] = T$, and $y_i(t) \in \mathbb{C}^T$ is the received signal of the i^{th} user at time t, $z_i(t) \sim \mathcal{CN}(0, I_T)$ is the i.i.d. complex Gaussian noise, h_i is the channel gain of the i^{th} user, which is assumed to be zero mean circularly symmetric complex Gaussian random variable with unit variance per dimension. We also assume that users have mutually independent channel gains. Moreover we assume a homogeneous network in which all users have the same SNR, i.e. $\rho_i = \rho$. We also assume that for each user there exits a low-rate but reliable and delay-free feedback channel to the base-station.

4.2 Scheduling via Single-bit Feedback

The base-station sets a threshold α for all users. Each user compares the absolute value of their channel gain to this threshold. Whenever the channel gain exceeds the threshold, a "1" will be transmitted to the base station; otherwise a "0" will be transmitted. [1] The base station receives feedback from all users and then randomly picks a user whose feedback bit was set to one for data transmission.[2] If all the feedback bits received by the base-station are zero, then no signal is transmitted in that interval.[3]

Our work is distinct from that of Gesbert and Alouini [42, 43] in the following manner. Even though the idea of thresholding the users' channel SNR's has been mentioned by Gesbert and Alouini, the requirements for their scheduling scheme are considerably different from ours. In particular, their method requires the users that have channel gains above a certain threshold to report those channel gains to the base station. This requires a feedback channel of variable-rate, but more importantly, a feedback channel that must still accommodate the transmission of real-valued variables back to the base station. So even though in their scheme, fewer parameters than before are transmitted, still the rate is considerable. In comparison, we are interested in a strictly limited-rate feedback scenario.

4.3 The Sum-Rate Capacity

Upon receipt of each set of feedback bit, the base-station only transmits to users whose channel gain is above the threshold α. Let $p = Pr[|h_i|^2 > \alpha] = e^{-\alpha}$ since the channel gains are all mutually independent, the probability of having k feedback bits equal to

[1] In this case, the user can simply stay silent.

[2] The scheduling to users with favorable channels may also be implemented via round robin. In the long run, both these strategies have the same average throughput per user. However, the round-robin version may be more appealing from a fairness point of view.

[3] In this case the base station can also randomly pick a user for data transmission, although for large number of users this has vanishing advantage over no transmission.

one obeys a binomial law, i.e.

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \tag{4.2}$$

The ergodic capacity upon receiving k ones by the base-station is:

$$\begin{aligned}\overline{C}_k &= \sum_{i=1}^{k} Pr[\text{the } i^{th} \text{ best user is selected}]\ C_i \\ &= \frac{1}{k}\sum_{i=1}^{k} C_i \end{aligned} \tag{4.3}$$

where $C_i = \int_0^\infty \log(1+\rho x) dF_i(x)$ and $F_i(x)$ is the CDF of the i^{th} highest absolute value of all channel gains. In other words if $\{X_1, \ldots, X_n\}$ is a permutation of $\{|h_1|^2, \ldots, |h_n|^2\}$ such that $0 \leq X_n \leq \cdots \leq X_1$, then $F_i(x) = Pr[X_i < x]$. When the channel gains are i.i.d., from Equation 2.4 we have:

$$F_i(x) = \sum_{l=0}^{i-1} \binom{n}{l} (F(x))^{n-l} (1-F(x))^l \tag{4.4}$$

where $F(x) = 1 - e^{-x}$ is the CDF of $|h_i|^2$ for $i = 1, \ldots, n$. Thus the sum-rate capacity of the network with one-bit feedback can be formulated as:

$$C_{1\_bit} = \sum_{k=1}^{n} p_k \overline{C}_k \tag{4.5}$$

4.4 The Optimal Threshold

The sum-rate capacity is a function of ρ, p and n. On the other hand the threshold α is uniquely determined by p from the following formula,

$$\alpha = F^{-1}(1-p) \tag{4.6}$$

For Rayleigh fading channel, the channel magnitude squared obeys an exponential law

$$\alpha = -\log p \tag{4.7}$$

In order to find the optimal threshold we choose p such that the sum-rate capacity C_{1_bit} is maximized. The cost function $C_{1\_bit}(p)$ is a weighted sum of functions of the form $p^k(1-p)^{n-k}$ which are all concave over the interval $[0,1]$, hence C_{1_bit} is a concave function of p and it has a unique maximum over the interval $[0,1]$. To calculate the value of p that maximizes the sum-rate capacity, we must solve $\frac{\partial C_{1\_bit}(p)}{\partial p}=0$ for p, i.e.,

$$\sum_{k=1}^{n}(k-np)p_k\overline{C}_k=0 \tag{4.8}$$

A closed form solution to this equation is in general not tractable. However, a numerical solution is possible with $O(n)$ complexity.

4.5 Asymptotic Analysis

In this section we explore the asymptotic behavior of the sum-rate capacity. We use Theorem 10 (Chapter 2), which states that if the probability measure associated with the random variable X_n is well concentrated around its mean value for large n, then Jensen's inequality for the Shannon function $\log(1+\rho x)$ is asymptotically tight. Note that X_n can be either a *discrete* or a *continuous* random variable.

When channel state information is fully available at the base station, the base station only transmits to the user with the best channel, hence the ergodic sum-rate capacity of the network can be calculated by the following formula:

$$\begin{aligned} C_{full\_CSI} &= C_1=\int_0^{\infty}\log(1+\rho x)dF_1 \\ &= n\int_0^{\infty}\log(1+\rho x)e^{-x}(1-e^{-x})^{n-1}dx \end{aligned}$$

Let $\mu_1=\int_0^{\infty}xdF_1$ and $\sigma_1^2=\int_0^{\infty}(x-\mu_1)^2dF_1$, then it is known (Equation 2.6 and Equation2.7) that: $\mu_1=\sum_{i=1}^{n}\frac{1}{i}$ and $\sigma_1^2=\sum_{i=1}^{n}\frac{1}{i^2}$, therefore $\mu_1\to\infty$ and $\frac{\sigma_1}{\mu_1}\to 0$ as

$n \to \infty$. From Theorem 10, it follows that:

$$\begin{aligned} C_{full\_CSI} &\stackrel{\circ}{=} \log(1+\rho\mu_1) \\ &\stackrel{\circ}{=} \log(\log n) + \log\rho. \end{aligned} \tag{4.9}$$

where $\stackrel{\circ}{=}$ indicates asymptotic equivalence, as defined earlier.

We are interested to investigate the behavior of the sum-rate capacity of the 1-bit feedback scheduling proposed in Section 4.2 in the asymptote of large number of users. This is accomplished via the following result.

Theorem 5 *The sum-rate capacity of a wireless network with 1-bit feedback and optimal choice of threshold, behaves as* $\log(\log n) + \log\rho$, *exactly the same as the sum-rate capacity of a fully informed network.*

Proof: Equation (4.8) can be re-written as:

$$C_{1\_bit} = \frac{1}{np}\sum_{k=1}^{n} kp_k\overline{C}_k \tag{4.10}$$

For a p satisfying Equation (4.10) we have

$$\begin{aligned} C_{1\_bit} &= \frac{1}{np}\sum_{k=1}^{n} kp_k\overline{C}_k \\ &= \frac{1}{np}\sum_{k=1}^{n} kp_k\left(\frac{1}{k}\sum_{i=1}^{k} C_i\right) \\ &= \frac{1}{np}\sum_{k=1}^{n}\sum_{i=1}^{k} p_kC_i \\ &= \sum_{i=1}^{n}\left(\frac{1}{np}\sum_{k=i}^{n} p_k\right)C_i \end{aligned} \tag{4.11}$$

Notice that $\pi_i = \frac{1}{np}\sum_{k=i}^{n} p_k,\ i = 1,\ldots,n$ is a valid p.m.f. because $\sum_{i=1}^{n}\pi_i = 1$, hence:

$$\begin{aligned} C_{1\_bit} &= \sum_{i=1}^{n}\pi_i C_i \\ &= \sum_{i=1}^{n}\pi_i \int_0^\infty \log(1+\rho x)dF_i \\ &= \int_0^\infty \log(1+\rho x)d\left(\sum_{i=1}^{n}\pi_i F_i\right) \\ &= \int_0^\infty \log(1+\rho x)dF_\pi \end{aligned} \tag{4.12}$$

where $F_\pi = \sum_{i=1}^{n}\pi_i F_i$ is a mixture probability measure of all order statistics of the exponential family. Now we show that F_π satisfies the required condition for Theorem 5.

$$\mu_\pi = \sum_{i=1}^{n}\pi_i\mu_i \tag{4.13}$$

where $\mu_i = \int_0^\infty x dF_i(x)$ is the mean of the i^{th} order statistics of the exponential family, and

$$\begin{aligned} \sigma_\pi^2 &= \int_0^\infty (x-\mu_\pi)^2 dF_\pi(x) \\ &= \int_0^\infty x^2 dF_\pi(x) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i \int_0^\infty x^2 dF_i(x) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i(\sigma_i^2+\mu_i^2) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i\sigma_i^2 + \sum_{i=1}^{n}\pi_i\mu_i^2 - \mu_\pi^2 \end{aligned} \tag{4.14}$$

where $\sigma_i^2 = \int_0^\infty (x-\mu_i)^2 dF_i$ is the variance of the i^{th} order statistics of the exponential family. It is a known fact (e.g. [20] Section 4.6) that for exponential distribution $F(x) = 1 - e^{-x}$ we have

$$\mu_i = \sum_{j=i}^{n} \frac{1}{j} = H_n - H_{i-1}$$

where $H_k = \begin{cases} \sum_{j=1}^{k} \frac{1}{j} & k > 0 \\ 0 & k = 0 \end{cases}$ and also,

$$\sigma_i^2 = \sum_{j=i}^{n} \frac{1}{j^2} = S_n - S_{i-1}$$

where

$$S_k = \begin{cases} \sum_{j=1}^{k} \frac{1}{j^2} & k > 0 \\ 0 & k = 0 \end{cases}$$

We have to show that $\frac{\sigma_\pi}{\mu_\pi} \to 0$

$$\begin{aligned} \mu_\pi &= \sum_{i=1}^{n} \pi_i \mu_i = \sum_{i=1}^{n} \pi_i (H_n - H_{i-1}) \\ &= H_n - \sum_{i=1}^{n} \pi_i H_{i-1} < H_n = \mu_1 \end{aligned} \tag{4.15}$$

it is known [44] that for all $k \geq 1$,

$$\log k + \gamma + \frac{1}{2(k+1)} < H_k < \log k + \gamma + \frac{1}{2k} \tag{4.16}$$

using Jensen's inequality we have

$$\begin{aligned} \mu_\pi &= H_n - \sum_{i=1}^{n} \pi_i H_{i-1} \\ &> H_n - \sum_{i=1}^{n} \pi_i H_i \\ &> H_n - \gamma - \sum_{i=1}^{n} \pi_i \log i - \frac{1}{2} \sum_{i=1}^{n} \frac{\pi_i}{i} \\ &> H_n - \gamma - \log \left(\sum_{i=1}^{n} i \pi_i \right) - \frac{1}{2} \sum_{i=1}^{n} \pi_i \\ &> H_n - \gamma - \frac{1}{2} - \log \left(\sum_{i=1}^{n} i \pi_i \right) \end{aligned} \tag{4.17}$$

on the other hand

$$\begin{aligned}
\sum_{i=1}^{n} i\pi_i &= \frac{1}{np}\sum_{i=1}^{n} i \sum_{k=i}^{n} p_k \\
&= \frac{1}{np}\sum_{k=1}^{n} p_k \sum_{i=1}^{k} i \\
&= \frac{1}{np}\sum_{k=1}^{n} p_k \left(\frac{k(k+1)}{2}\right) \\
&= \frac{\sum_{k=1}^{n} k^2 p_k + \sum_{k=1}^{n} k p_k}{2np} \\
&= \frac{(n-1)p}{2} + 1 \qquad (4.18)
\end{aligned}$$

from (4.15), (4.17) and (4.18) we get:

$$H_n - \log\left(np + 2 - p\right) - \gamma - \log(2\sqrt{e}) < \mu_\pi < H_n \qquad (4.19)$$

by inspecting Equation (4.8) we also notice that $p_{opt} = O\left(\frac{1}{n}\right)$, because in order to have equality, the number of positive and negative terms in Equation (4.8) should be of the same order in the asymptote of large n. Equivalently, the optimal threshold α scales logarithmically in the asymptote of large n (this fact can also be seen in Figure 4.1 in which the X-axis is in logarithmic scale). Therefore, (4.19) suggests that

$$H_n - \mu_\pi = O(1) \qquad (4.20)$$

or,

$$\mu_\pi \mathrel{\mathring{=}} \mu_1 \mathrel{\mathring{=}} \log n \qquad (4.21)$$

as $n \to \infty$. On the other hand:

$$\begin{aligned}
\sigma_\pi^2 &= \sum_{i=1}^{n} \pi_i \sigma_i^2 + \sum_{i=1}^{n} \pi_i \mu_i^2 - \mu_\pi^2 \\
&< \sigma_1^2 + \mu_1^2 - \mu_\pi^2 \\
&= \sigma_1^2 + (\mu_1 + \mu_\pi)(\mu_1 - \mu_\pi) \\
&< \sigma_1^2 + 2\mu_1(\mu_1 - \mu_\pi)
\end{aligned} \qquad (4.22)$$

we also notice that $S_n < S_\infty = \frac{\pi^2}{6} < 2$ (here $\pi = 3.1416\ldots$) thus:

$$\sigma_\pi^2 < 2 + 2\mu_1(\mu_1 - \mu_\pi) \tag{4.23}$$

hence from (4.20), (4.21) and (4.23) we have:

$$0 \le \left(\frac{\sigma_\pi}{\mu_\pi}\right)^2 < \frac{2}{\mu_\pi^2} + 2\left(\frac{\mu_1}{\mu_\pi}\right)\left(\frac{\mu_1}{\mu_\pi} - 1\right) \to 0 \tag{4.24}$$

as $n \to \infty$. Thus Equation (4.21) implies $\frac{\sigma_\pi}{\mu_\pi} \to 0$ as $n \to \infty$. Now we can apply Theorem 10 and Equation (4.21) to show that:

$$\begin{aligned} C_{1\_bit} &= \int_0^\infty \log(1 + \rho x) dF_\pi \\ &\overset{\circ}{=} \log(1 + \rho\mu_\pi) \\ &\overset{\circ}{=} \log(\log n) + \log\rho \end{aligned} \tag{4.25}$$

■

4.6 Simulation Results

Figure 4.2 shows the sum-rate capacity of a SISO network. As it can be seen in the figure, our proposed scheduling, with only 1-bit feedback, has the same double logarithmic growth rate as the fully informed network. The capacity loss is minimal. Scheduling with 1-bit feedback also captures most of the capacity of the fully informed network for a wide range of SNR, thus the scaling law proved in Theorem 5 is verified by the simulation. Figure 4.1 shows the optimal threshold for various of SNR values. It can be seen that the optimal threshold scales logarithmically with number of users (in Figure 4.1 the x-axis is in logarithmic scale).

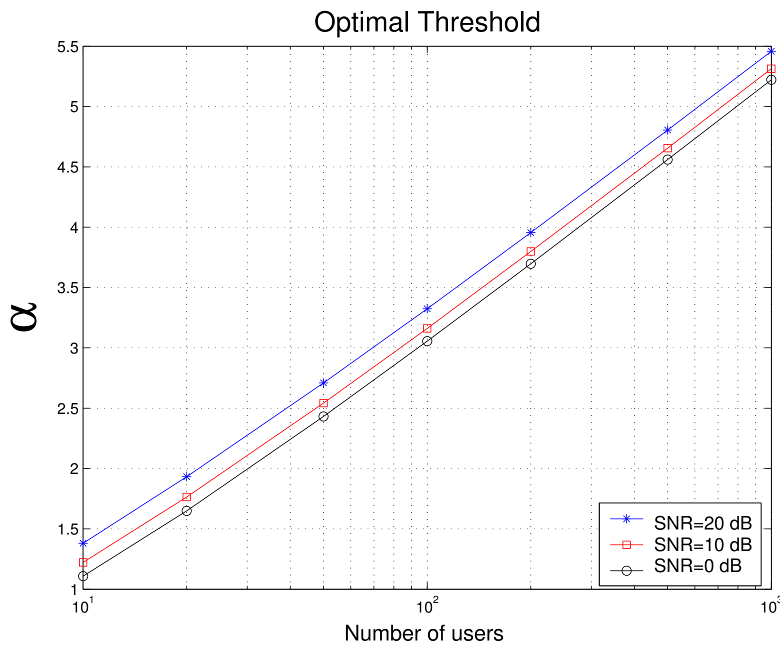

Figure 4.1. Optimal threshold vs. number of users for different SNR values

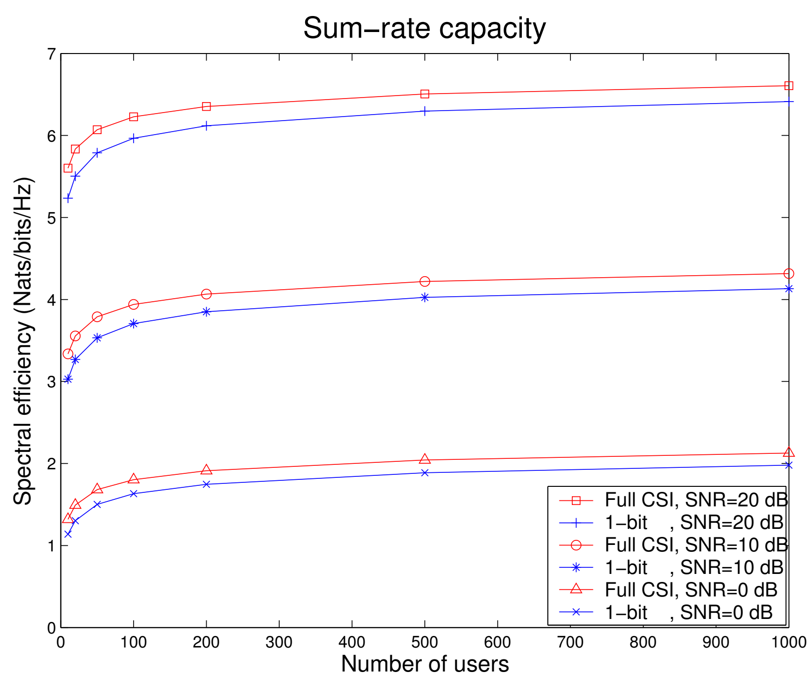

Figure 4.2. Comparison of sum-rate capacity for 1-bit and full CSI scheduling for different values of SNR

CHAPTER 5

OPPORTUNISTIC BEAMFORMING WITH LIMITED FEEDBACK

In the downlink of a multiuser MIMO system, if the base-station perfectly knows the channel states of all the users, then the so-called *dirty paper coding* is known to be the best strategy for maximizing the sum-rate capacity [45]. In a frequency division duplex (FDD) system the channel states are only available at the receiver, therefore in order to perform dirty paper coding, all users must feedback their complex-valued channel state matrices to the base-station, and this can impose a huge overhead on the system.

Sub-optimal schemes that exploit partial channel knowledge for downlink transmission are interesting from a practical point of view. The scheduling mechanism employed in downlink transmission can vary from *time-sharing* (transmitting to only one user at a time, e.g., [14, 41]) to *multicasting* , i.e., transmitting to several users at the same time [15]. It has been shown [46] that the sum-rate capacity of time-sharing scheduling is sub-optimal. However, these schemes are still of practical interest because of the simplicity of implementation.

Perhaps one of the more ingenious ways of using CSI is opportunistic beamforming [14] of Viswanath et al. This method applies to a multiuser system, where the users estimate and feedback their instantaneous downlink SNR to the base station. Then the base station will transmit to *the best* user. Because the channels between users and the base station is variable, loosely speaking, each user receives data when their channel *peaks*. In order for users to have reasonable delay, they must not be made to wait too long. To improve the delay profile [14] proposes an i.i.d. random phase on one of the two transmit antennas, thus creating an artificial fast fading.

There is also another way of using channel state information in a multi-antenna transmitter, namely (deterministic) beamforming. For example, one may use a conventional round robin scheduling between users, and when each user's turn comes, transmit along the eigen-direction of the channel to that user (either with adaptive power control or with adaptive rate control). It has been suggested [13] (Exercises) that in certain scenarios, in terms of overall network throughput, deterministic beamforming is inferior to opportunistic (random) beamforming.

Each of these methods nominally require unlimited reliable feedback. In Chapter 4 (also in [16]) we addressed the question of opportunistic scheduling with limited feedback, showing that only one bit of feedback per user is sufficient to capture most of the gain of multi-user diversity. Also, there exists a good amount of work on quantized beamforming (a nice summary of work in quantized beamforming is presented in [47]).

In this chapter we show how multiuser diversity and transmit diversity can co-exits together *in a practical scenario of limited feedback.* In other words, we ask the following question: in the presence of limited feedback, what combination of the two methods (opportunistic multi-user vs. deterministic beamforming) should we use, how can this combination be accomplished, and how well does it perform. At this point we should also address the notion of channel hardening with increasing number of transmit antennas [41]. First, the channel hardens only in the asymptote of large number of antennas. Second, it has recently been shown [48] that channel hardening due to multiple antennas is not always detrimental to multi-user diversity, which is also confirmed by our research.

5.1 System Model

We consider a network of n users each having N antenna for receiving data from the base-station. The base station has M antennas. For k^{th} user we assume the linear time invariant flat fading model:

$$\mathbf{y}_k(t) = \mathbf{H}_k \cdot \mathbf{x}_k(t) + \mathbf{n}_k(t)$$

where $\mathbf{y}_k(t) \in \mathbb{C}^N$ is the received signal and $x_k(t) \in \mathbb{C}^{M\times 1}$ is the transmitted signal for user k at time t. The transmit power is limited by ρ, i.e. $\mathbb{E}[\|\mathbf{x}\|^2] \leq \rho$, $\mathbf{n}_k(t)$ is a i.i.d. circularly symmetric complex Gaussian noise distributed according to $\mathbf{n}_k(t) \sim \mathcal{CN}(0, I_N)$ and $\mathbf{H}_k$ is a an $N \times M$ channel matrix whose i^{th} element, $\mathbf{H}_{ij,k}$ represents the channel gain between i^{th} transmit antenna at the base-station and the j^{th} receive antenna of the k^{th} user.

We assume the antennas at the base-station are not correlated, also we assume channel matrices across different users are independent. We assume that for k^{th} user the channel matrix $\mathbf{H}_k$ is perfectly known at the receiver but not necessarily known at the base-station and for each user there exits a feedback channel with limited rate to securely convey the channel state information to the base-station.

5.2 Scheduling with Limited Feedback of CSI

In Chapter 4, for the case where the base station has one antenna ($M = N = 1$), we propose an algorithm for downlink scheduling that only requires one bit of feedback from each user to the base-station. Furthermore, we proved that subject to judicious choice of the threshold, the above scheduling algorithm leads to the same sum-rate capacity growth as that of a scheduling with full knowledge of CSI at the base-station.

In a slow-fading environment, scheduling with full CSI (always choosing the user with the maximum channel gain) leads to excessive delays. The 1-bit scheduling scheme has a much better delay and fairness profile simply because in each time interval the base-station can choose from a pool of *eligible users* (those above the threshold) thus the base-station can prioritize the eligible users based on their previous utilization and on average this leads to better fairness.

Figure 4.2 compares the sum-rate capacity of the 1-bit scheduling and full CSI scheduling schemes for different values of SNR. The closeness of the curves in Figure 4.2 suggests that when the available rate in the feedback channel is more than one bit, using the extra feedback bits for quantizing the channel gains does not lead to a significant capacity improvement. *Therefore it is reasonable to use any extra feedback, over and above one bit, for other purposes.* Aside from multiuser diversity, there are other ways of using transmit-side channel state information, perhaps the most obvious being beamforming (array gain). We propose that any excess channel state information, over and above one bit, can be used to exploit beamforming gain.

There exists prior work on beamforming with limited feedback of channel knowledge. Reference [47] provides a comprehensive survey of feedback methods. When there are multiple antennas at the base-station, we suggest to combine our 1-bit opportunistic scheduling [16] and beamforming with limited feedback [49, 50]. We suggest that a combination of our 1-bit method and the limited feedback beamforming enjoys both the multiuser diversity and the transmit diversity and it has better performance over opportunistic (random) beamforming, assuming the same rate feedback is available for both methods.

When there are multiple antennas at the base-station and the rate of the feedback channel is limited to be L bits per channel realization, we can use the limited feedback methods for exploiting the beamforming gain [47]. A beamforming codebook $\mathcal{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_{2^L}\}$ ($\mathbf{u}_1, \cdots, \mathbf{u}_{2^L}$ are the beamforming vectors) is shared by all users and the bate-station. The k^{th} user finds the beamformer $\widehat{\mathbf{u}}_k$ that leads to the highest gain, i.e.

$$\widehat{\mathbf{u}}_k = \arg\max_{\mathbf{u}\in\mathcal{U}} \|\mathbf{H}_k u\|^2$$

then it compares the corresponding channel gain $\eta_k = \|\mathbf{H}_k \widehat{\mathbf{u}}_k\|^2$ to the threshold value

α advertised by the base-station. If the channel gain was above the threshold, The user sends its L-bit feedback information to the base-station, otherwise it does not transmit any feedback information. Thus the reception of L bits from user k by the base-station indicates that

1. The user k is *eligible* for transmission
2. The base-station should use the beamforming vector $\widehat{\mathbf{u}}_k \in \mathcal{U}$ for transmission to user k.

For scheduling, the base-station randomly selects one of the eligible users. When there is no eligible user in the network, it does not transmit to any user. [1]

5.3 Opportunistic Transmit Antenna Selection

When the beamforming vector is of size M (hence $L = \log_2 M$) then the best choice the code-book is to take u_l's as columns of identity matrix M.

$$\mathcal{U} = \left\{ \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix} \right\}$$

In this case, beamforming is equivalent to antenna selection in the base-station (since only one antenna at a time is active). Let $\mathbf{h}_{i,k}$ denote the i^{th} column of the channel matrix $\mathbf{H}_k$. The user k finds the column of its channel matrix with the maximum norm

$$\widehat{i_k} = \arg \max_{1 \leq i \leq M} \|\mathbf{h}_{i,k}\|^2$$

and then compares $\|\mathbf{h}_{\widehat{i_k},k}\|^2$ with the threshold value α advertised by the base station. If $\|\mathbf{h}_{\widehat{i_k},k}\|^2 > \alpha$, then the feedback bits $b_1 b_2 \cdots b_M$ (the binary digits of $\hat{i}$) are transmitted to

[1] The scheduling to users with favorable channels may also be implemented via round robin. In long run, both these strategies have the same average throughput per user. However, the round-robin version may be more appealing from a fairness point of view.

the base station, otherwise, no feedback is sent. Note that $\widehat{i_k}$ indicated the best transmit antenna for downlink transmission to the user k. Upon receipt of this information from eligible users, the base station randomly chooses one user for transmission from among all users whose feedback information has been successfully received. If the feedback bits for selected user, are $b_1 b_2 \cdots b_M$, the base station uses its $\hat{i}^{\ th}$ antenna for transmission, where $\hat{i} = \overline{b_1 b_2 \cdots b_M}$.

5.3.1 Sum-Rate Capacity

The equivalent channel gain for the user j is

$$\eta_j = \max_{1 \leq i \leq M} \|\mathbf{h}_{i,j}\|^2$$

Under Rayleigh fading assumption, each channel gain $\|\mathbf{h}_{i,k}\|^2$ is χ^2 distributed with N degrees of freedom, thus the CDF of η_j is

$$F(x) \ = \ (1 - e_N(x)e^{-x})^M \ , \tag{5.1}$$

where $e_N(x) = \sum_{p=0}^{N-1} \frac{x^p}{p!}$. Let $p = \Pr[\eta_k > \alpha]$. Since the channel gains are all mutually independent, the probability of having k feedback bits equal to one obeys a binomial law, i.e.

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.2}$$

The ergodic capacity upon having k user above the threshold is:

$$\begin{aligned} \overline{C}_k \ &= \ \sum_{i=1}^{k} Pr[\text{the } i^{th} \text{ best user is selected}] \ C_i \\ &= \ \frac{1}{k} \sum_{i=1}^{k} C_i \end{aligned} \tag{5.3}$$

where $C_i = \int_0^\infty \log(1 + \rho x) dF_i(x)$ and $F_i(x)$ is the CDF of the i^{th} highest equivalent channel gain. In other words if $\{X_1, \ldots, X_n\}$ is a permutation of $\{\eta_1, \ldots, \eta_n\}$ such that

$0 \leq X_n \leq \ldots \leq X_1$, then $F_i(x) = Pr[X_i < x]$. When the channel gains are i.i.d., it can be shown that [20]:

$$F_i(x) = \sum_{l=0}^{i-1} \binom{n}{l} (F(x))^{n-l} (1 - F(x))^l \tag{5.4}$$

Thus the sum-rate capacity of the network with limited feedback can be formulated as:

$$C_{LF} = \sum_{k=1}^{n} p_k \overline{C}_k \tag{5.5}$$

According to Equation (4.12), C_{LF} can be explained as follows

$$C_{LF} = \int_0^\infty \log(1 + \rho x) dF_\pi \tag{5.6}$$

where $F_\pi = \sum_{i=1}^n \pi_i F_i$ is a mixture probability measure of all order statistics of the family with parent CDF $F(.)$ and $\{\pi_i\}_{i=0}^n$ is a discrete probability measure defines as

$$\pi_i = \frac{1}{np} \sum_{k=i}^{n} p_k, \quad i = 1, \ldots, K \tag{5.7}$$

5.3.2 Optimal Threshold

The sum-rate capacity is a function of ρ, p and n. As explained in Section 4.6, the relation between the threshold α and the probability p is given by

$$\alpha = F^{-1}(1 - p) \tag{5.8}$$

The inverse function of the CDF given by (5.1) in general can not be explicitly calculated. When users have only one receive antenna ($N = 1$), however,

$$\alpha = -\log\left(1 - (1 - p)^{\frac{1}{M}}\right) \tag{5.9}$$

In order to find the optimal threshold we choose p such that the sum-rate capacity C_{LF} is maximized. The cost function $C_{LF}(p)$ is a weighted sum of functions of the form

$p^k(1-p)^{n-k}$ which are all concave over the interval $[0,1]$, hence C_{LF} is a concave function of p. Therefore it has a unique maximum over the interval $[0,1]$. To calculate the value of p that maximizes the sum-rate capacity, we must solve $\frac{\partial C_{LF}(p)}{\partial p} = 0$ for p, i.e.,

$$\sum_{k=1}^{n}(k-np)p_k\overline{C}_k = 0 \tag{5.10}$$

A closed form solution to this equation is in general not tractable. However, a numerical solution is possible with $O(n)$ complexity.

5.3.3 Asymptotic Analysis

In this section we show the proposed opportunistic scheme has the same capacity growth of *coherent opportunistic beamforming with full CSI* where the base-station has full channel knowledge of all users and the users with the highest channel norm is selected for transmission. The sum-rate capacity of this scheme provides an upper bound on the all opportunistic beamforming methods that user partial knowledge of the channel. For the simplicity of analysis we first assume that users have only one antenna ($N = 1$). For this scheduling scheme, the sum-rate capacity is obtain as follows:

$$C_{Full\_CSI} = \mathbb{E}[\log(1+\rho \max_{1\leq k\leq n} \|\mathbf{h}_k\|^2)] \tag{5.11}$$

where $\mathbf{h}_k$ is the $1 \times M$ channel vector of user k.

Theorem 6 *The sum-rate capacity of coherent opportunistic beamforming with full CSI available at the base-station scales as*

$$C_{Full\_CSI} \overset{\circ}{=} \log\log n + \log\rho$$

Proof: The random variable $Y_k = \|\mathbf{h}_k\|^2$ is distributed according to χ^2_{2M}. Using classical results in extreme value theory, it is shown in [8, 9] that the mean and the variance of

$Z_K = \max_{1\le k\le K} Y_k$ have the following asymptotic behavior

$$\mu_n = \mathbb{E}[Z_n] \overset{\circ}{=} \log n + \log\left(\frac{n^{M-1}}{(M-1)!}\right) + \gamma \tag{5.12}$$

$$\sigma_n^2 = \mathbb{E}\left[(Z_n - \mathbb{E}[Z_n])^2\right] \overset{\circ}{=} \frac{\pi^2}{6} \tag{5.13}$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Therefore $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$ as $n \to \infty$ thus Z_k satisfies the condition in Theorem 10 and we have

$$\begin{aligned} C_{Full\_CSI} &= \mathbb{E}[\log(1+\rho Z_n)] \\ &\overset{\circ}{=} \log(1+\rho\mathbb{E}[Z_n]) \\ &\overset{\circ}{=} \log\left(1+\rho(\log n + \log\left(\frac{n^{M-1}}{(M-1)!}\right)+\gamma)\right) \\ &\overset{\circ}{=} \log\log n + \log\rho. \end{aligned} \tag{5.14}$$

The next theorem explains that the sum-rate capacity for our proposed scheme (Equation 5.6) also has the same capacity growth as the scheduling with full CSI.

Theorem 7 *The sum-rate capacity of opportunistic transmit antenna selection scales the same as scheduling with full CSI (coherent beamforming), i.e.*

$$\lim_{n\to\infty} \frac{C_{LF}}{C_{Full\_CSI}} = 1$$

Proof: From Equation (5.6) we have $C_{LF} = \int_0^\infty \log(1+\rho x)dF_\pi$ where

$$F_\pi = \sum_{k=1}^{n} \pi_n F_k$$

and F_k's are the probability measures associated with order statistics with the parent distribution $F(x) = (1-e^{-x})^M$. if X is a random variable distributed according to $F(\cdot)$, we define the function $g(\cdot)$ such that $X = g(Y)$ where Y is an exponential random variable. We have

$$F(x) = \Pr[X < x] = \Pr[g(Y) < x] = \Pr[Y < g^{-1}(x)] = 1 - e^{-g^{-1}(x)}$$

thus

$$y = g^{-1}(x) = -\log(1 - F(x)) \tag{5.15}$$

We have $y' = \frac{f(x)}{1-F(x)} > 0$ thus $g^{-1}(x)$ and hence $g(x)$ are strictly increasing function which means they preserve order. Therefore if $0 < X_n \leq \cdots \leq X_1$ are the order statistics with parent distribution, $F(\cdot)$, $0 < Y_n \leq \cdots \leq Y_1$ with $Y_k = g^{-1}(X_k)$ are the order statistics of the exponential distribution.

$$C_{LF} = \sum_{k=1}^{n} \pi_k \mathbb{E}[\log(1 + \rho X_k)] = \sum_{k=1}^{n} \pi_k \mathbb{E}[\log(1 + \rho g(Y_k)]$$

This can be written as

$$C_{LF} = \int_0^\infty \log(1 + \rho \cdot g(x))\, d\nu_\pi(x)$$

where $\nu_\pi = \sum_{k=1}^{n} \nu_k$ is the mixture probability distribution of all order statistics of the exponential distribution. $\nu_k(x)$ is the CDF of Y_k and can be calculated as follows

$$\nu_k(x) = \sum_{l=0}^{k-1} \binom{n}{l} e^{-lx}(1 - e^{-x})^{n-l}$$

In the proof of Theorem 5, we showed that $\frac{\mu_\pi(\nu)}{\mu_1(\nu)}$ and $\frac{\sigma_\pi(\nu)}{\mu_\pi(\nu)} \to 0$, as $n \to \infty$ where

$$\mu_1(\nu) = \int_0^\infty x d\nu_1(x) = \mathbb{E}[Y_1] = \sum_{k=1}^{n} \frac{1}{k} \quad , \quad \mu_\pi(\nu) = \int_0^\infty x d\nu_\pi(x) = \sum_{k=1}^{n} \pi_k \mathbb{E}[Y_k]$$

and

$$\sigma_1^2(\nu) = \int_0^\infty (x - \mu_1(\nu))^2 d\nu_1(x) = \sum_{k=1}^{n} \frac{1}{k^2} \quad , \quad \sigma_\pi^2(\nu) = \int_0^\infty (x - \mu_\pi(\nu))^2 d\nu_\pi(x)$$

Now we show that the function $g(\cdot)$ belongs to $\mathcal{G}$ (defined in the appendix A, Definition 3), thus we can apply Theorem 13 to show that C_{LF} scales the same as C_{Full_CSI}. Recall $g(\cdot)$ is strictly increasing and $g^{-1}(0) = -\log(1 - F(0)) = 0$ hence $g(0) = 0$. To show that $g(\cdot)$ is concave, we show $y = g^{-1}(x) = -\log(1 - (1 - e^{-x})^M)$ is convex, we have $1 - e^{-y} = (1 - e^{-x})^M$ hence

$$y' e^{-y} = M e^{-x}(1 - e^{-x})^{M-1}$$

$$y''e^{-y} - (y')^2 e^{-y} = Me^{-x}(1-e^{-x})^{M-2}(Me^{-x}-1)$$

after some algebra we get

$$\begin{aligned}
\frac{y''}{y'}e^{-y}(1-e^{-x}) &= Me^{-x} - (1-(1-e^{-x})^M) \\
&= Me^{-x} - e^{-x}\sum_{i=0}^{M-1}(1-e^{-x})^i \\
&= e^{-x}\sum_{i=0}^{M-1}(1-(1-e^{-x})^i) > 0
\end{aligned}$$

The latter inequality means that $y'' > 0$ for all $x > 0$, hence g^{-1} is convex, therefore $g \in \mathcal{G}$. Further more, we prove $\lim_{x\to\infty}\frac{g^{-1}(x)}{x} = 1$,

$$\begin{aligned}
\lim_{x\to\infty}\frac{g^{-1}(x)}{x} &= \lim_{x\to\infty} -\frac{\log(1-(1-e^{-x})^M)}{x} \\
&= \lim_{x\to\infty} -\frac{\log\left(1-(1-Me^{-x}+\binom{M}{2}e^{-2x}+\cdots+(-1)^M e^{-Mx})\right)}{x} \\
&= \lim_{x\to\infty} 1 + \frac{\log\left(M-\binom{M}{2}e^{-x}+\cdots+(-1)^M e^{-(M-1)x}\right)}{x} \\
&= 1
\end{aligned}$$

We also have $\lim_{x\to\infty} g^{-1}(x) = \lim_{x\to\infty} g(x) = +\infty$, let $y = g(x)$

$$\lim_{x\to\infty}\frac{g(x)}{x} = \lim_{y\to\infty}\frac{y}{g^{-1}(y)} = 1$$

therefore $g(\cdot)$ satisfies all the conditions of Theorem 13, hence

$$\begin{aligned}
C_{LF} &= \int_0^\infty \log(1+\rho x)dF_\pi(x) \\
&= \int_0^\infty \log(1+\rho g(x))d\nu_\pi(x) \\
&\mathrel{\overset{\circ}{=}} \log(1+\rho g(\mu_\pi(\nu))) \\
&\mathrel{\overset{\circ}{=}} \log(1+\rho\mu_\pi(\nu)) \\
&\mathrel{\overset{\circ}{=}} \log(1+\rho\mu_1(\nu)) \\
&\mathrel{\overset{\circ}{=}} \log(1+\rho\log n) \\
&\mathrel{\overset{\circ}{=}} \log\log n + \log\rho
\end{aligned}$$

Thus from Theorem 6 we conclude

$$C_{LF} \overset{\circ}{=} C_{full\_CSI}$$

∎

5.4 Quantized Opportunistic (Random) Beamforming

Recall that the main premise of our proposed method is a balance/tradeoff between two main methods of using transmit CSI, in the limited feedback regime. One of them is deterministic beamforming, and the other is the opportunistic beamforming of Viswanath et al [14]. There exists some work on the quantization of feedback for deterministic beamforming, but to our knowledge, little work exists in the public-domain literature on the quantization effects in opportunistic beamforming. As a baseline for comparing our algorithms, in this section we develop a limited-feedback (quantized) version of the opportunistic beamforming.

In opportunistic beamforming, each user sends back to the base station its instantaneous receive SNR. For the quantized opportunistic beamforming, only a discrete index can be transmitted. The calculation of optimal quantization for this purpose is intractable due to the complicated error function involved, however, precise quantization is not critical because the boundaries of the quantization bins are important only to the extent that they give the same ordering of the users as given by optimal quantization. Therefore a reasonable but approximate quantization performs well with high probability.

The approximate quantization that we developed for our simulations and comparison purposes is chosen so that the quantization index takes all possible values with equal probability. Denote the CDF of SNR as $F_\gamma(x)$. We consider L quantization bits,

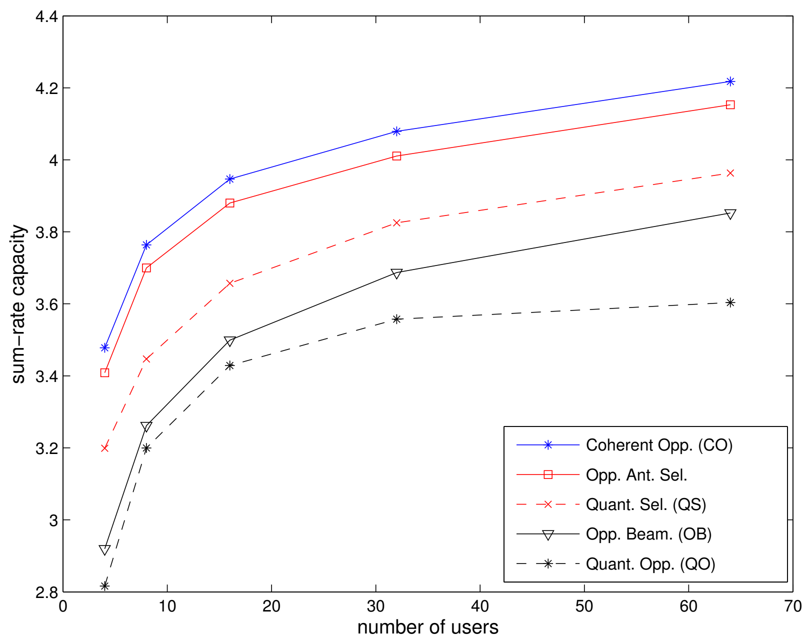

Figure 5.1. Comparison of the throughput of various methods at SNR=10dB

thus 2^L quantization levels. The quantization bins are defined by

$$F_\gamma\left(\frac{\ell}{2^L}\right) \leq \gamma < F_\gamma\left(\frac{\ell+1}{2^L}\right) \qquad \ell = 0, \ldots, 2^L - 1$$

5.5 Comparisons

5.5.1 Throughput

Figure 5.1 shows a comparison of the throughput of various methods considered in this chapter. The best performance is that of coherent opportunistic beamforming (CO), where the multiple-antenna transmitter has full knowledge of the (vector) channel of all users, and in each time interval beamforms towards the best user. Although this strategy has the best overall throughput, it suffers in two ways: first, it requires the highest amount of feedback compared with other methods, especially compared with the proposed limited-feedback methods. Second, in terms of fairness it is one of the worst techniques, as shown in the next subsection.

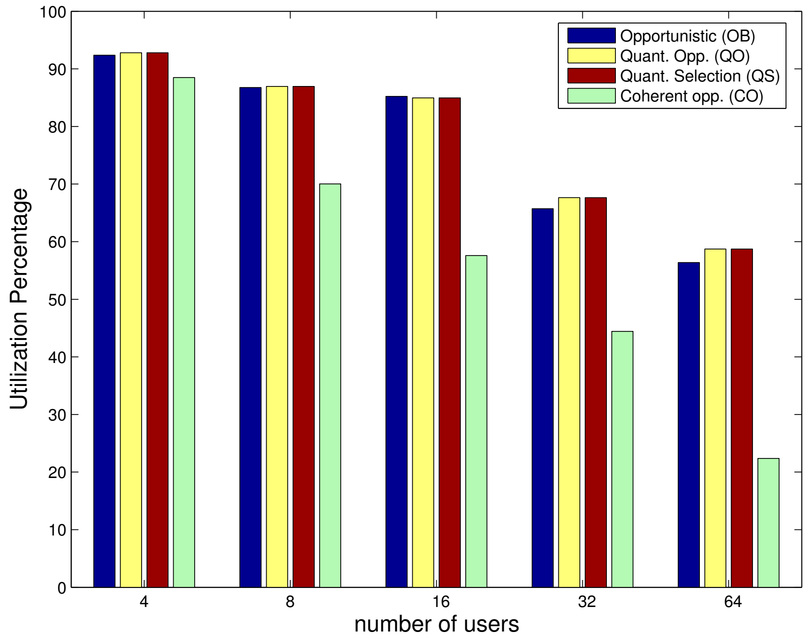

Figure 5.2. Normalized fairness

With a slight loss of throughput, one may use opportunistic antenna selection, where everything is like the previous case, except instead of full beamforming, the transmitter selects one of the antennas. In other words, with the knowledge of user (vector) channels, the transmitter picks the best transmit antenna and user among all possible such pairs.

Opportunistic beamforming (OB) as suggested by Viswanath et al [14] has performance below that of opportunistic antenna selection.

Finally, each of the last two cases can be quantized, as explained elsewhere, each resulting in some loss of throughput. The quantized versions shown in Figure 5.1 are done with only one bit per user, that selects the best antenna for that user. We note that using this one bit as antenna selection is only due to the fact that we have used two transmit antennas. The more general case is explained below.

5.5.2 Fairness

The simplest form of scheduling in a multi-user system is a deterministic round robin, which in fact has the best delay profile: every user is guaranteed a slot within a finite time. When there is flexibility on delay, for example in data communication, other modes of scheduling can be used that result in higher throughput, for example opportunistic beamforming. Even in the case of a winner-take-all strategy, however, one does not want to lose sight of the delay of users who do not "win" the competition for the channel, and this has been addressed for example in [14]. In this section, we define a metric for fairness and compare the methods of interest (in the limited feedback regime) according to this metric.

Assume that we apply opportunistic beamforming N_s times to a group of users. In traditional opportunistic beamforming, when the exact SNR of all users is known at the base station, only one user (with the highest SNR) is eligible for transmission. In the limited-feedback versions, more than one user may be eligible for reception, among eligible users one may choose randomly or according to other algorithms (e.g., whoever has been waiting the longest, etc.)

For our metric, in each realization of the channel, we note the ineligible users. In a sense these are our *underprivileged* users. Since the decision of privileged/unprivileged is a binary decision, we can model it with a 0-1 Bernoulli random variable. After N_s transmissions, we look at the most under-privileged user (the one who was eligible the least number of times) and calculate the expected value of the number of times it has been eligible.[2] This is a measure of the utilization of the worst user. We then normalize the utilization factor above by the utilization of round robin scheduling, to remove the effects of N_s and the number of users. We show a comparison of the fairness of various

[2]This is an order statistic which is difficult to calculate in closed form, because the random variables corresponding to different users are dependent, but it can be calculated numerically.

methods in Figure 5.2.

The four methods compared here are opportunistic beamforming with unlimited feedback (OB), quantized opportunistic beamforming (QO), our limited feedback scheme (QS), and coherent opportunistic beamforming (CO) where the base station knows all the (vector) channel gains of all users, and at each time interval deterministically beamforms towards the best user.

As observed in Figure 5.2, in terms of fairness alone, OB and its quantized version QO perform quite well, as well as our method QS. Deterministic beamforming does not do as well, especially when the number of users is high. The small variations between OB and QO, we believe, are due to the finite window effect and possibly also to the sample size of the simulation.

CHAPTER 6
OPPORTUNISTIC METHODS IN MULTIUSER BROADBAND NETWORKS

There is ever-increasing demand for higher data rates in the next generation wireless systems. In a multiuser environment, where a large number of users share the same media for communication, efficient broadband transmission techniques that provide higher spectral efficiency have been the subject of intense research in the past few years. Orthogonal frequency devision multiplexing (OFDM) is one of the well known multi-carrier techniques to combat ISI and is now an integral part of wireless standards such as 802.11a and HIPERLAN/2.

In a network of wireless users, multi-user diversity can be exploited to provide higher spectral efficiency and quality of service [12]. One way to exploit multiuser diversity gain is through opportunistic scheduling [12][14]. Previous work on opportunistic scheduling has been more focused on the frequency flat fading model [12][14][15]. However, in a network of OFDM users, only a few works have utilized opportunistic schemes to increase the capacity of the system.

One of the major problems in employing an opportunistic scheme in OFDM network is the large amount of feedback required to pass to the base-station. For example in 802.11a each user has 64 subchannels and a network of 100 users requires the base station to collect 6400 real numbers from all the users. Furthermore, this information should be received error free and with no delay. To address this issue, [51] proposed an opportunistic scheme in which adjacent subchannels are clustered into groups and then only the average SNR value of each cluster is fed back to the base station. But this still requires feeding back several real numbers to the base-station which may not be affordable, especially in ultra wide-band and/or fast-fading scenarios.

In this chapter we propose a simple subchannel allocation scheme in which only one bit per subchannel (or cluster) is fed back to the base station. Thus when each user has N subchannels, only N feedback bits per user are required. We show that even this limited feedback can increase the capacity to more than twice the capacity obtained by TDMA. Using analytical techniques developed in Chapters 2 and 4, we show that the growth rate of network capacity is identical to optimal opportunistic subchannel allocation with full channel information at the base station. Furthermore we investigate the effect of correlation among channel taps to show that even when there is correlation, significant gain in terms of sum-rate capacity can be achieved by our proposed scheme.

The organization of the chapter is as follows, in Section 6.1 we introduce the system model. Section 6.2 talks about the sum-rate capacity and its formulation in OFDM networks. In Section 6.3 our proposed limited feedback dynamic subchannel allocation algorithm is discussed. Section 6.4 contains the results about the sum-rate capacity growth in the asymptote of large number of users. In Section 6.5 we investigate the performance of our algorithm when there is temporal correlation between each user's channel taps.

6.1 System Model

For each user in the network, we consider a frequency selective linear time invariant model

$$y_{t,k} = \sum_{i=0}^{\nu} h_{i,k} x_{t-i,k} + w_{t,k} \tag{6.1}$$

where $x_{t,k}$ and $y_{t,k}$ are the input and the output for the k^{th} user ($k \in \{1, \ldots, K\}$) at time t respectively, w is the additive white complex Gaussian noise and uncorrelated among the users with zero mean and variance σ_w^2, $h_{i,k}$ is the i^{th} channel tap for user k and is distributed as $\mathcal{CN}(0,1)$ which is assumed to be uncorrelated among different users, although for each user, channel taps may or may not be correlated. ν is the memory of the channel and it is assumed to be the same for all users. We assume that the base-station uses OFDM for data transmission to each user. By applying cyclic prefix and IDFT, user k's channel is divided into N different sub-channels $H_{n,k}$ such that:

$$H_{n,k} = \frac{1}{\sqrt{N}} \sum_{t=0}^{\nu} h_{t,k} e^{-j\frac{2\pi nt}{N}} \tag{6.2}$$

We also assume that the total transmission power in the network is limited by $P_{\max}$.

6.2 Sum-rate Capacity

When all users share the same bandwidth, and the base-station has full information about every user's subchannels, then in order to maximize the sum-rate capacity of the network, the problem of subcarrier and power allocation to different users in the network must be solved jointly. However, this imposes a huge computational complexity at the base station. Especially if the wireless channel varies quickly, then then base-station requires an enormous computational power to rapidly compute the optimal solution for power and subchannel allocation among the users. Moreover the optimal dynamic joint power and subchannel allocation requires fast and reliable feed-forward and feedback channels for exchanging information between the users and the base station. Especially with large number of users in the network, sending this information back and forth between the users and the base station causes a huge overhead for the network. This motivates a low-complexity sub-optimum algorithm.

One may achieve economy of computation and communication through separation of subchannel and power allocation. It is possible to first select subchannels and then perform water-filling among all selected subchannels, but this again requires the base station to send back the optimal power allocation vector to all the users, together with the indices of their selected subchannels. Yet another suboptimal scheme is to equally allocated the total power among all subchannels and then perform the subchannel allocation among all users [52]. We adopt the latter approach in this paper.

Assuming full channel knowledge at the base station, maximizing the sum-rate capacity of the network reduces to allocating subchannel to users that have the best channel conditions. In order to avoid inter-carrier interference we allocate each frequency bin to a single user. Under this condition, maximum sum-rate capacity with equal power splitting among the subchannels is achieved when for each frequency bin we choose the user whose corresponding subchannel gain is maximum within that frequency band. The sum rate capacity in this case is given by

$$C_{full\_CSI} = \sum_{n=1}^{N} \log(1 + \mathrm{SNR} \cdot \max_{k} |H_{n,k}|^2) \tag{6.3}$$

where $\mathrm{SNR} = \frac{P_{\max}}{N\sigma_w^2}$ is the SNR per subchannel. This subchannel selection scheme is in fact a generalization of the opportunistic scheduling in flat-fading multiuser networks [12] over N different flat fading subchannels provided by OFDM.

6.3 Subchannel Allocation with Limited Feedback

The opportunistic scheme mentioned in the previous section and most of the similar subchannel allocation schemes [52] require full knowledge of the subchannel information to be available at the transmitter. However, from a practical point of view this is not affordable because a sum total of KN positive real numbers should be reliably transmitted to the base-station which is not affordable in practice . Svedman et. al [51] propose an alternative where, instead of feeding back the gain of each subchannel, each user's subchannels are divided into clusters and in each cluster the maximum value of the cluster is fed back to the base-station. This reduces the number of real values to KL assuming that there are L clusters. But this still requires feeding back several real numbers to the base station without any error and delay which is still not attractive from an implementation point of view.

We propose a simple scheme where, instead of feeding back the full information of the subchannels, only one-bit of information per subchannel (or cluster) is fed back to the base station for subchannel allocation. For user k, the n^{th} subchannel gain $|H_{n,k}|^2$ is compared to a threshold α_n, if the subchannel gain is above the threshold a *"1"* is transmitted back to the base station otherwise a *"0"* is transmitted. [1] So only N bits per user is required in feedback.[2] Upon receipt of all feedback bits from the users, the bases station allocates each subchannel to one of the users whose corresponding feedback bit is *"1"*. This assignment can be done via random selection or round robin scheduling among eligible users. Our claim is that by judicious choice of the threshold levels $\{\alpha_n\}$ most of the multiuser capacity gain is preserved. The choice of α_n's for each sub-channel can be done according to Section 4.4.

6.4 Capacity Scaling

When channel taps are uncorrelated, i.e. $\mathbb{E}[h_{t,k}h^*_{s,k}] = \delta_{t-s}$, we can use the analytical framework developed in Chapter 4 for determining the optimal threshold value and the evaluation of the sum-rate capacity. We notice that under the assumption of uncorrelated channel taps, the subchannel gains $\{|H_{n,k}|^2\}$ are i.i.d. random variables with exponential distribution. Hence the ergodic sum-rate capacity with full CSI at the base station is:

$$C_{full\_CSI} = N\mathbb{E}[\log(1 + \text{SNR} \max_k |H_{n,k}|^2)] \tag{6.4}$$

when subchannels are uncorrelated, similar to Theorem 5 we can show that

Theorem 8 *The sum-rate capacity of scheduling with one-bit per sub-channel scales the same as the capacity of opportunistic scheduling with full CSI.*

---

[1] When the feedback channel is contention-based and shared among all users, then in case a sub-channel is below the threshold level, no feedback data for that subchanel is transmitted, this leads to a better utilization of the available bandwidth on the feedback channel.

[2] One can also use the idea of clustering the subchannels to reduce the amount of feedback to L bits per user.

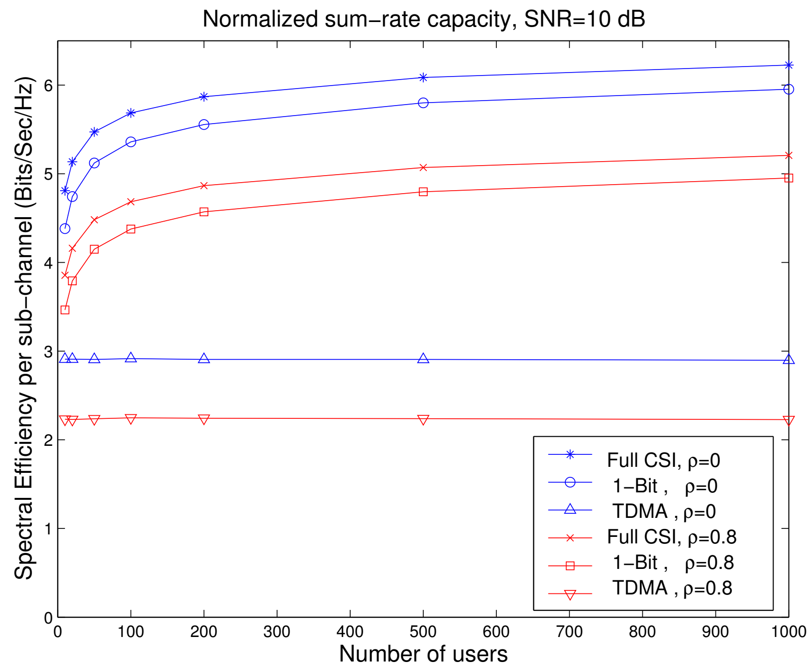

Figure 6.1. Sum-rate capacity (normalized by N) vs. number of users

Proof: Since the sub-channels are uncorrelated, for each sub-channel we can apply Theorem 5 to show that

$$\lim_{K\to\infty} \frac{C_{1\_bit}}{C_{full\_CSI}} = 1.$$

6.5 Subchannel Correlation

In this Section we assume that for each user the channel taps are correlated, but there is no dependence between different users' channels. The correlation model that we consider is an exponential decaying model described by

$$\mathbb{E}[h_{t,k}h^*_{s,k}] = \rho^{|t-s|} \tag{6.5}$$

This model well describes the correlation caused by a pulse-shaping filter that exits in many practical communications standards such as GSM.

Let $\eta_{n,k}$ be the power of the n^{th} subchannel of the k^{th} user which can be calculated as:

$$\begin{aligned}\eta_{n,k} &= \mathbb{E}[|H_{n,k}|^2] \\ &= \frac{1}{N}\mathbb{E}\left[\left(\sum_{t=0}^{N-1} h_{t,k}e^{-j\frac{2\pi tn}{N}}\right)\left(\sum_{s=0}^{N-1} h_{s,k}e^{-j\frac{2\pi sn}{N}}\right)^*\right] \\ &= \frac{1}{N}\mathbb{E}\left[\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} h_{t,k}h_{s,k}^* e^{-j\frac{2\pi(t-s)n}{N}}\right] \\ &= \frac{1}{N}\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} \rho^{|t-s|}e^{-j\frac{2\pi n}{N}(t-s)}\end{aligned}$$

by change of summing index to $u = t - s$ we get

$$\begin{aligned}\eta_{n,k} &= \sum_{u=-(N-1)}^{N-1} (1 - \frac{|u|}{N})\rho^{|u|}e^{-j\frac{2\pi n}{N}u} \\ &= \Re\{\beta_n\} - 1 \end{aligned} \quad (6.6)$$

where $\beta_n = \frac{1}{\sqrt{N}}\sum_{m=1}^{N-1} b_m e^{-j\frac{2\pi n}{N}m}$ is the discrete Fourier transform (DFT) of the sequence $b_m = (\sqrt{N} - \frac{m}{\sqrt{N}})\rho^m$. Assuming $\rho^N \ll 1$, after some algebra we obtain the following expression for $\eta_{n,k}$:

$$\eta_{n,k} \approx \frac{1-\rho^2}{1+2\rho\cos\theta_n+\rho^2} + \frac{2\rho\cos\theta_n + 4\rho^2 + 2\rho^3\cos\theta_n}{N(1+2\rho\cos\theta_n+\rho^2)^2} \quad (6.7)$$

where $\theta_n = \frac{2\pi n}{N}$. Notice that for a given n, $\{H_{n,k}\}$'s are i.i.d. across different users, hence $\eta_{n,k}$ does not depend on k. Thus correlation between taps leads to subchannels with different qualities. On the other hand exact calculation of the optimal threshold for this case is not mathematically tractable. So we propose a suboptimal solution for quantizing the subchannels with one bit. For the n^{th} frequency bin, we divide the subchannel gains by $\eta_{n,k}$ and then compare the normalized channel gain by the optimal threshold calculated in Section 4.4, if

$$\frac{|H_{n,k}|^2}{\eta_{n,k}} \geq \alpha_n$$

the feedback bit is set to "1", otherwise it is set to "0".

Figure 6.1 is the simulation result based on the proposed algorithm for SNR=10 dB. As can be seen in the figure for both uncorrelated and correlated cases, the sum-rate capacity growth of our scheme is the same as the opportunistic subchannel selection with full information available at the base station. Moreover the capacity achieved by our scheme is much higher than TDMA scheduling and only slightly lower than the full CSI sum-rate capacity.

CHAPTER 7
CONCLUSION AND FUTURE WORK

In this final chapter we summerize the contributions of this dissertation and discuss several avenues for future research in the area of opportunistic communications.

7.1 Contributions

- *Capacity analysis of antenna selection:* In Section 3.3 we study the behavior of the capacity of both transmit and receive antenna selection in the asymptote of large number of antennas. In high SNR scenario, the concept of *capacity gain* is developed. We evaluate the capacity gain for the uninformed transmitter (no CSI at the transmit side), fully informed (water-filling) and partially informed transmitter (antenna selection). We perform the same analysis fot these three cases using the concept of *channel gain* in the low-SNR scenario.

- *Fast algorithms for joint transmit-receive selection:* In Section 3.5 we propose two algorithms for joint transmit-receive antennal selection with quadratic and linear complexities (in number of antennas) that perform almost as well as the high complexity selection through exhaustive search.

- *Performance of antenna selection in keyhole channels:* In Section 3.6, we explore the performance of antenna selection in keyhole channels. We show that under the keyhole condition, selecting only one antenna at transmit side and a few antennas at receive side leads to an equivalent low-order SIMO system that has almost the capacity of the high order keyhole channel with no selection. Moreover we proove that the diversity of the system antenna selection is the same the baseline system.

- *Scheduling with 1-bit CSI feedback in multiuser SISO networks:* In Chapter 4 we propose a simple scheduling algorithm that only requires one-bit of information sent by the users to the base-station. We also prove that subject to a judicious choice of threshold, our method has the same sum-rate capacity growth as of scheduling with full CSI information.

- *Opportunistic beamforming with limited CSI feedback:* In Chapter 5, we extend the scheduling method proposed in Chapter 4 to the case where the base-station and/or the users are equipped with multiple antennas. We propose an opportunistic beamforming with limited CSI feedback that exploits both the multiuser diversity gain and the array gain of multiple antenna to improve the overall capacity. We prove the optimality of the capacity growth of our scheme and show its superior performance over conventional beamforming methods.

- *Opportunistic sub-channel allocation with limited CSI feedback for OFDM networks:* In Chapter 6 we extend the scheduling scheme proposed in Chapter 4 for OFDM networks. We propose a scheduling algorithm for dynamic sub-channel allocation with minimal feedback requirement and prove that most of multiuser diversity gain can be captured by this method.

- *A novel mathematical framework for asypmtotic capacity analysis of opportunistic systems:* In Appendix A, we prove theorems on the tightness of Jensen's inequality for a large class of functions. Throughout this thesis we frequently use these techniques for asymptotic analysis of the capacity of systems. The domain of the application of these results includes a large spectrum of problems.

7.2 Future Work

There are a variety of research directions for future work on opportunistic communications at both the link lavel and the network level.

- *Antenna Selection.*

  There are many unsolved problems in the area of antenna selection. Channel estimation for antenna selection is an important subject for future research. Analysis and code design for antenna selection still requires more investigation. Performance evaluation of antenna selection algorithms when the channel matrix is not perfectly known at the receiver is a seemingly important yet relatively unexplored problem. Combination of antenna selection with space-time signaling schemes has been noted by several investigators, but much work remains in this area and it is a worthy subject of future research.

- *Opportunistic Scheduling with Limited CSI*

  It is instructive to analyze the delay-throughput trade-off for our propsed scheme. A interesting research direction is to extend our method for multicast scheduling. Cross layer problems related to our proposed scheme can certainly be a fruitful area for research; the one-bit CSI feedback can not only improve the physical layer, but also can be used in transport layer for congestion control.

APPENDIX

APPENDIX A
ASYMPTOTIC TIGHTNESS OF JENSEN'S INEQUALITY

In this appendix we prove results that are used in asymptotic capacity analysis. Jensen's inequality is a useful tool in the analysis of the capacity in communications system; Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a measurable concave function and let X be a random variable, then

$$\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$$

The proof of Jensen's inequality can be found in standard textbooks on probability theory e.g. [53]. The equality holds if and only if the function $g(\cdot)$ is an affine function or the probability measure associated with the random variable X is a concentrated at a single point, i.e. $F_X(x) = \delta(x - x_0)$. For a family of random variables $\{X_n\}_{n=1}^{\infty}$, we are interested to see when Jensen's inequality becomes tight. Heuristically, if the function $g(\cdot)$ is slowly varying (hence locally linear)and the probability measure $F_{X_n}(\cdot)$ is well concentrated around the mean value in the asymptote of large n, then we expect to have a tight bound by Jensen's inequality. In order to investigate the asymptotic tightness of Jensen's inequality we need some preliminary results:

Lemma 2 *For all $x, y \geq 0$ we have:*

$$|\log(1 + \rho x) - \log(1 + \rho y)| \leq \log(1 + \rho|x - y|)$$

Proof: Since $g(x) = \log(1 + \rho x)$ is an increasing function, without loss of generality we can assume $x \geq y \geq 0$. Therefore we have $\rho^2 y(x - y) \geq 0$ this inequality can be re-written as $\frac{1+\rho x}{1+\rho y} \leq 1 + \rho(x - y)$ by taking the logarithm from both sides we arrive at the desired inequality. ■

Definition 2 *The family of random variables* $\{X_n\}$ *is said to be uniformly integrable if*

$$\lim_{c\to\infty} \limsup_{n} \int_c^\infty x \ dF_{|X_n|}(x) = 0$$

Uniform integrability is the sufficient condition for convergence in mean.

Theorem 9 *If* X_n *is a uniformly integrable random variable and* $\mathbb{E}[X] < \infty$*, then convergence in distribution, implies convergence in mean, i.e., if* $X_n \xrightarrow{i.p.} X$ *then* $\mathbb{E}[X_n] \to \mathbb{E}[X]$*.*

Proof: See [54]. ■

The following lemma provides an alternative way to prove the uniform integrability.

Lemma 3 *If* $\mathbb{E}[|X_n|] < \infty$ *for all* n*, then the random variable* X_n *is uniformly integrable if*

$$\lim_{c\to\infty} \limsup_{n} \int_c^\infty \Pr[|X_n| > t]dt = 0$$

Proof: For every n and $c > 0$ we have

$$c(1 - F_{|X_n|}(c)) \leq \int_c^\infty x \ dF_{|X_n|}(x) \leq \mathbb{E}[|X_n|] < \infty$$

hence $\lim_{c\to\infty} c(1 - F_{|X_n|}(c)) = 0$. Using integration by-part we have

$$\begin{aligned} 0 \leq \int_c^\infty x \ dF_{|X_n|}(x) &= -x(1 - F_{|X_n|}(x))\,|_c^\infty + \int_c^\infty \Pr[|X_n| > x] \ dx \\ &= -c(1 - F_{|X_n|}(c)) + \int_c^\infty \Pr[|X_n| > x] \ dx \\ &\leq \int_c^\infty \Pr[|X_n| > x] \ dx \end{aligned}$$

and this proves the lemma. ■

Lemma 4 *If* $X_n \xrightarrow{i.p.} 0$ *and* $a_n \to 0$*, then,* $a_n \cdot X_n \xrightarrow{i.p.} 0$

Proof: For every $\epsilon, \delta_1, \delta_2 > 0$, there exits N_1 such that for all for all $n > N_1$ we have $|a_n| < \delta_1$. Also there exits N_2 such that for all $n > N_2$, $\Pr[|X_n| > \frac{\epsilon}{\delta_1}] < \delta_2$, thus for all $n > N = \min\{N_1, N_2\}$,

$$Pr[|a_n X_n| > \epsilon] = \Pr[|X_n| > \frac{\epsilon}{|a_n|}] \leq \Pr[|X_n| > \frac{\epsilon}{\delta_1}] \leq \delta_2$$

thus $a_n X_n \xrightarrow{i.p.} 0$. ■

Now we prove the following theorems for asymptotic tightness of Jensen's inequality,

Theorem 10 *Let* $\{X_n\}$ *be a family of positive i.i.d. random variable with finite mean* μ_n *and variance* σ_n^2*, also* $\mu_n \to \infty$ *and* $\frac{\sigma_n}{\mu_n} \to 0$ *as* $n \to \infty$*, then for all* $\rho > 0$ *we have*

$$\frac{\mathbb{E}[\log(1 + \rho X_n)]}{\log(1 + \rho \mathbb{E}[X_n])} \longrightarrow 1. \tag{A.1}$$

Proof: Using the Chebyshev's inequality for all $\epsilon > 0$ we have:

$$\begin{aligned} \Pr\left[\left|\frac{1 + \rho X_n}{1 + \rho \mu_n} - 1\right| > \epsilon\right] &= \Pr\left[\left|\frac{X_n - \mu_n}{1/\rho + \mu_n}\right| > \epsilon\right] \\ &\leq \frac{\mathbb{E}[(X_n - \mu_n)^2]}{\epsilon^2 (1/\rho + \mu_n)^2} \\ &= \frac{1}{\epsilon^2}\left(\frac{\sigma_n}{\mu_n}\right)^2 \end{aligned}$$

hence $\frac{1+\rho X_n}{1+\rho \mu_n} \xrightarrow{i.p.} 1$. Using the *continuous mapping theorem*, we have

$$\log\left(\frac{1 + \rho X_n}{1 + \rho \mu_n}\right) \xrightarrow{i.p.} 0 \ .$$

On the other hand $\mu_n \to \infty$, hence $\frac{1}{\log(1+\rho\mu_n)} \to 0$ and we can invoke Lemma 4 to conclude

$$\frac{\log\left(\frac{1+\rho X_n}{1+\rho \mu_n}\right)}{\log(1 + \rho \mu_n)} = \frac{\log(1 + \rho X_n)}{\log(1 + \rho \mu_n)} - 1 \xrightarrow{i.p.} 0$$

Thus

$$\frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} \xrightarrow{i.p.} 1.$$

We show that the random variable $Z_n = \frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1$ is uniformly integrable.

$$\begin{aligned} I = & \int_c^\infty \Pr\left[\left|\frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1\right| > t\right] dt \\ = & \int_c^\infty \Pr\left[\frac{|\log(1+\rho X_n) - \log(1+\rho\log\mu_n)|}{\log(1+\rho\mu_n)} > t\right] dt \\ \leq & \int_c^\infty \Pr\left[\frac{\log(1+|X_n-\mu_n|)}{\log(1+\rho\mu_n)} > t\right] dt \qquad \text{using Lemma 2} \\ = & \int_c^\infty \Pr\left[\rho|X_n-\mu_n| > A_n^t - 1\right] dt \end{aligned}$$

where $A_n = 1+\rho\mu_n$. Using Chebyshev's inequality we have

$$I \leq \int_c^\infty \Pr\left[\rho|X_n-\mu_n| > A_n^t - 1\right] dt \leq \rho^2\sigma_n^2 \int_c^\infty \frac{dt}{(A_n^t-1)^2}$$

We use change of variable $u = A_n^t - 1$, $dt = \frac{du}{\log(A_n)(u+1)}$, $\alpha_n = A_n^c - 1$

$$\begin{aligned} I & \leq \rho^2\sigma_n^2 \frac{1}{\log(A_n)} \int_{\alpha_n}^\infty \frac{du}{u^2(u+1)} \\ & \leq \rho^2\sigma_n^2 \frac{1}{\log(A_n)} \int_{\alpha_n}^\infty \frac{du}{u^2} \\ & = \rho^2 \left(\frac{\sigma_n}{\mu_n}\right)^2 \frac{A_n^2}{\log(A_n)({A_n}^c-1)}. \end{aligned}$$

Since $\mu_n \to \infty$, $A_n \to \infty$ thus for all $c > 2$ $\frac{A_n^2}{\log(A_n)(A_n^c-1)} \to 0$. Also $\frac{\sigma_n}{\mu_n} \to 0$, hence $I \to 0$ and $Z_n = \frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1$ is uniformly integrable. Therefore from Theorem 9 we conclude

$$\frac{\mathbb{E}[\log(1+\rho X_n)]}{\log(1+\rho\mu_n)} \longrightarrow 1$$

as $n \to \infty$. ■

The Shannon function $g(x) = \log(1+\rho x)$ is not the only function for which Jensen's inequality is asymptotically tight. We introduce a larger class of function that have this property.

Definition 3 *We define $\mathcal{G}$ as the set of measurable functions $g : \mathbb{R}^+ \mapsto \mathbb{R}^+$ that are both increasing and concave with $g(0) = 0$.*

$$\mathcal{G} = \{g : \mathbb{R}^+ \mapsto \mathbb{R}^+ | \forall x > 0,\ g^{'}(x) > 0,\ g^{''}(x) < 0,\ g(0) = 0\}$$

In the following theorem we explore the properties of function in $\mathcal{G}$.

Theorem 11 *$\mathcal{G}$ has the following properties*

1. *$\mathcal{G}$ is closed under function composition.*
2. *for every $g \in \mathcal{G}$, $\frac{g(x)}{x}$ is a decreasing function*
3. *$g \in \mathcal{G}$ is a sub-additive function, hence*

$$\forall x, y > 0\ , \quad |g(x) - g(y)| < g(|x - y|)$$

Proof: For part 1 we note that if $g, f \in \mathcal{G}$ then $g \circ f(x) = g(f(x))$ is defined in $\mathbb{R}^+$. Also $(g \circ f)^{'}(x) = f^{'}(x) g^{'}(f(x)) > 0$ and $(g \circ f)^{''}(x) = f^{''}(x) g^{'}(f(x)) + (f^{'}(x))^2 g^{''}(f(x)) < 0$ therefore $g \circ f \in \mathcal{G}$. For proving part 2, we use the concavity of $g(\cdot)$. For all $x, y > 0$ and $\alpha \in (0, 1)$,

$$\alpha g(x) + (1 - \alpha) g(y) < g(\alpha x + (1 - \alpha) y)$$

let $y = 0$, then $\alpha g(x) < g(\alpha x)$. This can be written as $\frac{g(x)}{x} < \frac{g(\alpha x)}{\alpha x}$ which means that $g(\cdot)$ is a decreasing function. In part 3, we first prove that $g(\cdot)$ is sub-additive. Using part 2, for all $x, y > 0$ we have $\frac{g(x+y)}{x+y} < \frac{g(x)}{x}$ hence.

$$\frac{x}{x+y} g(x + y) < g(x)$$

similarly we get

$$\frac{y}{x+y} g(x + y) < g(y)$$

By adding these two inequalities, we arrive at $g(x+y) < g(x) + g(y)$, which means $g(\cdot)$ is a sub-additive function. Without loss of generality we assume $x \geq y \geq 0$, therefore $g(x) = g(y + (x-y)) < g(y) + g(x-y)$ hence $g(x) - g(y) < g(x-y)$. But $|g(x) - g(y)| = g(x) - g(y)$ because $g(\cdot)$ is increasing. Thus

$$|g(x) - g(y)| < g(|x-y|).$$

■

The following theorem establishes the conditions under which Jensen's inequality for any function $g \in \mathcal{G}$ is asymptotically tight.

Theorem 12 *Let $\{X_n\}$ be a family of positive i.i.d. random variable with finite mean μ_n and variance σ_n^2, so that $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$ as $n \to \infty$, then for every $g \in \mathcal{G}$ satisfying the condition $\lim_{x\to\infty} \frac{g(x)}{x} = K > 0$, we have*

$$\frac{\mathbb{E}[g(X_n)]}{g(\mathbb{E}[X_n])} \longrightarrow 1. \tag{A.2}$$

Proof: As shown in the proof of Theorem 10, $\frac{\sigma_n}{\mu_n} \to 0$ implies $\frac{X_n}{\mu_n} \to 1$. Using the Chebyshev inequality and Theorem 3 (part 3) for every $a > 0$ we have

$$\begin{aligned}
\Pr\left[\left|\frac{g(X_n)}{g(\mu_n)} - 1\right| > a\right] &= \Pr\left[\left|\frac{g(X_n) - g(\mu_n)}{g(\mu_n)}\right| > a\right] \\
&\leq \Pr\left[\frac{g(|X_n - \mu_n|)}{g(\mu_n)} > a\right] \\
&\leq \Pr\left[|X_n - \mu_n| > g^{-1}(ag(\mu_n))\right] \\
&\leq \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \left(\frac{\mu_n}{g^{-1}(ag(\mu_n))}\right)^2
\end{aligned}$$

We have $\mu_n \to \infty$ as $n \to \infty$. Also the condition $\lim_{x\to\infty} \frac{g(x)}{x} = K > 0$, implies that $g(x) \to \infty$ as $x \to \infty$. g^{-1} is an increasing continuous function therefore $g^{-1}(ag(\mu_n)) \to \infty$, thus

$$\lim_{n\to\infty} \frac{\mu_n}{g^{-1}(ag(\mu_n))} = \lim_{n\to\infty} \frac{g(g^{-1}(ag(\mu_n)))}{ag^{-1}(ag(\mu_n))} = \frac{K}{a}$$

Since $\frac{\sigma_n}{\mu_n} \to 0$ from A.3 we can conclude that $\frac{g(X_n)}{g(\mu_n)} \xrightarrow{i.p.} 1$. We need to show that the random variable $Z_n = \frac{g(X_n)}{g(\mu_n)} - 1$ is uniformly integrable. Using Lemma 3 and A.3 we have

$$\begin{aligned} I_n &= \int_c^\infty \Pr\left[\left|\frac{g(X_n)}{g(\mu_n)} - 1\right| > a\right] da \\ &\leq \frac{\sigma_n^2}{\mu_n^2} \int_c^\infty \frac{\mu_n^2 da}{\left(g^{-1}(ag(\mu_n))\right)^2} \end{aligned}$$

We use the change of variable $u = g^{-1}(ag(\mu_n))$, $da = \frac{g^{'}(u)}{g(\mu_n)} du$, $\alpha_n = g^{-1}(cg(\mu_n))$, also we note that $g^{'}(\cdot)$ is a decreasing function thus for $u \in (\alpha_n, \infty)$ we have $g^{'}(u) < g^{'}(\alpha_n)$, hence

$$\begin{aligned} I_n &\leq \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \mu_n^2 \int_{\alpha_n}^\infty \frac{g^{'}(u)}{u^2} \frac{du}{g(\mu_n)} \\ &< \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{\mu_n^2 g^{'}(\alpha_n)}{g(\mu_n)} \int_{\alpha_n}^\infty \frac{du}{u^2} \\ &= \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{\mu_n^2 g^{'}(\alpha_n)}{g(\mu_n)} \int_{\alpha_n}^\infty \frac{du}{u^2} \\ &= \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{1}{c} \cdot \frac{\mu_n}{g(\mu_n)} \cdot \frac{g(\alpha_n)}{\alpha_n} \cdot g^{'}(\alpha_n) \end{aligned}$$

By L'Hospital's rule, $\lim_{x\to\infty} g^{'}(x) = \lim_{x\to\infty} \frac{g(x)}{x} = K$. Also $\mu_n \to \infty$ implies $\alpha_n \to \infty$, therefore $\lim_{n\to\infty} \frac{\mu_n}{g(\mu_n)} \cdot \frac{g(\alpha_n)}{\alpha_n} \cdot g^{'}(\alpha_n) = K$, hence $\lim_{n\to\infty} I_n = 0$ which means $Z_n = \frac{g(X_n)}{g(\mu_n)} - 1$ is uniformly integrable thus $\mathbb{E}[Z_n] \longrightarrow 0$

■

Note: In the above theorem, if $\limsup \sigma_n < \infty$ then the condition $\lim_{x\to\infty} \frac{g(x)}{x} = K > 0$ can be weakened to $\lim_{x\to\infty} g(x) = +\infty$.

The following theorem combines the results from Theorem 10 and Theorem 12:

Theorem 13 *Under the assumptions of Theorem 12 we have*

$$\frac{\mathbb{E}[\log(1 + \rho g(X_n))]}{\log(1 + \rho g(\mathbb{E}[X_n])} \longrightarrow 1$$

Proof: The proof is straightforward and very similar to the proof of Theorem 10. ■

REFERENCES

[1] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[2] A. F. Molisch, M. Z. Win, and J. H. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. International Conference on Communications*, Helsinki, Finland, June 2001, pp. 570–574.

[3] I. Bahceci, T. M. Duman, and Y. Altunbasak, "Antenna selection for multiple-antenna transmission systems: performance analysis and code construction," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2669 – 2681, Oct. 2003.

[4] A. Gorokhov, "Antenna selection algorithms for MEA transmission systems," in *Proc. IEEE ICASSP*, Orlando, FL, May 2002, pp. 2875–2860.

[5] M. Gharavi-Alkhansari and A. Greshman, "Fast antenna selection in MIMO systems," *to appear in the IEEE Trans. on Signal Proc. special issue on space-time communication*, 2003.

[6] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms," *IEEE Trans. on Signal Processing*, pp. 2796 –2807, Nov. 2003.

[7] S. Sanayei and A. Nosratinia, "Asymptotic capacity gain of transmit antenna selection," in *Proc. WNCG Symposium*, Austin, TX, Oct. 2003.

[8] S. Sanayei and A. Nosratinia, "Asymptotic capacity analysis of transmit antenna selection," in *Proc. of IEEE International Symposium on Information Theory*, Chicago, IL, June 2004, p. 241.

[9] S. Sanayei and A. Nosratinia, "Capacity of MIMO channels with antenna selection," 2005, submitted to IEEE Trans. on Info. Theory.

[10] S. Sanayei and A. Nosratinia, "Capacity maximizing algorithms for joint transmit-receive antenna selection," in *Proc. of the Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2004.

[11] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: a bandwidth efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, pp. 70–77, July 2000.

[12] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. International Conference on Communications*, Seattle, WA, June 1995.

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

[14] P. Viswanath, D.N.C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[15] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. on Information Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[16] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with only 1-bit feedback," in *IEEE Wireless Communication and Networking Conference (WCNC)*, New Orleans, LA, March 2005.

[17] C. E. Shannon, "A mathematical theory of communications," *Bell Sys. Tech. Journal*, vol. 27, pp. 379–423, 623–656, 1948.

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

[19] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2619 – 2692, Oct. 1998.

[20] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, John Wiley and Sons, 1992.

[21] H. A. David, *Order Statistics*, Wiley, 1981.

[22] A. Rényi, "On the theory of order statistics," *Acta. Math Acad. Sci. Hung.*, vol. 4, 1953, 191-231.

[23] J. Galambos, *Asymptotic Theory of Order Statistics*, Krieger Pub. Co., 1987.

[24] G. J. Foschini and M. J. Gans, "On limits of wireless communication in fading environment when using multiple antennas," *Wireless Personal Communication*, vol. 6, pp. 311–335, March 1998.

[25] R. S. Blum and J. H. Winters, "On optimum MIMO with antenna selection," *IEEE Communications Letters*, vol. 6, pp. 322–324, Aug. 2002.

[26] D. Gore, R. Heath, and A. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Communications Letters*, vol. 6, no. 1, pp. 491–493, Nov. 2002.

[27] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO flat-fading channels: Theory and algorithms," *IEEE Trans. on Information Theory*, pp. 2687 –2696, Oct. 2003.

[28] A. Lozano, A. Tulino, and S. Verdu, "High-SNR power offset in multi-antenna communication," in *Proc. of IEEE International Symposium on Information Theory*, Chicago, IL, June 2004, p. 288.

[29] D. Bliss, K. Forsythe, A. Hero, and A. Yegulalp, "Environmental issues for MIMO capacity," *IEEE Trans. on Signal Processing*, vol. 50, no. 9, pp. 2128–2142, September 2002.

[30] S. Verdu, "Spectral efficiency in wideband regime," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[31] G. Kéri, "The Sherman-Morisson formula for the determinant and its application for optimizing the quadratic functions on conditinos sets given by extreme generators," in *Optimization Theory: Recent Developments from Mátraháza*, F. Giannessi, P. Pardalos, and T. Rapcsàk, Eds. Kluwer, 2001.

[32] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels: Insights into diversity-multiplexing tradeoff," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[33] S. Geman, "A limit theorem for the norm of the random matrices," *Annals of probability*, vol. 8, pp. 252–261, 1980.

[34] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multi-element transmit and receive antennas," *IEEE Trans. on Wireless Communications*, vol. 1, no. 2, pp. 361–368, April 2002.

[35] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.

[36] M. Z. Win and J. H. Winter, "Analysis of hybrid selection/maximal-ratio combining in rayleigh fading," *IEEE Trans. on Communications*, vol. 47, no. 12, pp. 1773–1776, Dec. 1999.

[37] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Trans. on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2003.

[38] Shahab Sanayei and Aria Nosratinia, "Space-time codes in keyhole channels: Analysis and design issues," in *Proc. of IEEE Global Telecommunications Conference*, Dallas, TX, Dec. 2004, vol. 6, pp. 3768 – 3772.

[39] M. Abramowitz and A. I. Stegun, *Handbook of Mathematical Functions*, New York, Dover, 1972.

[40] D. Tse, "Optimal power allocation over parallel Gaussian channels," in *Proc. of IEEE International Symposium on Information Theory*, Ulm, Germany, June 1997.

[41] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1893 – 1909, Sept. 2004.

[42] D. Gesbert and S. Alouini, "How much feedback is multi-user diversity really worth," in *Proc. International Conference on Communications*, Paris, France, June 2004.

[43] D. Gesbert and S Alouini, "Selective multiuser diversity," in *in Proc. of Intl. Symp. on Sig. Proc. and Info. Techn. (ISSPIT)*, Darmstadt, Germany, Dec 2003.

[44] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, 4th edition, 1996.

[45] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[46] B. Hassibi and M. Sharif, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels," in *Proc. of IEEE International Symposium on Information Theory*, Chicago, IL., June 27 - July 2 2004, p. 175.

[47] J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[48] M Kobayashi, G. Caire, and D. Gesbert, "Antenna diversity vs. multiuser diversity: quantifying the tradeoffs," in *International Symposium on Information Theory and its Applications*, Parma, Italy, October 2004.

[49] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Information Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[50] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. on Information Theory*, vol. 49, pp. 2562–2579, Oct. 2003.

[51] P. Svedman, S. K. Wilson, L. Cimini, and B. Ottersten, "A simplified feedback and scheduling scheme for OFDM," in *Proc. Vehicular Technology Conference*, Milan, Italy, May 2004.

[52] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. Vehicular Technology Conference*, May 2000, vol. 2, pp. 1085–1089.

[53] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 1*, Wiley, 1971.

[54] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, 1980.

VITA

Shahab Sanayei was born on February 4^{th}, 1973, in Mashad, Iran, son of Sara and Rahman Sanayei. He received his B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1995 and his M.S. degree in electrical engineering from Rice University, Houston, TX, in 2002. From 1994 to 1997 he was with the Intelligent Systems Laboratory in Sharif University of Technology. From 1997 to 2000 he was an R&D engineer with Towzin Electric Co., Tehran, Iran. Since Sepember 2002 he has been with the department of electrical engineering at the University of Texas at Dallas, where he has been working on his PhD degree in electrical engineering. His research interests include MIMO systems, Signal Processing, Stochastic Processes and Information Theory.