UNIVERSITY OF CALIFORNIA, SAN DIEGO

Feedback Reduction Techniques and Fairness in Multi-User MIMO Broadcast Channels with Random Beamforming

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Matthew Owen Pugh

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor Ery Arias-Castro
Professor William S. Hodgkiss
Professor Kenneth Kreutz-Delgado
Professor Laurence B. Milstein

2011

The dissertation of Matthew Owen Pugh is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____
Chair

University of California, San Diego

2011

DEDICATION

To William Leo Hollar

EPIGRAPH

You see, wire telegraph is a kind of a very, very long cat. You pull his tail in New York and his head is meowing in Los Angeles. Do you understand this? And radio operates exactly the same way: you send signals here, they receive them there. The only difference is that there is no cat.

—Albert Einstein

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First I would like to thank my loving and supporting parents Michael and Janet Pugh. Without them, I would obviously not be where I am today and I owe everything to them. I would also like to thank my amazing sister Jacqueline Pugh who makes me want to be the best brother and person possible. I am also thankful beyond words to my girlfriend Veronica Sooley for helping me achieve my dreams.

To my advisor Professor Bhaskar D. Rao I am deeply indebted and thankful. His guidance, thoughtfulness, and deep insights have made this dissertation a reality. I can never truly repay him for taking a risk on this previously wandering graduate student and transforming him into a successful researcher.

I am very thankful to my committee members Professor Ery Arias-Castro, Professor William S. Hodgkiss, Professor Kenneth Kreutz-Delgado and Professor Laurence B. Milstein. I have been fortunate to learn a great deal from each of these tremendous teachers and researchers. I would also like to thank Professor Jason Schweinsberg who taught me to how to appreciate, understand, and finally love probability. Without this knowledge this thesis would not be as complete.

My lab-mate Yogananda Isukapalli has been a great mentor and role-model. I learned a great deal about what it means to be a successful graduate student by following in his footsteps. To my other long-time lab-mates Nandan Das, Yichao Huang, Seong-Ho Hur, Eddy Kwon, Oleg Tanchuk and LiwenYu I am thankful for putting up with my long group meeting presentations and priceless constructive critiques. Of my lab-mates, I am especially thankful to Sagnik Ghosh and Sheu-Sheu Tan for being incredible friends and helping keep me sane with laughter.

When not reading papers or attempting research, I am thankful to all of my friends for giving some semblance of a social life. I am particular thankful to Ravi Agarwal, Steve Cho, Natan Jacobson, Omer Lang, Nickolaus Mueller, Jung Park and Jack Tzeng for their support and non-technical, but infinitely funnier, conversations over coffee, dinner or lunch. To my close high school friends Matt Carroll, Pat Carroll, Adam Darrow, Matt Darrow, Tu Nguyen, and Eric Ping I am thankful for their continued friendship .

Finally, I want to thank my grandfather William Hollar, to whom I dedicate this dissertation, for showing me the importance of hard work, dedication and wonder.

This work is supported in part by the University of California Discovery Grant #07-10241, Intel Corp., Qualcomm Inc., Texas Instruments Inc. and the Center for Wireless Communications at the University of California, San Diego.

This dissertation is a collection of papers that were published, submitted for publication, or in preparation for submission for publication. Chapter 2 is a reprint of a paper coauthored with Bhaskar D. Rao and appears in the March 2010 issue of *IEEE Transactions on Signal Processing* under the title *"Reduced Feedback Schemes Using Random Beamforming in MIMO Broadcast Channels"*. Sections of chapter 2 also appear in *"On the Capacity of MIMO Broadcast Channels with Reduced Feedback by Antenna Selection"* in *Asilomar Conference on Signals Systems, and Computers, March 2008* and *"Feedback reduction in MIMO broadcast channels with LMMSE receivers"* in *ICASSP 2010*. Both of these works were coauthored with Bhaskar D. Rao. The material in Chapter 3 is work which in preparation for submission to the *IEEE Transactions on Signal Processing* under the title *"Reducing Feedback in Broadcast Channels via Thresholding"* and material submitted to *Asilomar Conference on Signals, Systems, and Computers 2011* under the title *"Feedback Reduction by Thresholding in Multi-User Broadcast Channels: Design and Limits"*. These works are coauthored with Bhaskar D. Rao. Chapter 4 is material coauthored with Bhaskar D. Rao and is in preparation for submission to *IEEE Transactions on Information Theory* under the title *"Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models"*. The dissertation author is the primary researcher and author, and the co-author listed in these publications contributed to or supervised the research which forms the basis of this dissertation.

VITA

2005	B. S. in Electrical Engineering *summa cum laude*, University of California, Los Angeles
2005	B. S. in Applied Mathematics, University of California, Los Angeles
2006-2007	Teaching Assistant, University of California, San Diego
2008	M. S. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego
2007-2011	Research Assistant, University of California, San Diego
2011	Ph. D. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

Matthew Pugh and Bhaskar D. Rao, “Reduced Feedback Schemes Using Random Beamforming in MIMO Broadcast Channels,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, March 2010, pages 1821-1832.

Matthew Pugh and Bhaskar D. Rao, “Feedback Reduction by Thresholding in Multi-User Broadcast Channels: Design and Limits”, *Submitted to Asilomar Conference on Signals, Systems, and Computers 2011.*

Matthew Pugh and Bhaskar D. Rao, “Reducing Feedback in Broadcast Channels via Thresholding”, *In preparation for submission to IEEE Transactions on Signal Processing.*

Matthew Pugh and Bhaskar D. Rao, “Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models”, *In preparation for submission to IEEE Transactions on Information Theory.*

Matthew Pugh and Bhaskar D. Rao, “Distributed Quantization of Order Statistics with Applications to CSI Feedback”, *Data Compression Conference, March 2011.*

Matthew Pugh and Bhaskar D. Rao, “Feedback reduction in MIMO broadcast channels with LMMSE receivers”, *ICASSP 2010.*

Matthew Pugh and Bhaskar D. Rao, “On the Capacity of MIMO Broadcast Channels with Reduced Feedback by Antenna Selection”, *Asilomar Conference on Signals, Systems, and Computers, March 2008.*

FIELDS OF STUDY

Major Field: Engineering

Studies in communication theory, information theory, estimation theory, probability theory, stochastic processes, stochastic approximation theory and their applications in feedback reduction in multi-user MIMO broadcast channels.
Professor Bhaskar D. Rao, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Feedback Reduction Techniques and Fairness in Multi-User MIMO Broadcast Channels with Random Beamforming

by

Matthew Owen Pugh

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California, San Diego, 2011

Professor Bhaskar D. Rao, Chair

With the rise of cellular communications, the broadcast channel has gained prominence since it is an effective model of the cellular downlink channel. Multiple-input multiple-output (MIMO) communications has also gained in popularity due to its theoretically promised performance benefits over traditional single antenna systems. Combining these two theoretical objects yields the the MIMO, or vector, broadcast channel. To achieve the highest possible performance it is known that channel state information (CSI) from each user in the broadcast channel must be known at the transmitter. This requires each user to feed back this information, which is an unwanted overhead.

The first part of the dissertation discusses methods to reduce this feedback overhead by considering use of multiple receive antennas. Under the random beamforming transmit methodology, the feedback is reduced by considering only feeding back the largest SINR value observed at each user. Analysis of this reduced feedback scheme involves finding the distribution of the largest SINR over correlated random variables. Each user can also use the additional receive antennas to perform LMMSE reception. The distribution of the post-processed SINR is found and used to compute the system performance. Feedback can be reduced after LMMSE reception by feeding back only the largest post-processed SINR. Bounding techniques on the distribution function of the maximum post-processed SINR are used to evaluate the performance of this scheme. Fixed finite thresholds are shown to have no asymptotic effects on the system performance.

The second part of the thesis considers the design of thresholds as a function of the number of users to reduce feedback. The asymptotic properties of any successful threshold are derived, namely any threshold $T(n) \in o(\log n)$ asymptotically achieves the optimal sum-rate scaling rate while any threshold $T(n) \in \omega(\log n)$ loses all multi-user diversity. Under three proposed system performance metrics, the optimal threshold is obtained. These metrics are: constraining the average number of users providing feedback, constraining the probability that no user feeds back, and constraining the rate lost due to thresholding.

The final portion of the dissertation address the question of fairness in the random beamforming technique. To improve fairness the proportional fair sharing algorithm is proposed as a substitute for the greedy scheduling algorithm. Under the Rayleigh fading model, the asymptotic performance is found to be identical to the asymptotic performance of the greedy algorithm. The rate of convergence of the proportional fair sharing algorithm to the asymptotic performance limit is found. The distribution of the distance from the asymptotic limit is computed. Additional analysis is performed on the use of thresholds under proportional fair sharing.

Chapter 1

Introduction

With cellular technologies becoming ubiquitous, and the demand for multimedia content increasing, maximizing the throughput of the downlink channel is an important and challenging problem. In traditional point-to-point communication problems, transmission schemes that are agnostic of the channel conditions have been developed that can achieve near-optimal performance. Thus in a traditional setting, the need for CSI is not critical. In the multi-user communication scenario of the broadcast channel, CSI becomes absolutely necessary to achieve the best possible performance. The reason for this is that in the multi-user scenario, there exists a form of diversity which cannot be exploited without channel knowledge, and that diversity is multi-user diversity. When there are many users in a broadcast channel, each user will experience varying channel conditions, some of which are good and some of which are poor. With CSI available at the transmitter, or the base station in the cellular downlink channel, the transmitter can schedule users who are currently in strong channel conditions that facilitate the highest throughputs. By knowing the CSI and the channels that can support high data rates, the transmitter is exploiting the multi-user diversity. Without any CSI at the transmitter, such scheduling decisions cannot be made and all multi-user diversity is lost. Therefore, as previously alluded to, CSI is required in order to exploit the multi-user diversity of the broadcast channel. The transmitter is made aware of the CSI by each user making some measurement of the channel and feeding back this measurement. This feedback is an unwanted but necessary overhead to get the best possible system performance.

This dissertations analyzes various methods of reducing the feedback of channel state information (CSI) in multi-user multiple-input multiple-out (MIMO) broadcast channels. Much research has been performed on designing efficient channel metrics that capture the nature of the channel conditions and on effective means of conveying this metric back to the transmitter. Under the random beamforming transmit scheme where signal-to-interference plus noise ratio (SINR) is the channel metric which is fed back, various techniques are developed and analyzed to reduce this feedback. Feedback reduction can be achieved in a MIMO broadcast channel by exploiting the additional spatial diversity at each user created by the additional spatially separated receive antennas. Feedback in this case is reduced by considering only the best channel locally at each user over receive antennas or by reducing the total number of SINR values by employing an optimal linear receiver at each user. To further facilitate feedback reduction a threshold can be implemented at each user such that if a particular user's SINR is below the threshold, the SINR value is not fed back to the transmitter. Originally it was shown that no asymptotic throughput is lost by the implementation of any finite threshold. Additionally research was done to find the scaling rates on any possible threshold as a function of the number of users in the system such that optimal performance was obtained asymptotically. Performance metric were proposed and optimal threshold derived under these metric for non-asymptotic results. Finally, when exploiting multi-user diversity, the fairness of the system can be a concern. Some users may suffer from poor channel conditions for extended periods of time, and thus not be scheduled by a greedy scheduling algorithm such as that used by the random beamforming transmit scheme. To address this concern, use of the proportional fair sharing algorithm (PFS) is suggested and under the proposed system model the convergence to the performance of the greedy scheduling algorithm is proven. To once again reduce feedback, the application of a threshold to the PFS algorithm is considered and analyzed.

This chapter is organized as follows. Section 1.1 provides some background on fading channels, the need for channel state information in multi-user systems and the random beamforming transmit scheme under the i.i.d. Rayleigh fading channel model. Section 1.2.1 provides details on the methodologies used to reduce feedback and section 1.2.2 describes the design of thresholds to reduce feedback. Finally, section 1.2.3 discusses in more detail the proportional fair sharing algorithm and how it accommodates fairness.

1.1 Background

1.1.1 Background on Fading Channels

In a traditional point-to-point communication system with a single transmit and receive antenna, the probability of symbol error decays exponentially fast with respect to signal-to-noise ratio (SNR) when the channel is an additive white gaussian noise (AWGN) channel. For simplicity, consider a narrowband channel whose symbols are indexed by m. The input-output relationship of the AWGN channel is simply

$$y[m] = x[m] + n[m] \tag{1.1}$$

where $x[m]$ is the transmit symbol, $y[m]$ the receive symbol, and $n[m]$ the additive white gaussian noise. Under the AWGN model, errors occur only when the additive noise term is so large that it shifts the transmit symbol into the decision region of a different symbol at the receiver, and this leads to the exponential rate of decay of the symbol error probability as a function of SNR.

Most channels unfortunately do not behave as the AWGN channel. The environment in which the signal is propagated significantly affects the signal itself. For example, a signal that is transmitted may have a direct line of sight path to the receiver, as well as an indirect path that involves reflections off of near by environmental objects. Thus at the receiver, various time delayed copies of the original symbol are received and there is a self-inflicted interference. A common way of modeling this phenomenon for narrowband flat fading channels is the Rayleigh fading model given by the following equation

$$y[m] = h[m]x[m] + n[m] \tag{1.2}$$

where the transmit symbol is now multiplied by the fading coefficient $h[m]$. In a Rayleigh fading channel, $h[m]$ is a complex circularly symmetric random variable that models the effects of the environment. The distribution of the random variable comes arises from the central limit theorem. Fading channels are much more representative of channels encountered in reality and has adverse effects on the system performance. Under the Rayleigh fading model, the probability of error decays as $\frac{1}{SNR}$ rather than the exponential rate of decay of the AWGN channel. One might imagine that performance can be improved if $h[m]$ is learned via pilot symbols or some other mechanism. Unfortunately knowledge of $h[m]$ only improves performance by 3 db and the error probability still decays as the inverse of SNR. Such poor performance compared to the AWGN channel comes about because there are now two sources of error. An error can occur when the noise term is large, or when the norm of the multiplicative fading term is small. The probability of the later event dominates and leads to the inverse scaling.

MIMO systems can help combat fading channels and improve system throughput because these systems provide additional spatial diversity and degrees of freedom due to the multiple transmit and/or receive antennas. The trade-off between exploiting the spatial diversity and the degrees of freedom is rather complicated, but a simple example will now be shown that demonstrates how additional receive antennas can help reduce the probability of error. Let there be a single transmit antenna and N receive antennas in a point-to-point system experiencing a Rayleigh fading channel. The channel model for the i^{th} receive antenna follows from Equation 1.2

$$y_i[m] = h_i[m]x[m] + n_i[m] \tag{1.3}$$

and the fading coefficients are independent for each $i \in \{1, \ldots, N\}$. Using a maximum ratio combining (MRC) receive technique, the probability of error now scales as $\frac{1}{SNR^N}$, where the exponent of N is due to the receive spatial diversity. This simple example illustrates just one of the benefits of MIMO communication systems.

1.1.2 Multi-User MIMO Broadcast Channel Multi-User Diversity

The channel that is considered in this dissertation is the multi-user MIMO broadcast channel. In this channel, there is a transmitter, equipped with multiple transmit antennas, that sends messages to n users, each of which may be equipped with multiple receive antennas. The channel between each transmit-receive antenna pair for every user is assumed to follow an i.i.d Rayleigh fading model according to Equation 1.2. From an information theoretic point of view the n-dimensional capacity region is of fundamental interest in characterizing the maximum potential of the channel. The works [WSS06],[YC04] were able to characterize this region. An important point on the outer surface of the capacity region had been characterized earlier in [CS03]. The point analyzed was the sum-rate capacity, which is the point where the vector-valued function $f(\mathbf{x}) = \mathbf{x}$ intersects the boundary of the capacity region. The sum-rate capacity is an interesting and important metric for characterizing the maximum potential of a channel because it measures the maximum total data rate that is capable of being transmitted. Any other point on the capacity region may give a particular user a higher average rate, but it comes at the expense of reducing the sum total of all the rates among the users.

The theoretical capacity region is useful as a benchmark for any communication system since the capacity region is the theoretical maximum possible performance given the channel model. The problem with the proofs of the theoretical capacity and sum-rate capacity expressions is that they are often non-constructive. The proof of the maximum sum-rate point uses the now well known dirty paper coding technique, which is non-causal and computational expensive. The difficulties of implementing the techniques used in the proofs has led to much research dedicated to finding techniques that are computationally more feasible yet offer good performance The sub-optimality of these techniques is traded off with the ease of implementation.

One of the metrics that the suboptimal techniques attempt to match is the sum-rate scaling rate as a function of the number of users in the broadcast channel. Assume that the transmitter in the broadcast channel has M antennas. The number of transmit antennas limits the number of independent data streams that can be multiplexed simultaneously on any given time-frequency resource. This is to say that when there are M transmit antennas, at most M independent users can be simultaneously transmitted to. When there are $n > M$ users in the broadcast channel, a new form of diversity arises: multi-user diversity. If there are more users than transmit antennas, by scheduling any subset of M independent users, one would expect there to be roughly M times the throughput as opposed to a single user system. This is indeed the case, but higher performance can be achieved by carefully selection of the M users to schedule. The sum-rate analysis work performed in [CS03] shows that as a function of the number of users n, the optimal sum-rate of the broadcast channel scales as $M \log \log n$. This result says that in addition to the throughput gain due to multiple users and multiple transmit antennas, that the sum-rate grows as a function of the number of users. This growth due to the number of users is due to the multi-user diversity. When there are many users, some users will experience high quality channels with high probability. By transmitting to these users the sum-rate performance of the system is increased. It is clear that to schedule the users experiencing the best channel conditions requires knowledge of these channel conditions at the transmitter. The feedback of this channel state information (CSI) is essential in order to exploit the multi-user diversity and benefit from the growth rate of $\log \log n$. Transmit schemes of low computational complexity that can achieve the $\log \log n$ sum-rate growth are asymptotically nearly optimal yet are easily implementable. The random beamforming transmit scheme, the transmit scheme which is the basis of this dissertation, is one such scheme.

1.1.3 Random Beamforming

In the broadcast channel, as mentioned in Section 1.1.2, when the transmitter has M transmit antennas, up to M independent users can be scheduled at any given scheduling epoch. A central question is how to select which users to schedule. To exploit the multi-user diversity, this question then becomes one of how does one characterize a "good" channel. Saying a channel is "good" is more complicated than the point-to-point communication channel where everything can be characterized by SNR. Since the transmitter is sending to M different users, the messages interfere with each other and thus can decrease the SINR. Thus, scheduling M quality users becomes not only a problem of finding users with high SNRs, but making sure the scheduled users do not interfere with each other significantly. The random beamforming technique proposed in [SH05] confronts these issues.

The random beamforming technique assumes a Rayleigh block fading channel model. If there are M transmit antennas at the transmitter, then the transmitter randomly generates an M dimensional orthonormal basis in $\mathbb{C}^M$ drawn uniformly from some isotropic distribution. Each of the random basis vectors will act as a beamforming vector for one of the M messages that is to be sent during each scheduling interval. It is assumed that each user in the broadcast channel knows the randomly generated basis, whether the transmitter notifies them of this basis via a message or by common randomness shared between the users and the transmitter. Once the random basis is generated and known at each user, a pilot signal is sent on each transmit beamforming vector, the basis vectors, and the SINR is measured at each user. Notice that because the beamforming vectors are known, the SINR is capable of being measured. The SINR is a scalar value that represents the quality of the channel in each beamforming direction and is the essential CSI required to exploit multi-user diversity. If the transmitter is notified of the SINR values measured at each user, then to maximize the sum-rate throughput the scheduling algorithm selects the user with the highest SINR on each of the transmit beamforming vectors. It was shown in [SH05] that this random beamforming transmit scheme with greedy scheduling under Rayleigh fading achieves a sum-rate scaling that grows as $\log \log n$, and thus asymptotically fully exploits the multi-user diversity. Based on this very promising result, this dissertation focuses on how to reduce the amount of CSI, or in this case the number of SINR values. Additionally, the greedy algorithm under random beamforming and a Rayleigh fading model cannot guarantee fairness on any finite time frame so the proportional fair scheduling algorithm is considered as a replacement for the greedy scheduling algorithm and the performance is analyzed.

1.2 Contributions of the Thesis

The main results of this dissertation are contained in three parts. The first part considers feedback reduction in multi-user MIMO broadcast channels by exploiting spatial diversity. The second section utilizes thresholding mechanisms to reduce feedback. The final portion considers the fairness of the random beamforming technique. The three parts are summarized below.

1.2.1 Feedback Reduction Techniques Exploiting Spatial Diversity

For the most part, the work done in [SH05] focuses on when each user has a single receive antenna. If there are M transmit antennas at the transmitter, then each user makes M SINR measurements, one for each random transmit beamforming vector. Each of these M values is fed back to the transmitter, and if there are n users in the system, the total number of SINR values fed back is nM. A question addressed by this dissertation is what happens when each user, rather than having a single receive antenna, has $N \leq M$ receive antennas. Two approaches are taken to this case. The first considers each antenna individually. Thus at each user, their are NM SINR values to be measured, one for each antenna and each transmit beam. The second approach utilizes the fact that the N receive antennas are co-located at a user and thus combining can be performed. A linear minimum mean squared error (LMMSE) filter, which is the optimal linear receive filter, is applied at each user. The application of the LMMSE receive filter results in M values at the output of the filter, one for each transmit direction. These two approaches leverage the fact that there are additional receive antennas and thus more possible measurements and diversity. A question to be ask however is if all the measurements are required at the transmitter to exploit the multi-user diversity. This question is addressed by this work.

Individual Antenna Perspective

In a system where the transmitter has M antennas and each user has $N \leq M$ receive antennas, then as previously mentioned there are NM SINR measurements at each user. In the work [SH05], it is pointed out that not all NM values must be fed back to the transmitter to exploit the multi-user diversity. This result is based on the observation that only the largest SINR for each transmit beam at each user need be fed back. For each transmit vector, a user makes N SINR measurements, one for each receive antenna. Only the maximum of these N values need back fed back because the other $N-1$ values are not even the maximum locally at the particular user, so it is impossible to be the maximum SINR in the entire system. The scheduling algorithm at the transmitter is greedy, and thus these $N-1$ values becomes inconsequential. Therefore, as opposed to a total feedback load of nNM SINR values if each measurement was fed back, only nM values are fed back, M per user, and achieves the same performance as if every value was known at the transmitter. Now the question is if each user is required to feed back even M values.

Each user has NM SINR measurements and a method was just described such that the maximum SINR value over the receive antennas is fed back. In an effort to further reduce the amount of SINR fed back, what happens if only the maximum over the M maximums is fed back? That is to say, consider feeding back only the single largest SINR value among the NM SINR values and the transmit beam index associated with that value. This problem proves to be much more difficult than the previous feed back scheme. The primary reason is that when taking the maximum over transmit beams, the SINR random variables become correlated. The first step in solving this problem is to characterize the distribution of the maximum value over both transmit vectors and receive antennas. After this is done, the asymptotic sum-rate scaling rate must be found under this new distribution. Fortunately, the sum-rate scales as $\log \log n$, the theoretical optimum. Therefore, if rather than feeding back M values, if each user only feeds back the single largest measured SINR value and the associated beam index, the same asymptotic scaling rate can be achieved as full SINR feedback. Under this modified scheme, only n SINR values are fed back, so the amount of total feedback has been reduced by a factor of M as compared to the previous method. The details of this analysis can be found in Chapter 2.4.

LMMSE Receivers

In the previous Section 1.2.1, each antenna provided an additionally SINR observation for each user. The additional receive antenna can also increase system performance by allowing for more complicated receive architectures. One such architecture is the LMMSE receiver. At the output of the LMMSE receiver, there are M post-processed SINR values, one for each transmit beamforming vector. One would imagine that by implementing an LMMSE receiver at each user, and feeding back the resulting M SINR values, that the scaling rate of $\log \log n$ is achieved. This result however depends on the distribution of the SINR values that are fed back to the transmitter, and thus to determine the scaling rate of the system employing LMMSE receivers requires knowing the distribution of the post-processed SINR. This distribution is found and shown to exhibit the desired $\log \log n$ scaling. This scheme has a total feedback overhead of nM SINR values.

Akin to the individual antenna perspective, rather than feeding back the M post-processed SINRs, what happens if only the maximum of these M values is fed back along with the corresponding transmit beam index? Unfortunately, the closed form distribution of the maximum of the M post-processed SINR values could not be found due to the correlated nature of these random variables. To characterize the scaling rate of this reduced feedback scheme where only the largest post-processed SINR value is fed back, sufficiently tight bounds are found for the distribution function of the maximum of the M values. It is then shown that both the lower and upper bounds exhibit the $\log \log n$ scaling rate, and thus the true, but unknown, distribution of the maximum exhibits the same scaling rate. Therefore rather than feeding back all M post-processed SINR values, one can feed back only the largest among the M and the corresponding beam index and still achieve the same asymptotic sum-rate scaling rate. Chapter 2.5 contains the details for these results.

Finite Thresholds

To reduce the amount of feedback even further, a thresholding mechanism is proposed. The idea is that a threshold is chosen such that if the SINR value that a user would feed back falls below the threshold, then the SINR value is not fed back. Because the scheduling algorithm is known to be greedy, such a mechanism seems feasible because if the threshold is chosen properly, the feedback of SINR values below the threshold are eliminated and these values hopefully had a small probability of being the largest SINR values in the system. Thus the goal of the thresholding mechanism is to reduce feedback by the elimination of the feedback of small SINR values that would not be scheduled even had they been provided to the transmitter, and therefore not affecting the system performance. The draw back of the thresholding mechanism is that it is always possible that for any chosen threshold, that no SINR value exceeds it, at which point no SINR values are fed back to the transmitter. In such a scenario, the transmitter has no CSI, and thus the multi-user diversity cannot be extracted.

It is shown that for any finite threshold value T, that as the number of users asymptotically goes to infinity, that the rate lost due to the threshold goes to zero. Thus asymptotically, for any finite threshold, there is no difference between the thresholded system and the system without a threshold in terms of sum-rate. This result basically follows from the fact that for the maximum of n random variables that have unbounded support from above, as n grows to infinity, the maximum becomes arbitrarily large and will exceed any fixed finite threshold. Thus a fixed finite threshold is useful for reducing the total feedback load of the system while not comprising the asymptotic properties of the system. Chapter 2.6 contains these results.

1.2.2 Feedback Reduction Techniques Using Thresholds

In Section 1.2.1 it is described how any fixed finite threshold does not affect the asymptotic performance of the system. While this result is interesting, it does not provide insight into how an effective threshold can be designed. A meaningful threshold should be a function of the number of users in the system rather than some a priori fixed value. The reason for this is that as the number of users in the system increases, the threshold should be increased. To see why this is the case, when there are more users in the system, then there is an increased chance that at least one user is experiencing very good channel conditions, or in our case a large SINR value, and thus the threshold can be increased to reduce the amount of small SINR values that are fed back. Conversely, when there are not many users in the system, the threshold should be small since there is a significant probability that no user experiences a large SINR. When no user exceeds the threshold, no SINR is fed back to the transmitter and all multi-user diversity is lost. Two approaches are considered in this dissertation for the appropriate design of thresholds. The first class of results considers the asymptotic performance of any threshold as a function of the number of users. The second set of results proposes certain metrics and optimizes the thresholds accordingly.

Asymptotic Properties of Thresholds

If a threshold is a function of the number of users, how fast can the threshold grow? That is, intuitively it is argued that as the number of users in the system increases, so too should the threshold. It is possible however that if the threshold grows too fast as a function of the number of users, that eventually no user will exceed the threshold and the multi-user diversity cannot be utilized. This dissertation provides the answer to the appropriate scaling rate of any possible threshold such that multi-user diversity is not lost. More precisely, let $T(n)$ be a threshold which is a function of the number of users. It is shown that if $T(n) \in o(\log n)$ that multi-user diversity is still achieved and the sum-rate of the system still asymptotically grows as $\log \log n$.

The aforementioned result provides a sufficient condition for the scaling rate of any possible threshold as a function of the number of users. A natural question that follows is if it is possible for a thresholding function $T(n)$ to grow faster than $\log n$ and still achieve the $\log \log n$ sum-rate scaling. To address this question, the converse to the previous result is found. Namely it is shown for any thresholding function $T(n) \in \omega(\log n)$, asymptotically no user in the system will exceed the threshold almost surely, and thus asymptotically no users in the system provide feedback and all multi-user diversity is lost. Therefore $\log n$ is a kind of boundary on how fast the threshold can scale and still have the system achieve the optimal sum-rate scaling rate. These results are contained in Chapter 3.4.

Thresholds for a Finite Number of Users

The results concerning the growth rate of possible successful thresholding functions in Section 1.2.2 is a result of theoretical concerns, but how should the threshold be chosen when the system designer knows there are a finite number of users n in the system? The answer to this question depends on what the system designer deems most important in terms of system performance and what can be sacrificed. Obviously a system in which there is no threshold at all will yield superior performance since full CSI is known at the transmitter, but the sacrifice is that feedback load is maximum; costing uplink bandwidth and transmit power. To reduce the feedback overhead, three metrics are proposed and the thresholds are optimized under these metrics.

The first metric is to constrain the probability that no user feeds back for any transmit beam to be less than some design parameter γ_{outage}. When no user feeds back for a particular transmit beamforming vector, then multi-user diversity is lost, and this event is called an outage event. Therefore the challenge is to design the threshold $T(n)$ such that when there are n users in the system, the probability that no user exceeds $T(n)$ for any transmit beam is less than γ_{outage}. The threshold in this case if found as a function of n and γ_{outage}. The second proposed metric constrains the average amount of feedback in the system. Namely, the goal is to design the threshold such that on average only k of the n users in the system feed back their SINR information for each transmit beam. Lastly, the third metric constrains the difference between the rate of the thresholded and unthresholded systems. Any non-zero threshold will cause the system to incur some rate loss because there is always the probability that no user will experience an SINR larger than the threshold at any given scheduling epoch. Let R_{loss} be a design parameter such that the difference between the rates of the thresholded and unthresholded system is less than R_{loss}. Parameterized by n and R_{loss}, the optimal threshold is found. Chapter 3.5 provides the detailed analysis of these results.

1.2.3 Proportional Fair Sharing and Fairness

The random beamforming technique uses a greedy scheduling algorithm to select the users experiencing the best SINR for each transmit beamforming vector. Due to the i.i.d. nature of the Rayleigh fading channel no user is preferred over time. Thus, each user has an equal chance of being the maximum which will be scheduled and thus asymptotically each user will be scheduled an equal portion of the time and will asymptotically experience the same average rate. The asymptotic fairness hides the fact that the greedy scheduling algorithm may produce periods where a specific user is not scheduled for an arbitrarily long period of time. To confront this issue, this dissertation considers implementing a proportional fair sharing algorithm in place of the greedy scheduling algorithm.

Under the proportional fair sharing algorithm, the transmitter schedules the user with the best ratio of current rate to average rate for each transmit beam instead of just the user with the largest SINR. More precisely, let $R_{i,n}$ be the current rate of user i at the n^{th} scheduling epoch (which is a function, in the random beamforming scheme, of the SINR) and $1_{i,n}$ be an indicator random variable denoting whether user i was scheduled at scheduling period n. The average rate scheduled for user i up to scheduling period n is given by $\sum_{k=0}^{n-1} 1_{i,k} R_{i,k}$. The PFS algorithm thus schedules the user with the largest ratio of $R_{i,n}$ to $\sum_{k=0}^{n-1} 1_{i,k} R_{i,k}$. This algorithm promotes fairness by increasing the chance that users who have small averages rates are scheduled.

Generally there is a trade off between fairness and throughput. The greedy algorithm with random beamforming maximizes the rate at each scheduling period, but does not consider fairness at all, and thus for any finite time window, is not fair. The PFS algorithm on the other hand sacrifices some rate to increase the average throughput of users that have small average rates. The central question is how much rate is lost due to trying to be fair. The main result shown in this dissertation is that when the rates are drawn from an i.i.d. distribution (which they are in the random beamforming scheme under i.i.d. Rayleigh fading), asymptotically the average rate experienced by each user under the PFS algorithm asymptotically equals that of the greedy algorithm. That is to say that in the appropriate state-space, the two algorithms converge to the same point. These results are detailed in Chapter 4.4 and 4.5.

The equilibrium point in this state space is identified. The rate of convergence of the PFS algorithm to this equilibrium point is derived in Chapter 4.6.1. At any given time, the distance of the PFS algorithm to the equilibrium point is a random variable. The asymptotic distribution of this distance random variable is a normal random variable and the covariance matrix is found in Chapter 4.6.2. Additionally, taking a cue from the previous section, application of a threshold is considered under the PFS algorithm. A priori it is not clear how a threshold affects the equilibrium point of the PFS algorithm. Convergence under the proposed thresholding scheme is proven, and the new equilibrium point is found in Chapter 4.7.

Chapter 2

Reduced Feedback Schemes using Random Beamforming in MIMO Broadcast Channels

2.1 Outline

A random beamforming scheme first proposed in [SH05] for the Gaussian MIMO broadcast channel with channel quality feedback is investigated and extended. Considering the case where the n receivers each have N receive antennas, the effects of feeding back various amounts of signal-to-interference-plus-noise ratio (SINR) information are analyzed. Using the results from order statistics of the ratio of a linear combination of exponential random variables, the distribution function of the maximum order statistic of the SINR observed at the receiver is found. The analysis from viewing each antenna as an individual user is extended to allow combining at the receivers, where it is known that the linear MMSE combiner is the optimal linear receiver. Using the results in [GS98b] and [GS98a], the CDF for the SINR after optimal combining is derived. Analytically, using the Delta Method, the asymptotic distribution of the maximum order statistic of the SINR with and without combining is shown to be, in the nomenclature of extreme order statistics, of type 3. Applying the main result in [SH05], the throughput of the feedback schemes are shown to exhibit optimal scaling asymptotically in the number of users. Finally, to further reduce the amount of feedback, a hard threshold is applied to the SINR feedback. The amount of feedback saved by implementing a hard threshold is determined and the effect on the system throughput is analyzed and bounded. This chapter is a reprint of a paper coauthored with Bhaskar D. Rao and appears in the March 2010 issue of *IEEE Transactions on Signal Processing* under the title *"Reduced Feedback Schemes Using Random Beamforming in MIMO Broadcast Channels"*. Sections of this chapter also appear in *"On the Capacity of MIMO Broadcast Channels with Reduced Feedback by Antenna Selection"* in *Asilomar Conference on Signals Systems, and Computers, March 2008* and *"Feedback reduction in MIMO broadcast channels with LMMSE receivers"* in *ICASSP 2010*.

2.2 Introduction

The users in a broadcast channel experience varying levels of channel quality. For high throughput, it is useful for the transmitter to be fully aware of the channel to all the users. This represents a large amount of information that must be known at the transmitter, especially if the number of users n is large. A mechanism by which the transmitter is aware of the channel state information is for each user to feed back the observed channel. To send back the observed channel may impose unreasonable complexity, so schemes wherein the users send back partial channel information, but still realize the major benefits of multiuser diversity, are of interest. The questions that concerns this chapter are how to reduce the amount of feedback, the tradeoff between reduced feedback and performance, and what are the benefits of having multiple receive antennas. In this chapter, the random beamforming scheme suggested in [SH05] is considered. The random beamforming scheme has the transmitter with M transmit antennas produce M random orthonormal beams and send the messages on these beams. Because the throughput is a function of SINR, if each user sends back the SINR it experiences, then the transmitter transmits to the users that are currently experiencing the best channels for each beam. If each of the n users has N receive antennas, each receive antenna of each user can be considered as an individual user and the SINR values are fed back as such. In this case, the SINR measured at each antenna for each beam results in MN SINR values to be fed back per user. Although MN SINR values are measured at each user, if the maximum per beam is sent back, namely M values, the same performance can be achieved with reduced feedback.

Other novel techniques that utilize random beamforming have since been proposed. In [IMKT05], random weight vectors are used during a training period where the users feed back the required information so that the transmitter can choose the optimal random weight vectors and users for the current transmission. This concept is extended in [WCH07] where training is done on a set of random orthonormal bases and based upon the feedback, the transmitter will select the best beamforming vectors and users. The question of how many random beams to use based upon the available multiuser diversity is addressed in [JWZ08]. While each of these techniques extend the methodology of random beamforming, the focus of this chapter is to examine random beamforming schemes with minimal feedback. As such, only a single random orthonormal basis will be considered per fading block.

In this chapter, a feedback scheme is proposed where each user sends back the maximum SINR experienced across all the receive antennas and across all the transmitted beams. This reduces the number of SINR values fed back to the transmitter further from M to 1 per user. Even with this significant reduction in feedback, it is shown that asymptotically in the number of users, the distribution of the maximum SINR is of type 3 (see [Dav70], [Gal78]), the same type as that in [SH05]. Ultimately, using the methodology developed in [SH05], it is shown this reduced feedback scheme also exhibits the same asymptotic scaling as the original feedback scheme.

Next we consider an enhancement to the above schemes. In the above mentioned feedback schemes, the receive antennas are making individual SINR measurements. The availability of multiple receive antennas allows combining to be performed at the receiver to increase the received SINR. Transmit random beamforming is utilized in [JCK03] with each user performing receive beamforming using the left singular vectors of the channel matrix. The effective SNR of each user is fed back to the transmitter which then selects one user to transmit to based on a proportional fair scheduler while using waterfilling for power control. This scheme will not be pursued as it is well known ([TV05] for example) that the optimal linear receiver is the linear MMSE receiver and using a random orthonormal basis allows parallel transmission to multiple users. Feeding back the M SINR values per user after optimal combining increases the throughput of the system for the same amount of feedback. The amount of feedback is further reduced when only the maximum SINR seen across beams after optimal combining is sent back, i.e. the maximum of those M values. This reduces the amount of feedback to a single SINR value per user and the associated beam index. The behavior of such a scheme is considered and evaluated in the chapter. The distribution of the SINR values after optimal combining is derived and asymptotically the distribution of the maximum is shown to be of type 3 and to have optimal throughput scaling. Similar analysis of optimal combining in an interference limited regime is performed in [MPP07]. This work was then extended in [MPP08] to the general SINR setting, where for the the specific case of $M = 4$ and $N = 2$, they derive the SINR distribution and scaling laws for the LMMSE receiver.

As the number of users in the system grows, the amount of feedback also grows. In an effort to further reduce the amount of feedback, a thresholding mechanism can be employed in conjunction with the previously proposed schemes. If the maximum SINR value that was to be sent back to the receiver is under the threshold, then the SINR value is not sent back. Consequently, not every user provides feedback. This scheme is also considered and it is shown that for any finite thresholding value, the loss due to thresholding asymptotically goes to zero in the number of users.

This chapter extends the work in [PR08] in several ways, most notably by considering optimal combining, thresholding and finding a closed form expression for the distribution of the maximum SINR observed at a user when SINR measurements are taken at the antenna level.

The organization of the chapter is as follows: the system model used to analyze the problem and previous work conducted in [SH05] is discussed in Section II. Section III considers a new scheme where only the maximum SINR per user is fed back. In the analysis of the problem, the distribution of the maximum SINR per receive antenna is derived. The asymptotic performance is also analyzed with the help of the Delta Method. Section IV extends the results of Section III to the case where each user implements an LMMSE receiver. Section V investigates the effects of thresholding the SINR feedback on the total amount of feedback in the system as well as on system throughput. Section VI summarizes the results of this chapter.

2.3 System Model and Background

In this section the system model used to analyze the problem and previous work found in [SH05] are discussed. The results found in [SH05] will be built upon in the latter parts of this chapter.

2.3.1 System Model

A block fading channel model is assumed for the Gaussian broadcast channel. The transmitter has M transmit antennas and there are n receivers, each with N receive antennas. It is further assumed that $n \gg M$ and $N \leq M$, a reasonable assumption, for example, in a cellular system. Let $\mathbf{s}(t) \in \mathbb{C}^{M\times 1}$ be the transmitted vector of symbols at time slot t and $\mathbf{y}_i(t) \in \mathbb{C}^{N\times 1}$ be the received symbols by the i^{th} user at time slot t. The following model is used for the input-output relationship between the transmitter and the i^{th} user:

$$\mathbf{y}_i(t) = \sqrt{\rho_i} H_i \mathbf{s}(t) + \mathbf{w}_i(t), \quad i = 1, \ldots, n. \tag{2.1}$$

$H_i \in \mathbb{C}^{N\times M}$ is the complex channel matrix which is assumed to be known at the receiver, $\mathbf{w}_i \in \mathbb{C}^{N\times 1}$ is the white additive noise, and the elements of H_i and $\mathbf{w}_i$ are i.i.d. complex Gaussians with zero mean and unit variance (as defined by Edelman in [Ede89]). The transmit power is chosen to be M, i.e. $E\{\mathbf{s}^*\mathbf{s}\} = M$, the SNR at the receiver is $E\{\rho_i|H_i\mathbf{s}|^2\} = M\rho_i$ and ρ_i is the SNR of the i^{th} user. It is assumed that $\rho_i = \rho \ \forall i$.

2.3.2 Background

The random beamforming scheme developed in [SH05] forms the basis of this chapter. The key elements of this work are now described.

SINR Distribution

The transmission scheme, as developed in [SH05], involves generating M random orthonormal vectors ϕ_m $(M \times 1)$ for $m = 1, \ldots, M$, where the basis is drawn from an isotropic distribution. Let $s_m(t)$ be the m^{th} transmit symbol at time t, then the total transmit signal at time slot t is given by

$$\mathbf{s}(t) = \sum_{m=1}^{M} \phi_m(t) s_m(t). \tag{2.2}$$

The received signal at the i^{th} user is given by

$$\mathbf{y}_i(t) = \sum_{m=1}^{M} \sqrt{\rho} H_i(t) \phi_m(t) s_m(t) + \mathbf{w}_i(t). \tag{2.3}$$

Each receive antenna at each user is assumed to measure the SINR for each of the M transmitted beams and the maximum of the observed SINR values is fed back leading to the use of order statistics.

Assuming that the i^{th} user knows the quantity $H_i(t)\phi_m(t)$ from Equation (3) for all m, the SINR of the j^{th} receive antenna of the i^{th} user for the m^{th} transmit beam is computed by the following equation:

$$SINR_{i,j,m} = \frac{|H_{i,j}(t)\phi_m(t)|^2}{\frac{2}{\rho} + \sum_{k \neq m} |H_{i,j}(t)\phi_k(t)|^2}. \tag{2.4}$$

$H_{i,j}$ is the j^{th} row of the i^{th} user's channel matrix. Because the beamforming vectors are orthonormal and the entries of H_i are i.i.d. complex Gaussian with zero mean and unit variance, the numerator in Equation (2.4) is distributed as a $\chi^2(2)$ random variable and the denominator as an independent $\chi^2(2(M-1))$ random variable. The density of the SINR is given in [SH05]

$$f_s(x) = \frac{e^{-\frac{x}{\rho}}}{(1+x)^M}\left(\frac{1}{\rho}(1+x)+M-1\right)u(x) \tag{2.5}$$

From now on, for notational simplicity, the $u(x)$ will dropped from the distribution and density expressions with the understanding that all the random variables of interest are nonnegative. To find the distribution of the maximum SINR, the distribution function must be known and is given by the integration of the density in Equation (2.5) and is shown in [SH05] to be

$$F_s(x) = 1 - \frac{e^{-\frac{x}{\rho}}}{(1+x)^{M-1}}. \tag{2.6}$$

Scheduling Scheme

With the cdf of the SINR observed at each receive antenna known, one naive feedback scheme is for each antenna to send back all the SINR values measured for each transmit beam. This results in a total of nNM SINR values sent back to the transmitter, which then transmits to the antennas with the largest SINR for each transmit beam. Viewing each antenna as a separate user, the following approximation is shown in [SH05]:

$$R \approx E\left\{\sum_{m=1}^{M}\log\left(1+\max_{i=1,\ldots,nN} SINR_{i,m}\right)\right\} = ME\left\{\log\left(1+\max_{i=1,\ldots,nN} SINR_{i,m}\right)\right\} \tag{2.7}$$

The approximation sign is required because there is a small probability that the same antenna is the best for multiple transmit beams and each beam can only be used to serve one user. To find the approximate rate requires the use of order statistics.

The distribution of the maximum SINR for a given beam is given by classical order statistics ([Dav70], [Gal78]) and is

$$F_{max}(x) = \left[F_s(x)\right]^{nN}, \tag{2.8}$$

where $F_s(x)$ is the CDF of the SINR given in Equation (2.6). This equation can be used because the SINR values across antennas for a specific transmit beam are independent and marginally are identically distributed. Feeding back only the largest SINR value for each transmit beam per user reduces the system feedback to nM SINR values and is considered in [SH05]. Each SINR value fed back to the transmitter is distributed according to $[F_s(x)]^N$ since the maximum was taken over the N receive antennas prior to feedback. The transmitter takes the maximum over the n received SINR values for each beam resulting in maximum order statistic distributed according to $[F_s(x)]^{nN}$, which is identical to the distribution had every measured SINR value been fed back. Therefore, there is no benefit to feeding back all observed SINR values. For future reference, let us refer to this reduced feedback scheme of [SH05] as **Scheme A**.

2.4 Feeding Back the Largest SINR per User

Each user in **Scheme A** is feeding back the M SINR values associated with the largest SINR measured for each beam. To further reduce the amount of feedback, what happens if each user only feeds back the largest SINR value observed over all receive antennas and transmit beams as well as the associated beam index? For example, in Figure 2.1, when there are four transmit beams and 3 receive antennas per user, the quantity of interest is $\max_{i,j} SINR_{i,j}$, the maximum SINR element in the grid. This scheme will reduce the total amount of system feedback from nM SINR values to n SINR values and the corresponding beam indices. This feedback scheme is referred to as **Scheme B**. Case 2 of Section VI in [SH05] proposes a method where at most one beam is assigned to each user and the SINR metric for user i is given by

$$SINR_{i,j} = \frac{\phi_j^* H_i^* H_i \phi_j}{\frac{1}{\rho} + \sum_{k \neq j} \phi_k^* H_i^* H_i \phi_k},$$

which is the combined energy of transmit beam j over the sum of the inverse of the SNR and the sum of the energy from the other transmit beams. In **Scheme B**, the SINR is viewed at the antenna level for each beam while in Case 2 of [SH05], combining has been performed such that there is one SINR value per transmit beam. The analysis of optimal combining schemes will come in Section IV of

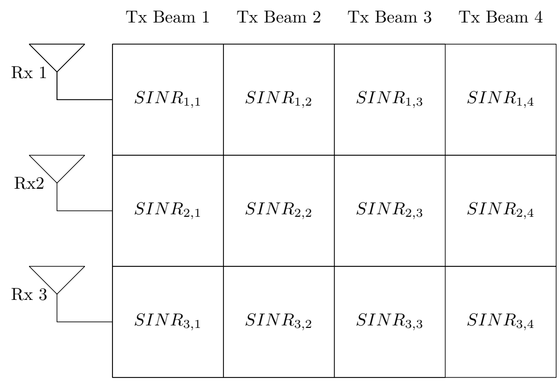

Figure 2.1: SINR Observations at a single user for M=4, N = 3

this chapter, while the current focus is on **Scheme B**. The analysis of this scheme poses some difficulties that do not arise when considering **Scheme A**. This section addresses the new difficulties that arise and then, in the same vein as the work of [SH05], the asymptotic performance of **Scheme B** is analyzed.

2.4.1 Analysis of Scheme B

Although **Scheme B** reduces the amount of feedback by a factor of M compared to **Scheme A**, it is suboptimal. This is due to the fact that the best SINR values for a particular beam may not be fed back due to the restriction that only one value can be sent back per user. This restriction, however, removes the mechanism which led to the approximation symbol in Equation (2.7). Although **Scheme B** may be suboptimal, in the interest of reducing the feedback as much as possible, this scheme is considered. Later in this section the asymptotic performance of the scheme in the number of users will be analyzed and will be shown to have optimal scaling properties.

The distribution of the SINR that is served by the transmitter is fundamentally different for **Scheme B** than **Scheme A** for two reasons. The first difference is that in **Scheme A**, the maximum is taken for each beam and the marginal distributions of the SINR were i.i.d. across receive antennas. However, the SINR values at a particular receive antenna for different transmit beams are coupled. Looking again at Figure 2.1, the SINR values in a given column are i.i.d., where as the SINR values in a given row are marginally identically distributed but are not independent. To see this, fix the antenna at a particular user and vary the beam index. Let $|H_{i,j}(t)\phi_m(t)|^2 = X_m$ and $c = \frac{2}{\rho}$. Then the SINR values at a receive antenna are given by $\frac{X_1}{c+\sum_{k\neq 1} X_k}, \ldots, \frac{X_M}{c+\sum_{k\neq M} X_k}$. The SINRs are coupled by the appearance of the numerator term of a particular SINR value appearing in the denominator of all the other SINR values. Because the SINR values are not independent, the order statistics used earlier cannot be applied. The second difference between the two schemes is that in **Scheme B** since each user feeds back the largest observed SINR over all receive antennas and transmit beams, the number of SINR values to maximize over at the transmitter for each beam is a random quantity. In terms of the distribution of the maximum order statistics, this changes the exponent reflecting the number of variables being maximized over. These fundamental differences will now be addressed.

Distribution of the Maximum SINR per User

If the distribution of the largest SINR value at a particular user for a fixed receive antenna can be found, then the maximum SINR per user can be found because the SINR random variables are independent across receive antennas. Using the notation defined earlier, let $|H_{i,j}(t)\phi_m(t)|^2 = X_m$ and $c = \frac{2}{\rho}$. Then for a fixed receive antenna, the SINR for each transmit beam is given by

$$SINR_1 = \frac{X_1}{c + \sum_{i\neq 1}^{M} X_i}, \ldots, SINR_M = \frac{X_M}{c + \sum_{i\neq M}^{M} X_i}$$

The distribution of interest is the maximum of these M SINR values. The key observation is that if the X_is are ordered, i.e. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(M)}$, then the maximum SINR for a fixed receive antenna is given by

$$SINR_{(M)} = \frac{X_{(M)}}{c + \sum_{i=1}^{M-1} X_{(i)}} \tag{2.9}$$

This makes intuitive sense, since to maximize the SINR, the largest signal power should be put in the numerator and all the other signal powers should be considered as interference and put in the denominator. To find the distribution of the quantity in Equation (2.9), results based on the ratio of the linear combination of order statistics are called upon. As mentioned previously, the X_is are distributed as a χ_2^2 random variable, which is equivalent to a exponential-$\left(\frac{1}{2}\right)$ random variable. The ratio of the linear combination of order statistics drawn from an exponential distribution have been studied in [Dav70], [MMA82]. After some manipulation to get the χ_2^2 random variables in the proper exponential form, the distribution function of the maximum SINR for a particular beam index is given by

$$\Pr\left(SINR_{(M)} \leq x\right) = F_{SINR_{(M)}}(x) = 1 - \sum_{i=1}^{M} \frac{[d_i(x)]_+^M \exp\left(-\frac{xc}{d_i(x)}\right)}{A_i(x)} \qquad (2.10)$$

where $d_i(x) = \frac{2[1-x(M-i)]}{M-i+1}$, $A_i(x) = d_i(x)\prod_{j\neq i}^{M}\left(d_i(x) - d_j(x)\right)$, and $[\cdot]_+$ is the positive part of the argument.

Equation (2.10) gives the distribution of the maximum SINR over the beams for a particular receive antenna at a specific user. **Scheme B** feeds back the largest SINR per user, so the maximum has to be taken over receive antennas as well. As with **Scheme A**, the random variables across receive antennas are independent, thus the distribution of the maximum SINR per user is given by $\left[F_{SINR_{(M)}}(x)\right]^N$. Figure 2.2 verifies the distribution of the maximum SINR measured at a user given by Equation (2.10) raised to the N^{th} power by comparing it with Monte Carlo simulation and it is observed that the theoretical distribution matches the simulation results exactly.

Order Statistics Over a Random Number of Observations

If each user feeds back the maximum observed SINR and the beam index that produced it, then the transmitter receives n SINR values and n beam indices over which to maximize. The number of SINR values to maximize over at the transmitter for a particular beam is, however, a random number.

For example, in Figure 2.3, there are nine users and at the transmitter, the number of users feeding back a particular beam index is random. In this case three users feed back beam index 1, two users feed back beam index 2, and so on. This effect is not taken into consideration in Case 2, Section VI of [SH05]. There it is mentioned that only the maximum SINR (which differs from the SINR

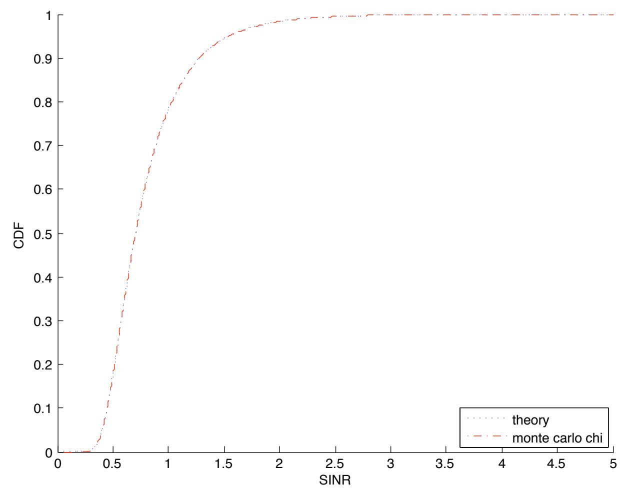

Figure 2.2: Monte Carlo simulated Maximum SINR per User versus Theoretical Distribution for M = 5, N = 2, SNR = 0 dB

metric currently under consideration) and the corresponding beam index need be fed back, yet the maximum is taken over all n users even though no information is known about many of the users for a particular beam since that information will not be fed back.

Because the beamforming vectors are a randomly generated orthonormal basis from an isotropic distribution and the matrices are composed of i.i.d. circularly symmetric complex Gaussian random variables, there is no preferred beam over time. That is, each beam has an equal probability of being the one that produced the maximum at any given user. The joint distribution of the number of SINR values fed back for each beam can then be viewed as a multinomial distribution where the probability of each beam being selected is $\frac{1}{M}$. The distribution of the order statistic of the maximum SINR is a distribution function raised to a power that is a random variable. Marginally, the selection of each beam is distributed binomially with probability $\frac{1}{M}$. Averaging over the binomially distributed exponent applied to the distribution $F_{SINR_{(M)}}(x)$ given by Equation (2.10) yields:

$$F_{SchemeB}(x) = \sum_{i=0}^{n} \left[F_{SINR_{(M)}}(x)\right]^{Ni} \binom{n}{i} \left(\frac{1}{M}\right)^{i} \left(1 - \frac{1}{M}\right)^{n-i}. \tag{2.11}$$

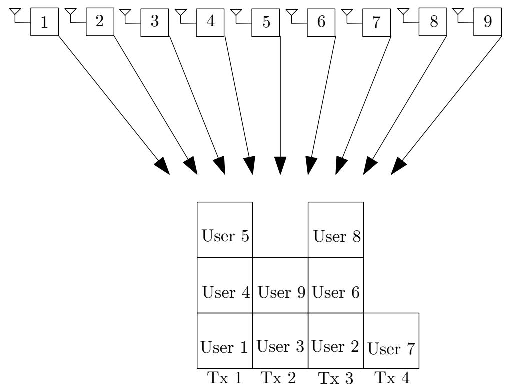

Figure 2.3: 9 User System only Feeding Back Beam Index and Largest SINR Over Transmit Beams and Receive Antennas

Figure 2.4 compares Equation (2.11) with monte carlo simulation for the binomial averaging and the simulation results match the theoretical findings. The distribution in Figure 2.4 starts at a value of approximately 0.1 rather than 0 since there is a non-zero probability that no SINR values will be fed back for a particular transmit beam. In fact the starting value is seen to be $\left(1-\frac{1}{M}\right)^n$ from Equation (2.11), which in this example evaluates to 0.107.

Throughput Analysis

Using integration by parts and Equation (2.7), the throughput of a scheme is expressed as

$$R = ME\left\{\log\left(1+X\right)\right\} = M\int_0^\infty \frac{1}{1+x}\left(1-F(x)\right)dx. \qquad (2.12)$$

where X is drawn from the distribution of the scheme being used. Numerical integration is very attractive since the closed form expression of the expectations in Equation (6) for the distributions derived earlier are not known to the authors. Figure 2.5 compares the throughput of **Scheme A** with **Scheme B** as a function of the number of users for various SNRs. Notice that **Scheme A** and **Scheme B** tend towards each other as the number of users increases. This seems reasonable since as the number of users in the system increases, it is expected that the maximum

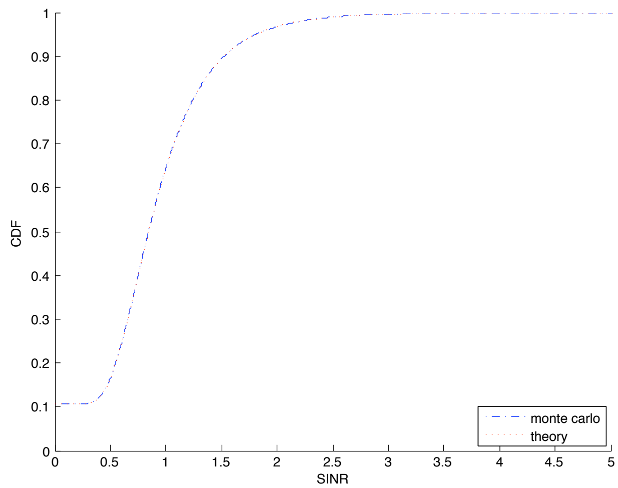

Figure 2.4: Comparing Monte Carlo simulated Distribution of the Maximum SINR Serviced with Maximum SINR Observed at User Fed Back to Theoretical Bounds for M = 5, N = 2, n = 10, SNR = 0 dB

SINR for each beam is distributed across users with high probability. When the maximum SINR for each beam is distributed over users, **Scheme B** captures the true maximum and the two schemes perform equivalently. Figure 2.5 also shows that as the SNR increases, eventually there is no performance gain due to the increased interference. Most noticeably, at SNRs of 20dB and 30dB, the performance is virtually indistinguishable.

2.4.2 Asymptotic Performance

The asymptotic performance of these schemes compared with the optimal sum-rate capacity that is achieved via dirty paper coding is of primary interest. The main result from [SH05] concerns the asymptotic scaling of **Scheme A** with one receive antenna per user and is restated here:

Theorem 1. *([SH05]) Let M and ρ be fixed and $N = 1$. Then*

$$\lim_{n \to \infty} \frac{R}{M \log \log n} = 1 \tag{2.13}$$

*where R is the throughput of **Scheme A**.*

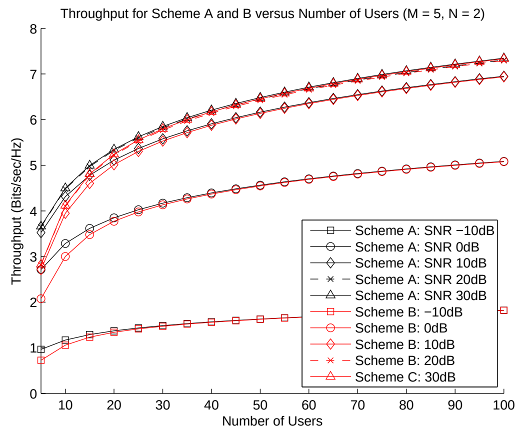

Figure 2.5: Throughput as a Function of Number of Users for Different Schemes for M = 5, N = 2 and Various SNRs

In the single receive antenna case, i.e. $N = 1$, the value of $M \log \log n$ comes from the fact that the transmitter is selecting the maximum SINR from n values for each beam. For a single receive antenna per user, it is known that the optimal sum-rate capacity scales as $M \log \log n$, and so the scaling is optimal. For **Scheme A** in the more general case where $1 < N \leq M$, the sequence of random variables for each transmit beam seen at the transmitter is $x_1, \ldots, x_{nN}$, thus the dominator in Theorem (1) becomes $M \log \log nN$, as pointed out in Section VI, Case 1 of [SH05]. The following corollary summarizes the more general case.

Corollary 1. *([SH05]) Let M, $N \leq M$, and ρ be fixed. Then for* ***Scheme A****, the throughput satisfies*

$$\lim_{n \to \infty} \frac{R}{M \log \log nN} = 1.$$

It should briefly be noted that a scaling rate of $M \log \log nN$ is essentially the same rate as $M \log \log n$ since N is a constant that inside the double logarithm becomes inconsequential in the limit as n goes to infinity, i.e.

$$\lim_{n \to \infty} (\log \log nN - \log \log n) = 0.$$

The denominator term in Theorem (1) is determined by the number of observations that the maximum is taken over. **Scheme B** feeds back only the largest SINR measured at each user at the antenna level. The sequence of random variables to be maximized over for each beam is then $x_1, \ldots, x_{M_i}$, where M_i is the number of SINR values fed back for the i^{th} beam. This led to the binomial type expression of the maximum SINR. What is the scaling when the number of terms to be maximized over is random?

Theorem (1) is concerned with the asymptotic scaling relative to the sum-rate capacity. Using the strong law of large numbers, as the number of users n grows, the total number of values fed back for each transmit beam converges to $\frac{n}{M}$. In **Scheme B** let $Y_{i,n}$ be the sequence of random variables in n denoting the number of SINR values fed back for the i^{th} beam in a system with n users. Let $\theta = \frac{1}{M}$ be the probability that a particular transmit beam is selected. The sequence of random variables $Y_{i,n}$ are binomially distributed with probability of success θ. Define $X_{i,n} = \frac{Y_{i,n}}{n}$. Then this sequence of random variables, by the central limit theorem, satisfies

$$\lim_{n \to \infty} \sqrt{n} \left(X_{i,n} - \frac{1}{M} \right) \Rightarrow \mathcal{N} \left(0, \frac{1}{M} \left(1 - \frac{1}{M} \right) \right). \tag{2.14}$$

Let $\sigma^2 = \frac{1}{M}\left(1 - \frac{1}{M}\right)$ and define the function $g(x, \theta) = \left[F_{SINR_{(M)}}(x)\right]^{nN\theta}$. The motivation for these equations is that as the number of users in the system increases, with high probability the number of SINR values fed back per beam approaches $n\theta$. The equation $g(x, \theta)$ is selected from the term being binomially averaged in Equation (2.11). The notion of each beam being the maximum for approximately $n\theta$ users with high probability is made rigorous by the Delta Method ([CB01])

$$\sqrt{n} \left[g(x, X_{i,n}) - g(x, \theta) \right] \Rightarrow \mathcal{N} \left(0, \sigma^2 \left[\frac{d}{d\theta} g(x, \theta) \right]^2 \right). \tag{2.15}$$

Plugging the values into this equation yields

$$\begin{aligned} &\left[\left(F_{SINR)_{(M)}}(x) \right)^{nNX_{i,n}} - \left(F_{SINR_{(M)}}(x) \right)^{n\frac{N}{M}} \right] \Rightarrow \\ &\mathcal{N} \left(0, n^2 M \left(1 - \frac{1}{M} \right) \left[F_{SINR_{(M)}}(x) \right]^{2\frac{nN}{M}} \left(\log \left(F_{SINR_{(M)}}(x) \right) \right)^2 \right) \end{aligned} \tag{2.16}$$

for a particular argument x. Because the distribution $F_{SINR_{(M)}}(x) < 1$ for any finite argument, the variance of the above distribution tends to zero as the number of users in the system increases. Therefore, estimating any point of the distribution of the extreme order statistic of the SINR where only the maximum SINR seen at each user is fed back can be approximated by $\left[F_{SINR_{(M)}}(x)\right]^{\frac{nN}{M}}$. This yields the following corollary:

Corollary 2. *The throughput of **Scheme B** for fixed M, $N \leq M$, and ρ scales as $M \log\log \frac{nN}{M}$, or*

$$\lim_{n\to\infty} \frac{R}{M \log\log \frac{nN}{M}} = 1.$$

Before proving the corollary, note that the effect of having the number of users grow to infinity is to provide more observation to take the maximum over. In **Scheme B** where the feedback is restricted to solely one SINR value, asymptotically in the number of users, the performance scales doubly logarithmically with $\frac{nN}{M}$ rather than nN as in **Scheme A**. As mentioned before, the multiplicative factor of $\frac{1}{M}$ difference between the schemes is inconsequential in the limit since the growth rate is double logarithmic.

Proof. It is well known ([Dav70], [Gal78]) that if there exists a limiting distribution of the maximum order statistic, the limiting distribution is one of three types. It will be shown that the distribution of the asymptotic order statistic of $F_{SINR_{(M)}}(x)$ in the terminology of [Dav70], is of type 3, i.e.

$$\Lambda_3(x) = e^{-e^{-x}} \quad -\infty < x < \infty \ . \tag{2.17}$$

A well-known condition for the asymptotic distribution to be of type 3 is

$$\lim_{x\to\infty} \frac{d}{dx}\left[\frac{1-F(x)}{f(x)}\right] = 0. \tag{2.18}$$

Carrying out the differentiation, another equivalent condition for the asymptotic distribution to be of type three is the following:

$$\lim_{x\to\infty} \frac{\left[F(x)-1\right] f'(x)}{\left(f(x)\right)^2} = 1 \tag{2.19}$$

where $f(x)$ is the density of $F(x)$, which exists since our distribution is continuous.

To show that $F_{SINR_{(M)}}$ is of type 3, it is shown that it satisfies Equation

(2.19). Since differentiation is a local property of a function and Equation (2.19) is concerned with the limit as the argument goes to infinity, significant simplification of the distribution $F_{SINR_{(M)}}$ can be made because the terms $[d_i(x)]_+^M = 0$ for $x \geq 1$ except when $i = M$. Therefore the distribution simplifies to

$$F_{SINR_{(M)}}(x) = 1 - \frac{e^{-\frac{x}{\rho}}}{\alpha\,(x+1)^{M-1}} \text{ for } x \geq 1 \tag{2.20}$$

where $\alpha = \prod_{j=1}^{M-1} \frac{j-M}{j-M-1}$. Equation (2.19) can be verified by taking the first and second derivatives of Equation (2.20), and thus $F_{SINR_{(M)}}$ is of type 3.

Notice that Equation (2.20) is equivalent to Equation (2.6) except for the α term. Following the analysis of [SH05], to satisfy the conditions of Uzgoren's theorem ([Uzg]), it must be shown that there exists a $u_n = O(\log n)$ such that

$$1 - F_{SINR_{(M)}}(u_n) = \frac{e^{-\frac{x}{\rho}}}{\alpha\,(x+1)^{M-1}} = \frac{1}{n}. \tag{2.21}$$

For sufficiently large n such that Equation (2.20) holds, the existence of a unique u_n satsifying Equation (2.21) is guaranteed since $\frac{e^{-\frac{x}{\rho}}}{\alpha(x+1)^{M-1}}$ is continuous and monotonically decreasing. To show that $u_n = O(\log n)$, notice that rearranging Equation (2.21) yields $\frac{u_n}{\rho} + (M-1)\log(1+u_n) = logn - (M-1)\log\alpha$, where $(M-1)\log\alpha$ is constant for fixed M. Thus the same conclusion as [SH05] is reached, that $u_n = \rho\log n - \rho(M-1)\log\log n + O(\log\log\log n)$, where the constant term can be absorbed by the O-notation. With the above results, the fact $F_{SINR_{(M)}}$ is of type 3, and the results from Appendix A, Theorem (1) holds exactly as in [SH05], except for one key difference. In [SH05], the rate scales with the number of users n, but from the above analysis, for a particular beam, the number of values being fed back is converging to $\frac{nN}{M}$, which yields Corollary (2). □

2.5 MMSE Receivers and Feedback

The previous feedback schemes measured the SINR at the individual antenna level. However, since each user in the system has N receive antennas, a more complex receiver structure can be utilized. It is known that the optimal linear receiver is the linear MMSE receiver. Using the system model defined in

Equations (2.1) - (2.3), the SINR after optimal combining for the i^{th} user and the j^{th} transmit beam is given by

$$SINR_{i,j} = \phi_j^* H_i^* \left[H_i \left(\sum_{k \neq j} \phi_k \phi_k^* \right) H_i^* + \frac{2}{\rho} I \right]^{-1} H_i \phi_j \tag{2.22}$$

where * denotes conjugate transposition and I is the identity matrix.

Once all the SINR values after optimal combining are computed for all the transmit beams at each receiver, the transmitter requires feedback for user selection. Prior to any maximization, there are M SINR values after optimal combining at each user. Immediately one could implement a scheme, call it **Scheme C**, similar to **Scheme A** and feed back the SINR values for each transmit beam after linear MMSE combining. Alternatively, if M SINR values represent too much data to feedback to the transmitter, a scheme similar to **Scheme B** can be adopted, call it **Scheme D**, where the maximum SINR after optimal combining can be sent back. In **Scheme D**, the total amount of feedback in the system is reduced to n analog SINR values and the corresponding n integer beam indices. To analyze **Scheme C** and **Scheme D** requires the analysis of the distribution of the SINR after optimal combining given by Equation (2.22), which leads to the following theorem.

Theorem 2. *The distribution function of the post-processing SINR given by Equation (2.22) for the system defined in Equations (2.1) - (2.3) is given by*

$$F_{MMSE}(x) = 1 - \frac{\exp\left(-\frac{x}{\rho}\right)}{(1+x)^{M-1}} \sum_{i=1}^{N} \frac{1 + \sum_{j=1}^{N-i} \binom{M-1}{j} x^j}{(i-1)!} \left(\frac{x}{\rho}\right)^{i-1}. \tag{2.23}$$

Proof. See Appendix B. □

As mentioned in the introduction, the distribution function for the interference limited regime is analyzed in [MPP07], while the SINR distribution is derived for the case $M = 4$ and $N = 2$ in [MPP08]. Figure 2.6 shows the theoretical distribution given by the above theorem versus a monte carlo empirical distribution for $M = 5$, $N = 2$, and $SNR = 10dB$ and they are in agreement.

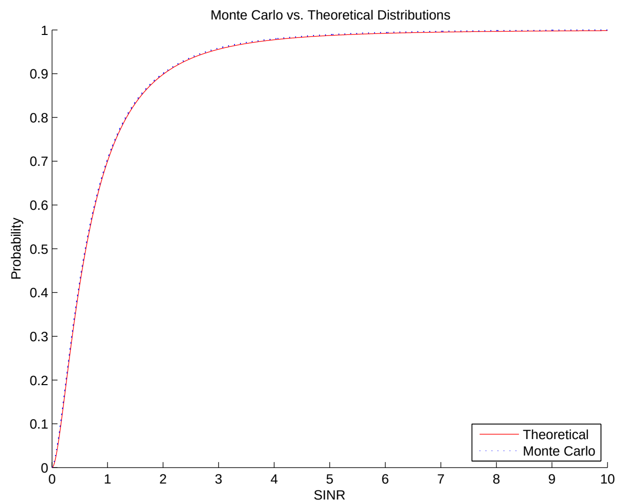

Figure 2.6: Theoretical vs. Empirical Distribution for SINR after Optimal Combining for M = 5, N = 2, SNR = 10dB

The analysis of **Scheme C** is straight forward because at the transmitter the values fed back from each user for each beam are independent and identically distributed according to the distribution given by Equation (2.23). Therefore the results from classical order statistics apply, namely the distribution of the maximum SINR selected by the transmitter for each beam is given by $[F_{MMSE}(x)]^n$.

The analysis of **Scheme D** becomes difficult in the same fashion that the analysis of **Scheme B** becomes difficult, namely the SINR values after optimal linear combining are correlated. Unlike the situation in **Scheme B** where the distribution of the maximum order statistic of the correlated random variables could be found, the distribution of the maximum order statistic of the SINR values after optimal combining at a particular user could not be found. Since the explicit distribution could not be found, bounds on the true distribution function will be used. The bounds of interest are the classical Fréchet Bounds

Theorem 3. *([Gal78], The Fréchet Bounds) Let $F(\mathbf{X})$ be an m-dimensional distribution function with marginals $F_j(x), 1 \leq j \leq m$. Then, for all $x_1, x_2, \ldots, x_m$*

$$\max\left(0, \sum_{j=1}^{m} F_j\left(x_j\right) - m + 1\right) \leq F\left(x_1, x_2, \ldots, x_m\right) \leq \min\left(F_1\left(x_1\right), \ldots, F_m\left(x_m\right)\right).$$

The Fréchet Bounds provide control of the unknown joint distribution of the SINR values after optimal combining, and can provide bounds on the order statistic because all the marginal distributions are identical:

$$\max\left(0, M\left(F_{MMSE}\left(x\right)-1\right)+1\right) \leq F_{MaxMMSE}(x) \leq F_{MMSE}(x) \tag{2.24}$$

where $F_{MaxMMSE}$ is the distribution of the maximum SINR at a user after optimal combining.

The asymptotic performance of **Scheme C** and **Scheme D** are of interest. To show that **Scheme C** has optimal scaling asymptotically, it must first be shown that $F_{MMSE}(x)$ given by Equation (2.23) is of type 3 (Equation (2.17)).

Theorem 4. *The limiting distribution of the extreme order statistics drawn from the distribution $F_{MMSE}(x)$ is of type 3.*

Proof. See Appendix C. □

Having established the previous theorem, the scaling rate is shown in Appendix C to satisfying the following corollary:

Corollary 3. *For fixed M, $N \leq M$, and ρ, the throughput for **Scheme C** asymptotically scales as $M \log\log n$, or*

$$\lim_{n\to\infty} \frac{R}{M \log\log n} = 1.$$

Compared with **Scheme A**, the asymptotic scaling is $\log\log n$ rather than $\log\log nN$, but in the limit the extra factor of N becomes insignificant. The loss of the factor of N comes from the fact that in **Scheme C** only one SINR value (after optimal combining) is fed back per user and in **Scheme A** the SINR value that was the maximum of N SINR values (without optimal combining) was fed back, resulting in the extra factor of N.

Theorem (4) establishes that $F_{MMSE}(x)$ is of type 3. The bounds in Equation (2.24) are in terms of $F_{MMSE}(x)$ and thus are also of type 3 implying the true unknown distribution of the maximum order statistic is of type 3. As with **Scheme B**, the number of SINR values after optimal combining that are maximized over at the transmitter in **Scheme D** is a random quantity, leading to binomial averaging. The binomial averaging can be applied to both the bounds given by Equation (2.24) and the Delta Method arguments will yield an exponent in both bounds asymptotically approaching $\frac{n}{M}$. To get a handle on the asymptotic scaling, consider the lower and upper bounds in Equation (2.24). The upper bound scales as $M \log\log n$ as shown in Appendix C. Substituting the distribution into the lower bound yields $\max\{0, 1 - MR(z)\}$, where $R(z)$ is defined by Equation (B.2) in Appendix B. For sufficiently large n, $\max\{0, 1 - MR(z)\} = 1 - MR(z)$ since $R(z)$ is monotonically decreasing to zero, so by the same methods as Appendix C, the lower bound scales as $M \log\log n$. Both the lower and upper bounds have the same asymptotic scaling rate, yielding the following corollary:

Corollary 4. *For fixed M, $N \leq M$, and ρ, the throughput for **Scheme D** asymptotically scales as $M \log\log n$, or*

$$\lim_{n \to \infty} \frac{R}{M \log\log n} = 1.$$

The upper bound in Equation (2.24) is very loose. Although it could not be shown, empirical evidence suggests that the SINR values after optimal combining are negatively associated. The definition of negative association is as follows:

Definition 1. *([JDP83], Negative Association) Random variables $X_1, X_2, \ldots, X_k$ are said to be negatively associated if for every pair of disjoint subsets A_1, A_2 of $\{1, 2, \ldots, k\}$,*

$$Cov\left\{f_1\left(X_i, i \in A_1\right), f_2\left(X_j, j \in A_2\right)\right\} \leq 0$$

whenever f_1 and f_2 are both increasing or both decreasing.

Negative association is a multivariate generalization of negative correlation. The key property of interest is that if $X_1, \ldots, X_N$ are negatively associated random variables, then

$$\Pr\left(X_1 \leq x_1, \ldots, X_N \leq x_N\right) \leq \prod_{i=1}^{N} \Pr\left(X_i \leq x_i\right).$$

Applying this to our distribution, we conjecture a tighter upper bound is given by

$$F_{MaxMMSE}(x) \leq \left[F_{MMSE}(x)\right]^M \tag{2.25}$$

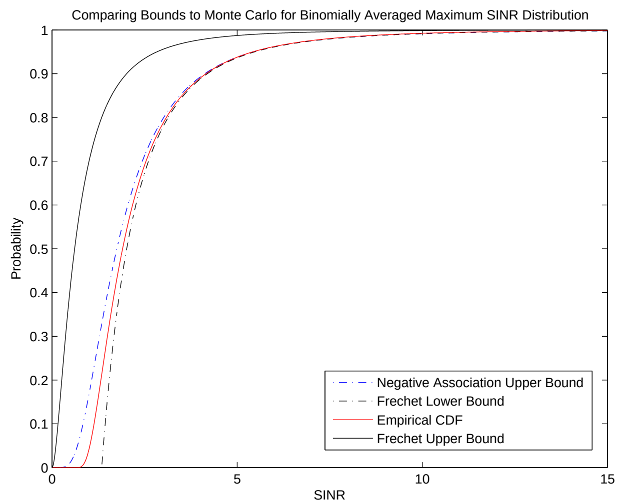

Figure 2.7: Theoretical vs. Empirical Distribution for SINR after Optimal Combining for M = 5, N = 2, SNR = 10dB

Figure 2.7 shows the bounds versus monte carlo simulations. Although the original upper bound due to Fréchet is weak, it sufficed in showing the optimal asymptotic scaling since the bound was of type 3. The tighter hypothesized bound due to negative association may be of use in numerical computations.

Figure 2.8 shows **Scheme C** and the upper and lower bounds for **Scheme D**. The upper bound is the conjectured bound due to negative association. The figure shows that the bounds on **Scheme D** converge as the number of users increases and that they converge to the rate of **Scheme C**. As with **Scheme A** compared with **Scheme B**, this convergence is expected as the maximum for each transmit beam will be distributed over users with high probability as the number of users increases. The throughput also becomes saturated as the SNR increases and the interference becomes dominant. Comparing Figure 2.5 and Figure 2.8, utilizing the extra receive antenna for optimal combining rather than just additional observations yields significant throughput gains at the expense of receive complexity.

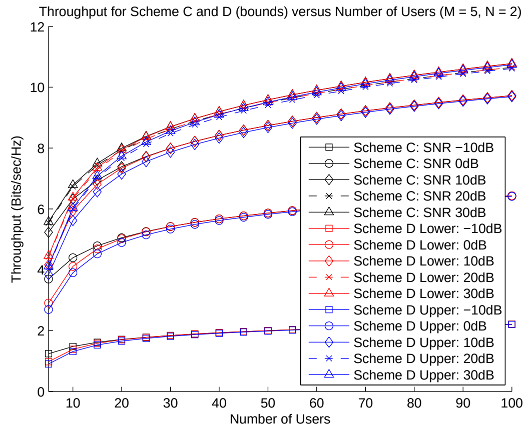

Figure 2.8: Throughput as a Function of Number of Users for Different Schemes for M = 5, N = 2 and Various SNRs

2.6 Thresholding the Feedback

Feedback is a precious commodity and the amount of feedback may have to be reduced to a bare minimum. One method of further reducing the amount of feedback is to apply a threshold at the receivers, and only send back the SINR values that exceed the threshold. Intuitively, if the SINR is below a reasonable threshold, it is very unlikely to be the maximum selected by the transmitter. This idea is briefly mentioned in [SH05] for **Scheme A**, where it is mentioned that as n grows large, the SINR value need only be fed back if it exceeds some constant threshold η, which is independent of n, to maintain the scaling laws. The rate lost in a system with a finite number of users is not considered. The analysis in this section determines the effects on throughput of applying any fixed threshold to a system with an arbitrary number of users, and then applying that analysis to show asymptotically there is no loss in throughput by applying any finite threshold. Additionally, by applying an appropriate threshold, one can trade off throughput for a reduction in feedback.

It is of interest to see how thresholding the SINR affects the distributions over which the expectations are taken to determine the throughput of the previous schemes. No closed form expression for the expectation could be found due to the complicated nature of the distributions. The effect of thresholding is to truncate the underlying cdf of the SINRs observed at the users in the following manner:

$$F_{threshold}(x) = \begin{cases} F(x_{th}) & x \leq x_{th} \\ F(x) & x > x_{th} \end{cases} \tag{2.26}$$

where x_{th} is the threshold. All SINR values less than the threshold are not fed back, which can be viewed as a mapping to zero throughput, resulting in the above truncation.

For concreteness, consider **Scheme A**. Applying Equation (2.12) yields

$$R \approx M \int_0^\infty \frac{1}{1+x} \left(1 - F_s(x)^{nN}\right) dx \tag{2.27}$$

After thresholding, the throughput of the scheme is

$$\begin{aligned} R_{threshold} &\approx M \left[\int_0^{x_{th}} \frac{1}{1+x} \left(1 - \left[F_s(x_{th})\right]^{nN}\right) dx + \int_{x_{th}}^\infty \frac{1}{1+x} \left(1 - F_s(x)^{nN}\right) dx \right] \\ &= M \left[\log(1 + x_{th}) \left(1 - \left[F_s(x_{th})\right]^{nN}\right) + \int_{x_{th}}^\infty \frac{1}{1+x} \left(1 - F_s(x)^{nN}\right) dx \right], \end{aligned}$$

which is a constant term plus an integral.

The loss in throughput due to thresholding can be expressed as

$$\begin{aligned} R - R_{threshold} &= M \int_0^\infty \frac{1}{1+x} \left[F_{threshold} - F_{max}\right] dx \\ &= M \int_0^{x_{th}} \frac{1}{1+x} \left[F_{max}(x_{th}) - F_{max}(x)\right] dx + \\ &\quad M \int_{x_{th}}^\infty \frac{1}{1+x} \left[F_{max}(x) - F_{max}(x)\right] dx \\ &= M \left[\log(1 + x_{th}) F_{max}(x_{th}) - \int_0^{x^{th}} \frac{1}{1+x} F_{max}(x) dx \right] \\ &\leq M \log(1 + x_{th}) F_{max}(x_{th}), \end{aligned}$$

where F_{max} could be any of the maximum distributions derived earlier. As the number of users grows, the upper bound on the throughput loss goes to zero for all the distributions discussed, since for a finite threshold, $F_{max}(x_{th}) \to 0$ as the number of users increases. Therefore, asymptotically, if the threshold is finite, the scaling laws derived earlier hold.

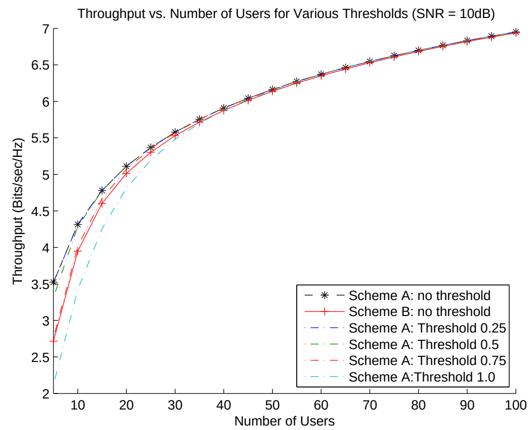

Figure 2.9: Throughput for Various Threshold Levels

For **Scheme A**, the number of SINR values not sent back due to thresholding is $nMF_s(x_{th})$, that is the $\Pr[\text{SINR} \leq x_{th}] = F_s(x_{th})$ multiplied by the number of possible values that could be sent back. Viewing it another way, it is a binomial experiment with nM trials and the probability of success being $F_s(x_{th})$. Similarly, using the cdf of the SINR after optimal combining, the number of SINR values saved for **Scheme C** is on average $nMF_{MMSE}(x_{th})$. For **Scheme B** the amount of savings is on average $n\left[F_{SINR_{(M)}}(x_{th})\right]^N$ and the savings for **Scheme D** can be bounded using the Fréchet Bounds.

Figure 2.9 shows **Scheme B** without thresholding against **Scheme A** for various levels of thresholding. For small thresholds, there is almost no change in throughput performance, whereas for the higher thresholds, large losses are suffered for small numbers of users. As the number of users is increased, there is no loss in performance as expected. **Scheme B** performs better than **Scheme A** for the threshold of 1 for small to moderate number of users, but does not uniformly bound the thresholded version of **Scheme A**. It is not quite a far comparison since **Scheme B** deterministically feeds back only one SINR value and the associated beam index, while applying a threshold to **Scheme A** yields a reduction in feedback that is a random variable. In this vein, it may be of interest to design the threshold to achieve an average amount of feedback per fading period.

2.7 Conclusion

This chapter is concerned with a random beamforming scheme that puts a premium on the amount of feedback in the system. To first reduce the amount of feedback, a metric was chosen that captures the quality of the channel, and that metric was the SINR, and has been motivated by its use in the formula for throughput of the channel. With a good metric chosen, a few schemes that feed back this SINR information are considered.

First, the scheme in [SH05] (**Scheme A**) was reviewed to provide background and context for different novel schemes. **Scheme B** suggests utilizing knowledge at the receiver and send back the maximum SINR observed at each user over all beams. This reduces the amount of feedback in the system to n SINR values. Since each user has N receive antennas, it is possible to perform optimal combining at the receivers. The distribution for the SINR after optimal combining is derived. There are two schemes for utilizing this optimally combined SINR value: either all M SINRs could be fed back per user (**Scheme C**), or send back only the maximum SINR at each user leading to system total of n feedback values (**Scheme D**). The asymptotic scaling of all the schemes is shown to be essentially $M \log \log n$.

Finally, thresholding was considered as another means of reducing the amount of feedback. If the SINR value to be sent back in any scheme lies below the set threshold, that SINR value is not sent back at all. Asymptotically in the number of users, the throughput lost due to thresholding at any finite level is shown to go to zero.

This chapter is a reprint of a paper coauthored with Bhaskar D. Rao and appears in the March 2010 issue of *IEEE Transactions on Signal Processing* under the title *"Reduced Feedback Schemes Using Random Beamforming in MIMO Broadcast Channels"*. Sections of this chapter also appear in *"On the Capacity of MIMO Broadcast Channels with Reduced Feedback by Antenna Selection"* in *Asilomar Conference on Signals Systems, and Computers, March 2008* and *"Feedback reduction in MIMO broadcast channels with LMMSE receivers"* in *ICASSP 2010.* Both of these works are also coauthored with Professor Bhaskar D. Rao.

Chapter 3

Reducing Feedback in Broadcast Channels via Thresholding

3.1 Outline

In a Gaussian multiple-input multiple-output (MIMO) broadcast channel with random transmit beamforming and a greedy scheduling algorithm, the question of which receivers should feed back their channel state information (CSI) is investigated. A thresholding mechanism is applied to the CSI such that receivers observing CSI lower than the threshold do not feed back. Thresholding is of interest because it reduces the unwanted overhead, which is necessary to achieve multiuser diversity, of CSI feedback in broadcast channels. Utilizing the theory of convergence of types, it is shown that if the thresholding function $T(n)$ scales slower than $\log n$, i.e. $T(n) \in o(\log n)$, that asymptotically in the number of users, no performance is lost compared to the system without a threshold. Conversely, using the asymptotic stability of order statistics, it is shown that if the thresholding function $T(n)$ grows faster than $\log n$, i.e. $T(n) \in \omega(\log n)$, eventually no user in the system will feed back their CSI and all multiuser diversity is lost. For a finite number of users, utilizing the closed form SINR distribution expressions, thresholds are designed to meet specific design criterion as a function of the total number of receivers in the system. Three design criterion are considered: limiting the probability that no receivers feed back information for any transmit direction, limiting the average number of receivers feeding back SINR information and limiting the rate loss due to thresholding. The material in this chapter is work which in preparation for submission to the *IEEE Transactions on Signal Processing* under the title *"Reducing Feedback in Broadcast Channels via Thresholding"* and material submitted to *Asilomar Conference on Signals, Systems, and Computers 2011* under the title *"Feedback Reduction by Thresholding in Multi-User Broadcast Channels: Design and Limits"*.

3.2 Introduction

The multiuser broadcast channel has been the subject of intense study for many years as it has numerous direct applications to modern technologies such as the cell phone downlink channel. Great strides have been taken in characterizing the theoretical limits of the broadcast channel, especially the vector Gaussian broadcast channel. Specifically there has been an explosion of work regarding the sum-rate capacity of the vector Gaussian broadcast channel starting with [CS03] and then followed up by [VT03], [VJG03], and [YC04]. The capacity region of the Gaussian vector broadcast channel was characterized in [WSS06]. While of great theoretical importance, many of the schemes used to achieve the theoretical optimal capacity are too complex or impractical to implement in real systems. One of the greatest hinderances in achieving the theoretical optimum is having complete CSI for all of the users in the multiuser broadcast channel available at the transmitter. It is well known that without CSI at the transmitter, the system cannot achieve multiuser diversity, which is a significant benefit of multiuser systems. In order to reap the benefits of the more advanced multiuser MIMO architecture, CSI feedback from the receivers to the transmitter is essential.

To address this critical problem, researchers have developed different methods of quantifying important information to feedback and how to best feed back this information. The problem becomes more complicated when one also includes considerations about the reliability of the CSI due to channel estimation errors and channel dynamics. In [IR07],[IAR09] and [IR10] the effects of imperfect channel estimation, feedback delay, and finite-rate quantization are considered on bit, symbol, and packet error probabilities in multiple-input single-output (MISO) channels. The effects of imperfect channel estimation and feedback delay on transmit beamforming is considered in [MZ07] and [MLS08].

In this chapter, channel estimation is assumed to be perfect and the feedback channel is delay and error free. Much reseach has been conducted under these assumptions. In [Jin06], the channel vector from the multi-antenna transmitter to the single receive antenna users is quantized using a random vector quantizer and the performance is analyzed while [LHS03] considers codebook design using the Grassmannian manifold. The asymptotic optimality of zero-forcing beamforming is studied in [YG06a]. The question of how much feedback is required so that the system does not reach a sum-rate ceiling is considered in [BK06].

The first steps towards thresholding feedback are in the works [FFM03, GA03, GA04] which consider one bit feedback schemes where a binary decision is made as to whether or not the channel is in a good or bad state for transmission. These works were later extended in [VHO07], where the feedback channel is used multiple times per coherence time to allow for refinement of the scheduling decision. The rate enhancement due to additional feedback bits is analyzed in [SN07]. Additional work along these lines can be found in [YAHA05] as well as an overview of feedback systems in general in [LJL$^+$08].

The tranmission scheme analyzed in this chapter is the random beamforming scheme proposed in [SH05]. The main idea is for the transmitter to transmit on the basis vectors of a random orthonormal basis and each receiver measures the SINR along each of the random directions. The receivers feed back their measured SINR values and the transmitter schedules the users with the largest SINR value for each orthonormal direction. In this transmission scheme, all of the channel information is captured in the SINR values. This scheme is extended in [PR10a] in several ways, including more advanced receive structures, namely LMMSE receivers, when each user is allowed multiple receive antennas. The main result of these works is that random beamforming is simple yet, asymptotically in the number of users, achieves the optimal sum-rate throughput scaling.

In the aforementioned works, all the users in the system feed back some form of CSI. This chapter addresses the question of which user should feed back CSI. To answer this question, the random beamforming scheme is considered. In the random beamforming scheme, the statistics of the SINR values, which is the CSI, are known (e.g. see [SH05, PR10a]). With knowledge of the CSI statistics, consider the example where a particular user in the system observers a small SINR value. Should that user feed back this small SINR value? If there are a large number of users in the system, then it is highly probable that someone will have a larger SINR value. In this case it does not affect the system if the small SINR value is not fed back due to the scheduling algorithm always selecting the largest SINR. On the other hand, it is always possible that every other user in the system has a smaller SINR value, in which case not feeding back will prevent the system from achieving the best throughput. The goal of this chapter is to design a threshold that achieves certain performance metrics as a function of the number of users in the system such that a user feeds back its observed SINR value if the value is above the threshold and does not feed back if the SINR is below the threshold. By allowing users to not feed back, transmit power can be saved as well as relieving congestion on the radio resource.

The contributions of this chapter can be broken into two parts. First, conditions on the scaling rate of the threshold $T(n)$ are considered as the number of users n in the system asymptotically grows to infinity. It is shown that if the threshold scales as $o(\log n)$, that asymptotically there is no loss in performances as compared to the case where every user feeds back their SINR information. Conversely, it is shown that if the threshold $T(n)$ scales as $\omega(\log n)$, that eventually no user in the system will feed back their SINR information, i.e. asymptotically no user in the system exceeds the threshold. In this case the transmitter has no CSI information and it is known that multiuser diversity cannot be achieved. The second part of the chapter considers explicit design of the thresholding function as a function of the number of users in the system when the number of users is finite. For a finite number of users, it is always possible that every user will be beneath any selected threshold and thus no one will feed back. Therefore, the design of the threshold will be with respect to achieving certain design criterion. In this chapter, three metrics are considered: limiting the probability that no user in the system feeds back for any transmit direction, limiting the average number of users feeding back and limiting the rate loss due to thresholding.

The organization of the chapter is as follows: the system model and random beamforming transmit scheme are discussed in Section 3.3. Section 3.4 analyzes the asymptotic scaling rate of the thresholding function. Section 3.5 addresses the design of the thresholding function in the case of a finite number of users. Finally, Section 3.6 summarizes the results of this chapter.

3.3 System Model and Random Beamforming

In this section the system model is described. Under the given system model, the random beamforming technique first proposed in [SH05] is outlined. The key implications of this work will be discussed as well as extensions found in [PR10a].

3.3.1 System Model

A block fading channel model is assumed for the Gaussian broadcast channel. The transmitter has M transmit antennas and there are n receivers (users), each with N receive antennas. It is assumed that $n \gg M$ and $M \geq N$, which is typically the case in a cellular system. Let $\mathbf{s}(t) \in \mathbb{C}^{M\times 1}$ be the transmitted vector of symbols at time slot t and $\mathbf{y}_i(t) \in \mathbb{C}^{N\times 1}$ be the received symbols by the i^{th} user at time slot t. The following model is used for the input-output relationship between the transmitter and the i^{th} user:

$$\mathbf{y}_i(t) = \sqrt{\rho_i} H_i \mathbf{s}(t) + \mathbf{w}_i(t), \quad i = 1, \ldots, n. \tag{3.1}$$

$H_i \in \mathbb{C}^{N\times M}$ is the complex channel matrix which is assumed to be known at the receiver, $\mathbf{w}_i \in \mathbb{C}^{N\times 1}$ is the white additive noise, and the elements of H_i and $\mathbf{w}_i$ are i.i.d. complex Gaussians with zero mean and unit variance. This is the standard Rayleigh fading model, where the channel gain between every transmit-receive antenna pair is a complex circularly symmetric normal random variable. The transmit power is chosen to be M, i.e. $E\{\mathbf{s}^*\mathbf{s}\} = M$, so that the normalized power per antenna is one. The signal-to-noise ratio (SNR) at the receiver is $E\{\rho_i|H_i\mathbf{s}|^2\} = M\rho_i$ and ρ_i is the SNR of the i^{th} user. It is assumed that the network is homogeneous, thus $\rho_i = \rho \ \forall i$.

3.3.2 Random Beamforming

The random beamforming transmit scheme proposed in [SH05] is the scheme analyzed in this chapter. The key elements of this work, as well as extensions found in [PR10a], are now described.

The random beamforming transmit scheme involves generating an M dimensional random orthonormal basis from an isotropic distribution, with basis vectors $\phi_m \in \mathbb{C}^{M\times 1}$ for $m = 1, \ldots, M$. Let $s_m(t)$ be the m^{th} transmit symbol at time t, then the total transmit signal at time slot t is given by

$$\mathbf{s}(t) = \sum_{m=1}^{M} \phi_m(t) s_m(t). \tag{3.2}$$

The received signal at the i^{th} user is given by

$$\mathbf{y}_i(t) = \sum_{m=1}^{M} \sqrt{\rho} H_i(t)\phi_m(t)s_m(t) + \mathbf{w}_i(t). \tag{3.3}$$

In the original formulation in [SH05], each of the N receive antennas at each user measure the SINR for each of the M transmit directions and the maximum of the observed SINR values for each transmit direction is fed back. Looking at Figure 3.1, which shows the SINR values for each receive antenna/transmit direction pair, this scheme feeds back the maximum SINR in each column. The transmitter then schedules the users experiencing the largest SINR for each transmit direction, leading to maximal order statistics.

It is assumed that the i^{th} user knows the quantity $H_i(t)\phi_m(t)$ from Equation 3.3 for each transmit direction. With this knowledge, the SINR observed at the

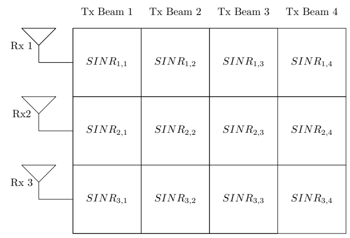

Figure 3.1: SINR observations at a single user for $M = 4, N = 3$

j^{th} receive antenna of the i^{th} user for the m^{th} transmit direction is given by the following equation:

$$SINR_{i,j,m} = \frac{|H_{i,j}(t)\phi_m(t)|^2}{\frac{2}{\rho} + \sum_{k \neq m} |H_{i,j}(t)\phi_k(t)|^2}. \tag{3.4}$$

$H_{i,j}$ is the j^{th} row of the i^{th} user's channel matrix. The orthonormal beamforming vectors and i.i.d. zero mean and unit variance complex Gaussian elements of H_i imply the numerator in Equation 3.4 is distributed as a $\chi^2(2)$ random variable and the denominator is an independent $\chi^2(2(M-1))$ random variable.

When each receiver has a single receive antenna, i.e. $N = 1$, then under the assumed system model, it is shown in [SH05] that the distribution of the SINRs at the receive antennas is given by

$$f_{SINR}(x) = \frac{e^{-\frac{x}{\rho}}}{(1+x)^M} \left(\frac{1}{\rho}(1+x) + M - 1 \right) u(x), \tag{3.5}$$

where $u(x)$ is the unit step function. From now on, the $u(x)$ will be dropped from the SINR distribution and SINR density expressions with the understanding that all the SINR random variables of interest are nonnegative. Integrating the density in Equation 3.5 yields the distribution function for the SINR random variables:

$$F_{SINR}(x) = 1 - \frac{e^{-\frac{x}{\rho}}}{(1+x)^{M-1}}. \tag{3.6}$$

The main result of [SH05] is that asymptotically random beamforming has the optimal sum-rate throughput scaling. More specifically, the sum-rate throughput of the random beamforming technique R asymptotically scales at a rate of $M \log \log n$, which is the theoretical optimal scaling rate. This theorem is now stated:

Theorem 5. *([SH05]) Let M and ρ be fixed and $N = 1$. Then*

$$\lim_{n\to\infty} \frac{R}{M \log\log n} = 1 \tag{3.7}$$

where R is the throughput of the random beamforming technique.

If there are $N \geq 1$ receive antennas at each user, the additional antennas can be utilized in different ways. In [SH05], each receive antenna at a given user is considered an independent user, in which case all the SINRs observed at each antenna are distributed according to the distribution function in Equation 3.6. An extension considered in [PR10a] is to feedback only the largest SINR observed across all receive antennas and over all transmit directions, thus limiting the feedback to a single SINR value per user (and the index of the transmit direction that generated the largest SINR value) and reducing the total amount of feedback in the system. Once again referring to Figure 3.1, this scheme corresponds to feeding back the largest SINR value in the array. It is shown in [PR10a] that the distribution of the SINR value fed back is given by $\left[F_{SINR_{(M)}}(x)\right]^N$ where $F_{SINR_{(M)}}(x)$ is given by

$$F_{SINR_{(M)}}(x) = 1 - \sum_{i=1}^{M} \frac{[d_i(x)]_+^M \exp\left(-\frac{xc}{d_i(x)}\right)}{A_i(x)} \tag{3.8}$$

and where $d_i(x) = \frac{2[1-x(M-i)]}{M-i+1}$, $A_i(x) = d_i(x)\prod_{j\neq i}^{M}(d_i(x) - d_j(x))$, and $[\cdot]_+$ is the positive part of the argument.

The extra receive antennas at each user can also be utilized by implementing a more complicated receive architecture such as an LMMSE receiver. With this architecture, the receiver will measure the SINR after LMMSE reception along each transmit direction and feed back the values. Under the given system model, it is shown in [PR10a] that the distribution of SINR after LMMSE reception is given by

$$F_{MMSE}(x) = 1 - \frac{\exp\left(-\frac{x}{\rho}\right)}{(1+x)^{M-1}} \sum_{i=1}^{N} \frac{1 + \sum_{j=1}^{N-i} \begin{pmatrix} M-1 \\ j \end{pmatrix} x^j}{(i-1)!} \left(\frac{x}{\rho}\right)^{i-1}. \tag{3.9}$$

Using these distributions, it is shown in [PR10a] that the results of Theorem 5 hold for these alternative feed back schemes and architectures. With the previous results now reviewed, the properties of the distribution functions can be used to ask questions about designing thresholds to limit which users feed back CSI and the effects these thresholds have on the system performance.

3.4 Thresholding Asymptotically in the Number of Users

Let $T(n)$ be the function that determines the threshold as a function of the number of users n in the system. For a fixed n, users experiencing a SINR less than $T(n)$ do not report back their observed SINRs and users experiencing SINR values larger than $T(n)$ feed back their observed values. In this section, the question analyzed is how fast can the thresholding function $T(n)$ grow and still have the thresholded system achieve the same optimal asymptotic sum-rate scaling as the non-thresholded system? The theory of extreme value distributions and the convergence of types are essential to answering this question and will now be briefly reviewed.

3.4.1 Extreme Value Distributions

The convergence of types theorem is critical for the analysis of extreme value distributions and is now stated.

Theorem 6. *(Convergence of Types)*
Suppose $U(x)$ and $V(x)$ are two distribution function neither of which concentrates at a point. Suppose for $n \geq 1, F_n$ is a distribution, $a_n \geq 0, b_n \in \mathbb{R}, \alpha_n \geq 0, \beta_n \in \mathbb{R}$ and

$$F_n\left(a_n x + b_n\right) \to U(x)$$
$$F_n\left(\alpha_n x + \beta_n\right) \to V(x)$$

weakly. If

$$\tfrac{\alpha_n}{a_n} \to A > 0$$
$$\tfrac{\beta_n - b_n}{a_n} \to B \in \mathbb{R}$$

then

$$V(x) = U(Ax + B).$$

Let $X_1, \ldots, X_n$ be i.i.d. random variables with common distribution function $F(x)$ and let $M_n = \bigvee_{i=1}^{n} X_i$. The main theorem concerning extreme value distributions states that if a nondegenerate limiting distribution exists for M_n, then the limiting distribution converges to one of three classes of types:

Theorem 7. *(Gnedenko, 1943)*
Suppose there exists $a_n > 0, b_n \in \mathbb{R}, n \geq 1$ such that

$$P\left[(M_n - b_n)/a_n \leq x\right] = F^n(a_n x + b_n) \to G(x),$$

weakly as $n \to \infty$, where G is assumed nondegenerate. Then G is of the type of one of the following three classes:

$$\text{Type 1) } \Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp\{-x^{-\alpha}\} & x \geq 0 \end{cases} \quad \text{for some } \alpha > 0.$$

$$\text{Type 2) } \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{for some } \alpha > 0.$$

$$\text{Type 3) } \Lambda(x) = \exp\{-e^{-x}\} \quad x \in \mathbb{R}$$

The asymptotic scaling of the sum-rate throughput for the random beamforming scheme is established by showing the SINR values scheduled fall in the domain of attraction of the Type 3 extreme value distribution given in Theorem 7. Formally, $F_{SINR} \in D(\Lambda)$, i.e. the distribution function for the SINR, lies in the domain of attraction of Type 3. Once the extreme value distribution is established, the constants a_n and b_n are determined and used to show that the sum-rate throughput of the random beamforming scheme scaled optimally. The selection of the sequence a_n and b_n is not unique and the following lemma shows the relationship between any selection of sequences that satsify $F \in D(\Lambda)$.

Lemma 1. *Suppose $F \in D(G)$, where G is one of the limiting distributions from Gnedenko's Theorem, and let $a_n > 0, b_n \in \mathbb{R}, \alpha_n > 0$, and $\beta_n \in \mathbb{R}$ satisfy $F^n(a_n x + b_n) \to G(x)$ and $F^n(\alpha_n x + \beta_n) \to G(x)$. Then $a_n \sim \alpha_n$ (i.e. $\lim_{n\to\infty} a_n/\alpha_n = 1$) and $(\beta_n - b_n) \in o(a_n)$.*

Proof. The existence of a_n, b_n, α_n and β_n is guaranteed by Gnedeno's Theorem since $F \in D(G)$. Also by Gnedenko's Theorem, G is the only possible limiting distribution that F approaches since $F \in D(G)$. Since $F^n(a_n x + b_n) \to G(x)$ and $F^n(\alpha_n x + \beta_n) \to G(x)$, by the Convergence of Types Theorem and letting G play the role of U and V yields the constants $A = 1$ and $B = 0$. Thus, also by the Convergence of Types Theorem, $\frac{\alpha_n}{a_n} \to 1$ and $\frac{\beta_n - b_n}{a_n} \to 0$, which gives the desired result. □

3.4.2 Sufficient Conditions on the Scaling Rate of the Thresholding Function

The effects of thresholding on the distribution function F_{SINR} must be understood to design a threshold that achieves optimal asymptotic scaling of the sum-rate throughput. In [PR10a], it is observed that that the threshold truncates the distribution function of the SINR values fed back to the transmitter in the following manner:

$$F_{SINR}^{threshold}(x) = \begin{cases} F_{SINR}(x) & x \geq x_{threshold} \\ F_{SINR}(x_{threshold}) & 0 \leq x < x_{threshold} \\ 0 & x < 0 \end{cases} \tag{3.10}$$

Equation 3.10 states that the probablity mass less than $x_{threshold}$, the threshold value, concentrates at zero, and this is the probability that the user does not feed back. The mass concentrates at zero because this is the lowest possible value that the nonnegative SINR random variables can attain. With this mapping, the values that do not exceed the threshold do not affect the scheduling decision since they are assumed to be the smallest possible value. To generalize the truncation from nonnegative random variables, let $x_1 = \inf\{x : F(x) > 0\}$ for an arbitrary distribution function F, i.e. the left end point of the support of F. Thresholding an arbitrary distribution function produces the following truncation:

$$F^{threshold}(x) = \begin{cases} F(x) & x \geq x_{threshold} \\ F(x_{threshold}) & x_1 \leq x < x_{threshold} \\ 0 & x < x_1 \end{cases} \tag{3.11}$$

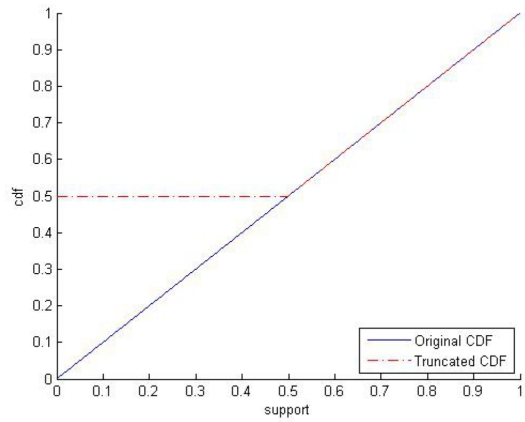

Figure 3.2: Uniform Distribution Truncated at $x_{threshold} = 0.5$

Figure 3.2 shows the truncation of the uniform distribution at $x_{threshold} = 0.5$. The probability mass less than $x_{threshold} = 0.5$ maps to the left end point of the support, which in this case is $x_1 = 0$.

The effect of thresholding by truncation on a distribution function has been mathematically quantified. Time will now be spent to show that this mathematical definition corresponds with the goal of reducing feedback by implementation of a threshold. Let X be a random variable drawn from a distribution function $F \in D(\Lambda)$. Then the distribution function from Equation 3.11 corresponds to the following random variable Y:

$$Y = \begin{cases} X & X \geq x_{threshold} \\ x_1 & X < x_{threshold} \end{cases} \tag{3.12}$$

From an engineering perspective, this formulation is motivated by the following scenario. Let there be n users, where user i makes an observation X_i drawn i.i.d. from $F \in D(\Lambda)$. Each user decides to send back their observation X_i if it exceeds some threshold $x_{threshold}$ to a central unit, and the central unit has the task of finding the maximum observation. The users that did not report their observations because they did not exceed the threshold should not effect the decision of finding the maximum. Thus if a user did not report a value, their observation is assumed to be x_1. By taking on the lowest possible value, it cannot affect the task of finding the maximum value among the users that did feed back, and this motivates Equation

3.11.

Remark: If $x_1 = -\infty$, then $F^{threshold}(x)$ given by Equation 3.11 is not a proper distribution function because there is non-zero probability mass at $-\infty$. This will not pose a problem in the following analysis. Additionally, any point x_2 chosen such that $x_1 \leq x_2 < x_{threshold}$ could be selected since such a selection would not affect the task of finding the maximum among the fed back values, all of which exceed x_2.

In [PR10a], it is shown that any finite threshold does not affect the asymptotic performance. The goal here is to design a threshold as a function of the number of users, $T(n)$, that grows as the number of users grows, and still has the desired asymptotic scaling. The following theorem is the main result of this section and provides conditions on the scaling rate of the threshold as a function of the number of users that guarantee optimal asymtotic scaling of the sum-rate throughput.

Theorem 8. *Consider the system when $N = 1$, i.e. each user has a single receive antenna, under the model and random beamforming scheme described in Section 3.3. If the thresholding function $T(n) \in o(\log n)$, then the sum-rate throughput of the system satisfies $\frac{R}{M \log \log n} \to 1$.*

Before presenting the proof of Theorem 8, the following lemma is required.

Lemma 2. *Suppose $F \in D(\Lambda)$ with the sequence of coefficients $a_n > 0$ and $b_n \in \mathbb{R}$ satisfying $F^n(a_n x + b_n) \to G(x)$ and the left end point of the support is x_1. Let X_i be drawn i.i.d. from F and let $M_n = \bigvee_{i=1}^{n} X_i$. Also let $T(n)$ be the threshold as a function of n. Then $F^{threshold} \in D(\Lambda)$ if and only if $\frac{T(n)-b_n}{a_n} \to -\infty$.*

Proof. $\Leftarrow$: Suppose $\frac{T(n)-b_n}{a_n} \to -\infty$. Let Y_i be drawn i.i.d. from $F^{threshold}$ and let

$M_n^{th} = \bigvee_{i=1}^{n} Y_i$. Then

$$P\left[\frac{M_n^{th} - b_n}{a_n} \leq x\right] = \left[F^{threshold}(a_n x + b_n)\right]^n$$
$$= \begin{cases} F^n(a_n x + b_n) & a_n x + b_n \geq T(n) \\ F^n(T(n)) & x_1 \leq a_n x + b_n < T(n) \\ 0 & a_n x + b_n < x_1 \end{cases}$$

As $n \to \infty$, by the hypothesis $\frac{T(n)-b_n}{a_n} \to -\infty$, so $a_n x + b_n \geq T(n)$ for all $x \in \mathbb{R}$, and thus $\left[F^{threhsold}(a_n x + b_n)\right]^n \to F^n(a_n x + b_n) \to G(x)$ as $n \to \infty$ since $F \in D(\Lambda)$. To show that this result holds regardless of which sequence of coefficients were chosen, suppose α_n and β_n are another sequence of coefficients satisfying $F^n(\alpha_n x + b_n) \to G(x)$. By Lemma 1, $\frac{\alpha_n}{a_n} \to 1$ and $\frac{\beta_n - b_n}{a_n} \to 0$. Substituting in the new coefficients yields

$$\frac{T(n) - \beta_n}{\alpha_n} = \left(\frac{1/a_n}{1/a_n}\right) \cdot \left(\frac{T(n) - \beta_n + (b_n - b_n)}{\alpha_n}\right)$$
$$= \left[\frac{1}{\frac{\alpha_n}{a_n}}\right] \cdot \left(\frac{T(n) - b_n}{a_n} - \frac{\beta_n - b_n}{a_n}\right)$$
$$\to \frac{T(n) - b_n}{a_n} \to -\infty$$

Thus any coefficients satisfying $F \in D(\Lambda)$ work.

$\underline{\Rightarrow}$: Suppose $F^{threshold} \in D(\Lambda)$. Then there exists $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\left[F^{threshold}(a_n x + b_n)\right]^n \to G(x) = e^{-e^{-x}}$ for all $x \in \mathbb{R}$. As before

$$\left[F^{threshold}(a_n x + b_n)\right]^n = \begin{cases} F^n(a_n x + b_n) & a_n x + b_n \geq T(n) \\ F^n(T(n)) & x_1 \leq a_n x + b_n < T(n) \\ 0 & a_n x + b_n < x_1 \end{cases}$$

Let $\lim_{n\to\infty} \frac{T(n)-b_n}{a_n} = c \in [-\infty, \infty]$, then in the limit as $n \to \infty$, the above expression becomes

$$\left[F^{threshold}(a_n x + b_n)\right]^n \to \begin{cases} G(x) & x \geq c \\ G(c) & x < c \end{cases}$$

where the left end point of the support need not be of concern because the left end point of the support of $G(x)$ is $-\infty$. The only way the above expression can converge to $G(x)$ for all $x \in \mathbb{R}$ is if $c = -\infty$, and thus $\lim_{n\to\infty} \frac{T(n)-b_n}{a_n} \to -\infty$, the desired result. □

Lemma 2 shows that if $T(n)$ scales sufficiently slowly, the convergence to the extreme value distribution $\Lambda(x)$ is unaffected. Having established Lemma 2, Theorem 8 can now be proven.

Proof. When $N = 1$, the distribution of the SINR is given by Equation 3.6 in Section 3.3. It is also shown in [SH05] that $F_{SINR} \in D(\Lambda)$ and that the so called *auxiliary function* f satifies

$$f(t) = \frac{1 - F_{SINR}(t)}{F'_{SINR}(t)} \to \rho$$

where $F'_{SINR}(t)$ is the density given by Equation 3.5. Thus $1 - F_{SINR}(x) \sim \rho F'_{SINR}(x)$. Using techniques from [Res87] (section 1.5), suitable coefficients a_n and b_n can be found so that $F^n_{SINR}(a_n x + b_n) \to G(x)$. The first step is to solve for b_n by solving $\rho F'_{SINR}(b_n) = n^{-1}$. Plugging in the density and taking the log of both sides produces

$$\begin{aligned} & -\log e^{-b_n/\rho} - M\log(1+b_n) + \log(1+b_n+\rho(M-1)) = -\log n \\ \Rightarrow\ & -\frac{b_n}{\rho} - M\log(1+b_n) + \log(1+b_n+\rho(M-1)) = -\log n \\ \Rightarrow\ & \frac{b_n}{\rho} + M\log(1+b_n) - log(1+b_n+\rho(M-1)) = \log n. \end{aligned}$$

As $n \to \infty, b_n \to \infty$, so $\frac{1}{\rho}b_n$ dominates the above expression and $\frac{1}{\rho}b_n \sim \log n \Rightarrow b_n \sim \rho \log n$. It is known, e.g. [Res87], that a suitable choice of a_n for this case is $f(b_n)$, where f is the aforementioned auxiliary function, thus a suitable sequence is $a_n = \rho$ for all n. The b_n sequence can be written as $b_n = \rho \log n + r_n$ where r_n is the remainder and $r_n \in o(\log n)$ in this case. Plugging in this formulation of b_n yields

$$\begin{aligned} & \tfrac{\rho\log n + r_n}{\rho} + M\log(1+\rho\log n + r_n) - \log(1+\rho\log n + r_n + \rho(M-1)) = \log n \\ & \Rightarrow r_n + \rho M\log(1+\rho\log n + r_n) - \rho\log(1+\rho\log n + r_n + \rho(M-1)) = 0 \end{aligned}$$

and solving for r_n produces

$$\begin{aligned} r_n &= -\rho M \log\log n - \rho M \log\log\left(\rho + \frac{1+r_n}{\log n}\right) + \\ &\quad \rho \log\log n + \rho \log\log\left(\rho + \frac{1+r_n+\rho(M-1)}{\log n}\right) \\ &= -\rho(M-1)\log\log n + o(1). \end{aligned}$$

Plugging the remainder expression into b_n yields

$$b_n = \rho \log n - \rho(M-1)\log\log n + o(1)$$

which is essentially the same expression found in [SH05], where it is called u_n. This sequence proved to be the key fact used in determining the sum-rate scaling result in their work.

With a suitable choice of $\{a_n\}$ and $\{b_n\}$ established, Lemma 2 can be utilized to determine the scaling rate of the threshold so that the asymptotic sum-rate throughput scales optimally. Observe that

$$\begin{aligned} \lim_{n\to\infty} \frac{h_n - b_n}{a_n} &= \lim_{n\to\infty} \frac{T(n) - (\rho\log n - \rho(M-1)\log\log n + o(1))}{\rho} \\ &= \lim_{n\to\infty} \left(\frac{T(n)}{\rho} - \frac{\rho\log n - \rho(M-1)\log\log n + o(1)}{\rho}\right) \\ &= \lim_{n\to\infty} \left(\frac{T(n)}{\rho} - (\log n - (M-1)\log\log n + o(1))\right) \\ &\to -\infty. \end{aligned}$$

Therefore $T(n) \in o(\log n)$ allows $\left[F_{SINR}^{threshold}(a_n x + b_n)\right]^n \to G(x)$ and all the conclusions derived in [SH05] using $F_{SINR} \in D(\Lambda)$ and $\{b_n\}$ hold. That is, selecting $T(n) \in o(\log n)$ still achieves the optimal sum-rate scaling rate. □

3.4.3 Necessary Conditions on the Scaling Rate of the Thresholding Function

In the previous section, it is shown how the threshold must scale in order that F remains in the domain of attraction $D(\Lambda)$. This provides a sufficient condition for the scaling rate of the threshold such that all previously derived results based on the extreme value distribution hold. The goal of this section is finding a necessary condition for the scaling rate of the threshold $T(n)$ such that the asymptotic scaling rate of the sum-rate throughput of the random beamforming scheme is optimal. To state it another way, it may be possible for $T(n)$ to grow faster than the rate required to keep $F \in D(\Lambda)$ and still have the sum-rate throughput of the truncated scheme be optimal.

A different method of analysis will be used to address the necessary conditions. The method is based upon the work in [RT73]. The following definition is essential for the analysis that follows:

Definition 2. *[RT73] Let a sequence of random variables $\{X_n, n \geq 1\}$ be unbounded above and let $M_n = \bigvee_{i=1}^{n} X_i$. The sequence $\{M_n\}$ is almost surely relatively stable (or stable for short) if there exist constants c_n such that $M_n/c_n \to 1$ almost surely as $n \to \infty$.*

In this section it will be shown that if $X_1, \ldots, X_n$ are drawn i.i.d. from F_{SINR}, then the maximum of the SINRs, M_n, is stable, and to find the associated constants c_n. With the known sequence c_n, another sequence d_n can be chosen such that $\lim_{n\to\infty} \frac{c_n}{d_n} \to 0$, or $c_n \in o\left(d_n\right)$. If d_n is substituted for c_n, then $\lim_{n\to\infty} \frac{M_n}{d_n} \to 0$ almost surely. This shows that the maximum of the SINRs is growing slower than the sequence d_n. If the thresholding function $T(n)$ is selected such that $T(n) \sim d_n$, eventually no user in the system will exceed the threshold and thus no users feed back. If no user feeds back SINR information, there is no channel state information at the transmitter, and in this case it is well known that multiuser diversity cannot be achieved. Thus, a threshold $T(n) \sim d_n$ will cause the random beamforming scheduling algorithm to asymptotically fail.

Having discussed how stability plays into the analysis of the random beamforming scheduling algorithm, the following theorem states how fast is too fast for the thresholding function to scale.

Theorem 9. *Let the SINR observed at each user be distributed i.i.d. according to F_{SINR}. Then if $T(n) \in \omega(\log n)$, asymptotically no user exceeds the threshold and consequently all multiuser diversity is lost.*

Proof. It must now be shown that M_n when $X_i \sim F_{SINR}$ is in fact stable and find the appropriate constants for stability. The following theorem of Resnick's provide necessary and sufficient conditions for stability.

Theorem 10. *[RT73] Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with common distribution function F, where $F(x) < 1$ for all x. There exist suitable normalizing constants $c_n, n \geq 1$ such that $M_n/c_n \to 1$ almost surely as $n \to \infty$ if and only if*

$$\int_1^\infty \frac{dF(x)}{1 - F((1-\epsilon)x)} < \infty$$

for all $1 > \epsilon > 0$.

The condition that $F(x) < 1$ for all x means that the random variable is unbounded from above, which is required in Definition 2. To verify Theorem 10 when $X_i \sim F_{SINR}$, substitute in the distribution function and density from Equations 3.5 and 3.6 into Theorem 10:

$$\begin{aligned}
\int_1^\infty \frac{\frac{e^{-x/\rho}}{(1+x)^M}\left(\frac{1}{\rho}(1+x)+M-1\right)}{1-\left(1-\frac{e^{-(1-\epsilon)x/\rho}}{(1+(1-\epsilon)x)^{M-1}}\right)}dx &= \int_1^\infty \frac{\frac{e^{-\frac{x}{\rho}}}{(1+x)^M}\left(\frac{1}{\rho}(1+x)+M-1\right)}{\frac{e^{-(1-\epsilon)x/\rho}}{(1+(1-\epsilon)x)^{M-1}}}dx \\
&= \int_1^\infty \frac{e^{-\frac{x}{\rho}\epsilon}}{(1+x)^M}(1+(1-\epsilon)x)^{M-1}\left(\frac{1}{\rho}(1+x)+M-1\right)dx \\
&\leq \int_1^\infty \frac{e^{-\frac{x}{\rho}\epsilon}}{(1+x)^M}(1+x)^{M-1}\left(\frac{1}{\rho}(1+x)+M-1\right)dx \text{ since } 0<\epsilon<1 \\
&= \int_1^\infty \frac{e^{-\frac{x}{\rho}\epsilon}}{1+x}\left(\frac{1}{\rho}(1+x)+M-1\right) < \infty
\end{aligned}$$

Thus M_n, the maximum of n SINR observations, is stable. In the proof of Theorem 10 in [RT73], it is shown that the coefficients $c_n \sim \mu_n \equiv F^{-1}(1 - n^{-1})$ are suitable for stability to hold. Since distribution functions need only be right continuous, the inverse is defined as $F^{-1}(x) = \inf\{y | F(y) \geq x\}$. In the SINR distribution of interest, the distribution function from Equation 3.6 is continuous. Similar to example (viii) in [RT73], the coefficients c_n will be found for this distribution. Define the function $R(x) = -\log(1 - F(x))$ and note $\mu_n \equiv$

$R^{-1}(\log n)$. Thus for F_{SINR}, $R(x)$ can be calculated:

$$\begin{aligned}
R(x) &= -\log(1 - F_{SINR}(x)) \\
&= -\log\left(\frac{e^{-x/\rho}}{(1+x)^{M-1}}\right) \\
&= -\log(e^{-x/\rho}) + \log(1+x)^{M-1} \\
&= \frac{x}{\rho} + (M-1)\log(1+x) \\
&\sim \frac{x}{\rho} \text{ as } x \to \infty
\end{aligned}$$

and therefore $R^{-1}(x) \sim \rho x$ and a suitable sequence c_n is $R^{-1}(\log n) = \rho \log n$. Thus, if the threshold function $T(n)$ is selected such that $\lim_{n\to\infty} \frac{\rho \log n}{T(n)} \to 0$, or $T(n) \in \omega(\log n)$, then eventually no user in the system will feed back and all multiuser diversity (and thus optimal sum-rate scaling) is lost, proving the theorem. □

Theorem 9, along with the conclusion from Section 3.4.2 provide an interesting result. Section 3.4.2 states that if $T(n) \in o(\log n)$, then optimal scaling still occurs, but the previous result states that if $T(n) \in \omega(\log n)$, that optimal scaling fails. Thus, any threshold $T(n) \in O(\log n)$ provides a kind of boundary between thresholds that allow optimal scaling and those that do not. When $T(n) \in O(\log n)$, stability is achieved, as shown before, and $\lim_{n\to\infty} \frac{M_n}{\log n} \to 1$ almost surely. In this case, it is unknown how the asymptotic scaling is effected.

3.5 Thresholds for a Finite Number of Users

The previous sections have established how fast thresholds can scale as a function of the number of users in the system so that asymptotically the sum-rate throughput still maintains the optimal scaling properties. While of theoretical interest, in any real system the number of users is finite. So how should the threshold be chosen in this case? The design of an appropriate threshold depends on the metrics that are to be optimized. This section discusses how to design the threshold for a system with n users to address the following metrics:

1. What should the threshold be so that the probability that no user feeds back information for any transmit beam is less than some design parameter γ_{outage}?

2. What should the threshold be so that on average only k users feed back SINR information for every transmit beam?

3. What should the threshold be so that the rate loss compared to the unthresholded system is less than some design parameter R_{loss}?

The first question is motivated by the fact that for any threshold, as long as one users feeds back SINR information for each transmit beam, then the maximum SINR is conveyed at the transmitter. In the event that the threshold is chosen so that no user feeds back SINR information for a transmit beam, multiuser diversity is lost on that beam, thus the probability that no user feeds back information needs to be controlled. The design parameter γ_{outage} is the "multiuser diversity outage probability", which is the maximum proportion of the time that the system is allowed to have no user feedback CSI information for any transmit direction. The second question will follow as a simple consequence of the analysis of the first question and constrains the average number of users feeding back. The final question is motivated by the fact that feedback is a precious resource, so perhaps one is willing to sacrifice some throughput to reduce the amount of feedback. As a function of the amount of throughput that is sacrificed for the reduction in feedback, an appropriate threshold can be designed. The following subsections provide solutions to the threshold design questions just raised.

3.5.1 Designing a Threshold under Outage Constraints

The first design criterion corresponds to each transmit beam having fed back SINR information $1 - \gamma_{outage}$ percent of the time. The system is in multiuser diversity outage if at least one transmit beam has no feedback information, and this situation should occur less than γ_{outage} percent of the time. The key to the analysis of designing a threshold to meet this requirement is to recall from Equation 3.10 the effect thresholding has on the underlying distribution function. Two feedback schemes will be considered. The first feedback scheme has each user feed back their observed SINR value for each transmit direction. The second feedback scheme has each user feed back only the largest observed SINR value over all transmit directions, and the index of the transmit beam that generated the maximum.

SINR Information for each Transmit Direction

First, consider the feedback scheme where information is fed back for each transmit beam and $N = 1$, i.e. the number of receive antennas per user is one. Since the SINR is always positive, if $0 \leq SINR < x_{threshold}$, then the user does not feed back the SINR information for that particular transmit beam, and this happens with probability $F_{SINR}(x_{threshold})$. For a system with n users and M transmit beams, let $Y_{i,j}, i \in \{1, \ldots, n\}, j \in \{1, \ldots, M\}$ be a Bernoulli random variable indicating that user i feeds back SINR information for transmit beam j. From the previous statement, $Y_{i,j}$ are i.i.d. with the following distribution:

$$\Pr\left[Y_{i,j} = 1\right] = 1 - \Pr\left[Y_{i,j} = 0\right] = 1 - F_{SINR}\left(x_{threshold}\right). \tag{3.13}$$

Define the random variable $Z_j = Y_{1,j} + \cdots + Y_{n,j}$, which denotes the total number of users that feedback information for transmit beam j. Since the $Y_{i,j}$ are i.i.d. for all i, Z_j is a binomial random variable, i.e.

$$Z_j \sim B\left(n, 1 - F_{SINR}\left(x_{threshold}\right)\right).$$

Outage occurs for transmit beam j when $Z_j = 0$, and $\Pr\left[Z_j = 0\right] = (1 - (1 - F_{SINR}(x_{threshold})))^n = \left(F_{SINR}(x_{threshold})\right)^n$ and the probability that the system is in outage is

$$\begin{aligned}
\Pr[\text{outage}] &= \gamma_{outage} = 1 - \Pr[\text{not in outage}] = 1 - \Pr\left[Z_1 \neq 0, \ldots, Z_N \neq 0\right] \\
&= 1 - \prod_{i=1}^{M} \Pr\left[Z_i \neq 0\right] = 1 - \prod_{i=1}^{M} \left(1 - \Pr\left[Z_i = 0\right]\right) \\
&= 1 - \left(1 - \left(F_{SINR}(x_{threshold})\right)^n\right)^M.
\end{aligned}$$

Thus, for a fixed number of users n, a fixed number of transmit antenna M, and

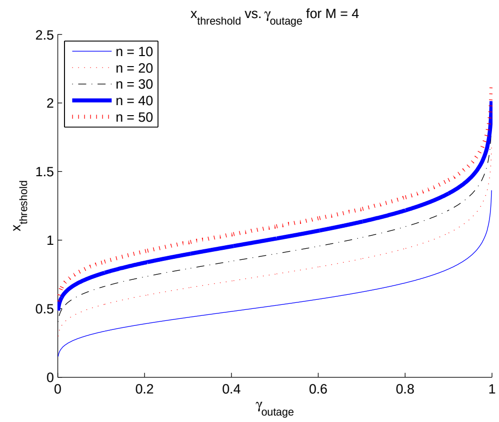

Figure 3.3: Threshold as a function of γ_{outage} for $M = 4$ and $\rho = 1$

design parameter $\gamma_{outage} \in [0, 1]$, the threshold can be determined:

$$\gamma_{outage} = 1 - (1 - (F_{SINR}(x_{threshold}))^n)^M$$
$$\Rightarrow \quad x_{threshold} = F_{SINR}^{-1}\left(\left(1 - (1 - \gamma_{outage})^{\frac{1}{M}}\right)^{\frac{1}{n}}\right)$$

where F_{SINR}^{-1} is the inverse distribution function. The solution to the inverse distribution function can be expressed in terms of Lambert functions. An alternative solution to evaluating the inverse is to use an iterative algorithm. Because the distribution function is a function of a scalar and monotonic, iterative algorithms are well suited for the computation of the inverse. Figure 3.3 shows the value of $x_{threshold}$ as a function of γ_{outage} for varying numbers of users in the system. As expected, for γ_{outage} approaching zero, the threshold also has to go to zero, and for γ_{outage} approaching one, the threshold begins to blow up. For larger numbers of users in the system, the threshold can be set higher and achieve the same outage probability due to the increased maximum order statistics.

If the receivers are allowed to have $N \geq 1$ number of receive antenna, as mentioned in Section 3.3.2, there are different ways to handle feedback. Originally in [SH05], one can consider each receive antenna as an independent user. In this case, the previous analysis follows verbatim except substituting n with nN since the system effectively has nN users under this methodology. Another approach considered in [PR10a] is to have each user perform LMMSE reception, and then

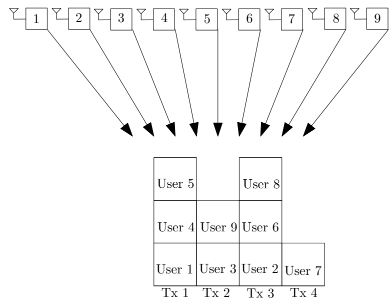

Figure 3.4: Example of Maximization at the Transmitter under Alternate Feedback Scheme

feedback the post-processed SINR for each transmit direction. In this case, the previous analysis is identical except substituting F_{SINR} with F_{MMSE} from Equation 3.9. In this case, the closed form solution of the inverse distribution function does not exist, and an iterative approach is required.

SINR Information for only the Best Transmit Direction

An alternative feedback scheme considered in [PR10a] has each user only feed back the maximum SINR across all transmit beams, as well as the index for the transmit beam that generated the maximum. Recall from Figure 3.1, this corresponds to feeding back the largest value in the array. For the time being, let $N = 1$. The number of SINR values fed back for a particular transmit beam is a random variable under this scheme. To help understand this randomness, consider Figure 3.4 which shows an example for a system with nine users and four transmit beams. Each user feeds back the largest SINR observed across all transit beams, so in this example, users 1,4 and 5 report back their SINR values and the fact that transmit beam 1 generated the maximum. The number of users that feed back SINR values corresponding to transmit beam 1 is not known a priori and is thus random. Since each user is restricted to feed back only one SINR value, and due to the symmetry of the system model, the number of SINR values for each transmit beam received at the transmitter is multinomially distributed. Using the traditional balls and bins analogy for the multinomial, the n users are the n balls, the M transmit beams are the M bins, yielding a probability vector

$p = \left[\frac{1}{M}, \ldots, \frac{1}{M}\right] \in \mathbb{R}^M$. The effect of thresholding on this problem is that rather than each user's feedback being in a bin associated with a transmit beam, that value may not actually be there if it did not exceed the threshold. Leveraging the balls and bins analogy again, each user can be thought of as putting a ball into a bin, but then there is a chance that the ball may be invisible and not counted due to the SINR not exceeding the threshold.

Recall the probability mass function of the multinomial of interest is

$$\Pr\left[Y_1 = y_1, \ldots, Y_M = y_m\right] = \begin{cases} \frac{n!}{y_1! \cdots y_M!} \left(\frac{1}{M}\right)^{y_1} \cdots \left(\frac{1}{M}\right)^{y_M} & \text{when} \sum_i^M Y_i = n \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

where Y_i are the number of users who sent back information for the i^{th} transmit beam. For any given realization of the multinomial, the probability of an outage event occuring is of interest. Let X_i, drawn from $F_{SINR_{(M)}}$ (as given by Equation 3.8), be the value that is to be fed back by the i^{th} user. In this problem, if $X_i < x_{threshold}$, then the ball is considered invisible. An outage event occurs when all of the balls in at least one bin are invisible. Suppose, for concreteness, that four users are to feed back information regarding transmit beam one. Then there are four balls in bin one, and $Y_1 = 4$. The case where all of the balls are invisible corresponds to the error event, and this event occurs with probability $1 - \left(F_{SINR_{(M)}}(x_{threshold})\right)^4$, where the exponent of four came from the fact that $Y_1 = 4$. This argument can be extended to all the other bins. Thus, conditioned on knowing the number of balls in each bin yields the following probability:

$$\begin{aligned} &\Pr\left[\text{not in outage} | Y_1 = y_1, \ldots, Y_M = y_m\right] \\ &= \prod_{i=1}^{M} \left(1 - \left(F_{SINR_{(M)}}(x_{threshold})\right)^{y_i}\right). \end{aligned} \tag{3.15}$$

Removing the conditioning produces the total probability of not being in outage

$$\begin{aligned} \Pr & \left[\text{not in outage}\right] \\ = & \sum \Pr\left[\text{not in outage} | Y_1 = y_1, \ldots, Y_M = y_m\right] \cdot \Pr\left[Y_1 = y_1, \ldots, Y_M = y_m\right] \\ = & \sum \Pr\left[Y_1 = y_1, \ldots, Y_M = y_m\right] \cdot \prod_{i=1}^{M} \left(1 - \left(F_{SINR_{(M)}}(x_{threshold})\right)^{y_i}\right) \end{aligned}$$

where the summation is taken over all possible configurations of the multinomial. With the probability of not being in outage calculated, the threshold can be solved for via the following equation:

$$1 - \gamma_{outage} = \sum \Pr\left[Y_1 = y_1, \ldots, Y_M = y_m\right] \cdot \prod_{i=1}^{M} \left(1 - \left(F_{SINR_{(M)}}\left(x_{threshold}\right)\right)^{y_i}\right). \tag{3.16}$$

The closed form inverse does not exist, so once again iteratve approaches must be taken. Equation 3.16 thus provides a way of finding the threshold under a multiuser diversity outage constraint when using a feedback algorithm that only feeds back local maxima. For the case where $N > 1$, the extension of the previous feedback scheme considered in [PR10a] takes the maximum over transmit beams and antennas to reduce the feedback per use to a single SINR value (and the beam index). In this case, the prior analysis follows substituting in $\left(F_{SINR_{(M)}}\right)^N$ for $F_{SINR_{(M)}}$, which arises due to the additional maximization over receive antennas.

3.5.2 Designing a Threshold Constraining the Average Number of Users Feeding Back

Because the SINR information is distributed across geographically separated receivers, it is impossible to design a threshold that can guarantee that exactly $k < n$ users feed back information for each transmit beam without further sharing of information. The best one can hope to do is use the statistical information about the channel metric, the observed SINR, to create a threshold that guarantees on average only k users feed back. First, consider the feedback schemes from Section 3.5.1 where each user feeds back SINR information for each transmit direction. Consider the case where $N = 1$. From Section 3.5.1 it is known that the probability of feeding back is a Bernoulli random variable, and the number of users feeding back for a particular transmit beam is a binomial random variable with parameters n and $1 - F_{SINR}\left(x_{threshold}\right)$. With these observations, the design of the threshold under the average number of users feeding back per transmit beam metric is straightforward. Once again, let $Z_j \sim B\left(n, 1 - F_{SINR}\left(x_{threshold}\right)\right)$ be the

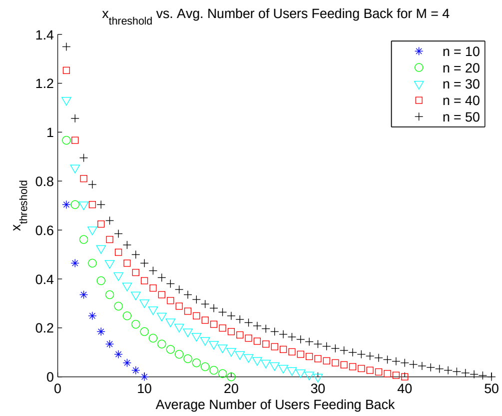

Figure 3.5: Threshold as a function of the Average Number of Users Feeding Back for $M = 4$ and $\rho = 1$

number of users feeding back for transmit beam j. Then

$$\begin{aligned} k &= E\left[\# \text{ of users feeding back for transmit beam} j\right] \\ &= E[Z_j] = n\left(1 - F_{SINR}\left(x_{threshold}\right)\right) \\ &\Rightarrow x_{threshold} = F_{SINR}^{-1}\left(\frac{n-k}{n}\right). \end{aligned} \tag{3.17}$$

Figure 3.5 shows the threshold as a function of the constraint on the average number of users feeding back for varying number of users in the system. As the average number of users allowed is increased, the threshold decreases as expected to accomodate more users exceeding the threshold. When the average number of users equals the number of users in the system, the threshold is equal to zero.

If one is interested in constraining the average total amount of feedback to k (rather than on average k values per transmit beam), by the symmetry of the problem, substitute k with $k' = \frac{k}{M}$ in Equation 3.17. This is saying that each transmit direction gets on average $\frac{k}{M}$ values, so for all M transmit beams, the total on average is k. These thresholds extend to the case when $N > 1$ in two ways. If one considers feeding back the maximum observed SINR over the receive antennas for each transmit direction, then the previous analysis holds substituting $F_{SINR}(x)$ with $(F_{SINR}(x))^N$. If one allows for LMMSE reception, the previous results hold by using the distribution function given in Equation 3.9.

As in Section 3.5.1, consider the feedback scheme where only the local maximum SINR over the transmit beams is fed back. Recall that $F_{SINR_{(M)}}$ denotes the distribution function of the SINR values fed back under this scheme. The analysis in this case is similar to the previous case, but now one must consider that the transmit beam whose SINR value is being fed back is a random variable that is uniform over the transmit beams, i.e. every transmit direction is equally likely to produce the local maximum SINR. From Figure 3.1, this corresponds to the fact that when finding the maximum value in the array, it is equally likely to occur in any given column. Let Z_j' be the number of values fed back for transmit beam j. Then $Z_j' \sim B\left(n, \frac{1-F_{SINR}(x_{threshold})}{M}\right)$, where the factor of $\frac{1}{M}$ comes from the previously mentioned fact that associated transmit beam is a uniform random variable over the M transmit beams. The threshold design is given by

$$\begin{aligned} k &= E\left[\text{\# of users feeding back for transmit beam}\right] \\ &= E[Z_j'] = n\left(\frac{1 - F_{SINR_{(M)}}\left(x_{threshold}\right)}{M}\right) \\ &\Rightarrow x_{threshold} = F^{-1}_{SINR_{(M)}}\left(1 - \frac{kM}{n}\right). \end{aligned} \tag{3.18}$$

Notice that in this case it is required that $k \leq \frac{n}{M}$. This constraint makes sense under this feedback scheme since without thresholding, on average there will be $\frac{n}{M}$ values fed back per transmit beam. These results can be extended to the case where $N > 1$ and the maximum SINR is taken over transmit directions and receive antennas in a straight forward manner. Substituting $\left(F_{SINR_{(M)}}\right)^N$ for $F_{SINR_{(M)}}$ produces $x_{threshold} = F^{-1}_{SINR_{(M)}}\left(1 - \frac{kM}{n}\right)^{\frac{1}{N}}$.

3.5.3 Designing a Threshold with Rate Loss Constraints

A third design constraint of interest is to select a threshold that bounds the rate lost in the system as compared to the system without a threshold. Let R be the throughput of the unthresholded system and consider the feedback scheme where $N = 1$ and the SINR is fed back for each transmit direction as in Section 3.5.1. From [PR10a], it is known for this scheme that R can be expressed as

$$R \approx M \int_0^\infty \frac{1}{1+x}\left(1 - \left(F_{SINR}(x)\right)^n\right) dx, \tag{3.19}$$

where the approximation symbol becomes very accurate for a moderate number of users in the system. Equation 3.19 is equivalent to writing out the integral equation to evaluate
$ME\left[\log\left(1+\max\left\{X_1,\ldots,X_n\right\}\right)\right]$, where the X_i are drawn i.i.d. from F_{SINR}. The exponent of n on the distribution function is the effect of the maximum order statistics. The sum-rate throughput of the thresholded system is more complicated. The thresholded distribution given in Equation 3.10 cannot be directly plugged into the above Equation 3.19. This is because the thresholded distribution defined maps the rate to $x_l = 0$ when the SINR does not exceed the threshold. This mapping works when analyzing the question of optimal sum-rate scaling because the primary concern is whether multiuser diversity is achieved or not. With a finite number of users, even if no user exceeds the threshold and thus there is no feedback, this does not mean the system achieves zero throughput. If no user feeds back for a particular direction, the scheduler can randomly select a user and transmit to it. This random selection does not benefit from multiuser diversity since there is no CSI, but it can achieve a non-zero throughput.

The thresholded sum-rate throughput consists of two parts: the rate when a transmit direction receives CSI and the rate when no user feeds back information for a transmit direction and thus a random user is scheduled on that direction. Let $F_{threshold}$ be the distribution of the SINR for a transmit direction under the scheme where it schedules the maximum user if at least one user feeds back SINR information, and schedules a random user if no CSI is fed back. Thus, as with Equation 3.10, when at least one user feeds back CSI information for a transmit direction, the true maximum is observed and the thresholded distribution is given by

$$F_{threshold}(x) = F_{SINR}^n \text{ for } x \in [x_{threshold}, \infty). \tag{3.20}$$

Now consider the situation when no CSI is fed back. Conditioned on the fact that no SINR value exceeds the threshold,when the scheduling algorithm selects a user at random to transmit to, from [Dav70] the conditional distribution of the SINR of the randomly selected user is given by

$$F_{random}(x) = \frac{F_{SINR}(x)}{F_{SINR}(x_{threshold})} \text{ for } x \in [0, x_{threshold}). \tag{3.21}$$

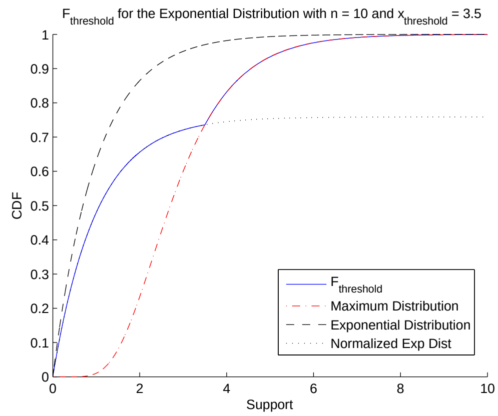

Figure 3.6: Thresholded Distribution Function for the Exponential Distribution with $n = 10$ and $x_{threshold} = 3.5$

The distribution is defined on $[0, x_{threshold})$ since this distribution occurs when no users exceeds the threshold. The denominator in Equation 3.21 renormalizes the distribution F_{SINR} so that F_{random} is a true distribution function. The probability that no user exceeds the threshold is given once again by $F_{SINR}(x_{threshold})^n$. Putting together these facts with Equation 3.20 produces the total distribution function for this thresholded scheme:

$$F_{threshold} = \begin{cases} F_{SINR}(x)^n & \text{for } x \in [x_{threshold}, \infty) \\ F_{SINR}(x_{threshold})^n \cdot F_{random}(x) & \text{for } x \in [0, x_{threshold}) \end{cases} \tag{3.22}$$

Figure 3.6 verifies this equation with the exponential distribution for $n = 10$ and $x_{threshold} = 3.5$. Substituting the distribution in Equation 3.22 for $F_{SINR}(x)^n$ in Equation 3.19 provides the sum-rate throughput of this scheme $R_{threshold}$:

$$R_{threshold} \approx M \int_0^\infty \frac{1}{1+x} \left(1 - F_{threshold}(x)\right) dx, \tag{3.23}$$

where the integral can be multiplied by M since all the transmit beams statistics are independent and identically distributed.

Define the throughput of the system when random users are selected on each transmit beam as R_{random}. Then from Equation 3.19, R_{random} is given by

$$R_{random} \approx M \int_0^\infty \frac{1}{1+x} \left(1 - F_{SINR}(x)\right) dx. \tag{3.24}$$

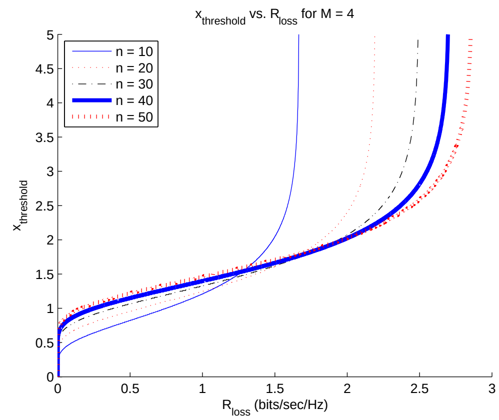

Figure 3.7: Threshold as a function of R_{loss} for $M = 4$ and $\rho = 1$

Let $R_{loss} = R - R_{threshold} \in [0, R_{random}]$ be the design parameter that quantifies the amount of rate that can be sacrificed in order to reduce feedback. The upper bound on R_{loss} is R_{random} since in the worse case scenario this throughput can be achieved. Since F_{SINR} for this system is continuous and monotonically increasing, R_{loss} is monotonic increasing as a function of $x_{threshold}$ and there exists one solution for $x_{threshold}$ that solves the above equation for R_{loss}. R_{loss} is a complicated function of $x_{threshold}$, but due to its monotonicity, an iterative algorithm can be used to find $x_{threshold}$ to sufficient accuracy. Figure 3.7 shows the threshold as a function of the design parameter R_{loss}. For small values of R_{loss}, the larger the number of users, the higher the threshold can be set and still achieve the same rate loss, which is expected since for a larger number of users, the probability of exceeding the threshold increases. All the thresholds eventually approach infinity as R_{loss} approaches the unthresholded sum-rate throughput given by Equation 3.19.

Adapting the preceeding arguments for the feedback schemes considered in Section 3.5.1 when $N > 1$ is straightforward. For the case where each antenna is considered an individual user, change the exponent of F_{SINR} in Equation 3.22 from n to nN. When LMMSE receivers are implemented, substitute F_{MMSE} in Equations 3.21 and 3.22. The more difficult scheme to analyze is when only the local maximum SINR values and the associated beam index are fed back as in Section 3.5.1. Recall the multinomial distribution given by Equation 3.14. Define

Y_i' to be the number of users that feed back CSI for the i^{th} transmit beam and exceed the threshold. Thus, Y_i' is equal to Y_i minus the the number of values that would have been fed back but did not exceed the threshold. The difficulty in analyzing the rate of this scheme from a theoretical perspective is two-fold. First, $\sum Y_i' \leq n$ since some values may not exceed the threshold, and hence the Y_i' do not form a multinomial. This is a problem since the exponent of the distribution is no longer known. For example, consider a system with $n = 20$ users, and suppose $\sum Y_i' = 15$ and for concreteness $Y_1 = 5$. When theoretically analyzing the rate, the exponent for the first transmit beam could have been the maximum of five values that were fed back, or it could have been a maximum of up to 10 values ($Y_1 + n - \sum Y_i'$), since there is no information available on five of the users. The second problem is that if no information is available for a specific transmit direction, then a random user is scheduled. Suppose that the user randomly selected fed back CSI information, but for a different transmit direction. For example, suppose $N = 1$ and the SINR values at the randomly selected user are 1 and 2 for transmit directions 1 and 2 respectively. If the threshold was $x_{threshold} = 0.5$, both exceed the threshold, but since only the local maximum is fed back, the value of 2 on the first transmit direction is never observed. Thus, when this user is randomly selected for the first transmit direction, the SINR value is greater than the threshold, but is conditionally known to be less than 3 since the that value is known at the scheduler. Therefore the distribution of the randomly selected users' SINRs is quite complicated. For these reasons, the analysis will not be performed on this scheme. Simulations can be used to design an appropriate threshold.

3.6 Conclusion

This chapter addresses the question of which user should feed back the observed SINR given statistical knowledge of the channel metric. When every receiver feeds back their SINR information, it is shown in [SH05] that random beamforming achieves the optimal sum-rate throughput scaling. In [PR10a], it is shown that any finite threshold such that users experiencing SINRs below this threshold do not feed back does not affect this optimal scaling rate. The questions raised in this chapter discuss how the thresholds can be designed as a function of the number of users in the system to achieve certain performance metrics.

The first performance metric considered is how fast can the threshold grow yet still exhibit the optimal scaling. It is shown that any thresholding function $T(n)$ as a function of the number of users n that grows slower than $\log n$, i.e. $T(n) \in o(\log n)$, achieves the optimal scaling rate. Conversely, any thresholding function $T(n) \in \omega(\log n)$ causes the system to lose all multiuser diversity.

The second part of the chapter considers how to optimally design thresholds for systems with a finite number of users. Using the statistics of the channel metric, thresholds for an n user system under three different design criterion are found. Under the first criterion, a threshold is designed to limit the probability that no user in the system feeds back CSI information for any transmit beam to be less than the design parameter γ_{outage}. The second criterion considers choosing a threshold such that on average only k out of n users in the system feed back CSI information for each transmit beam. Finally, a threshold is designed to limit the rate lost compared to a random scheduling algorithm to be below a design parameter R_{loss}.

The material in this chapter is work which in preparation for submission to the *IEEE Transactions on Signal Processing* under the title *"Reducing Feedback in Broadcast Channels via Thresholding"* and material submitted to *Asilomar Conference on Signals, Systems, and Computers 2011* under the title *"Feedback Reduction by Thresholding in Multi-User Broadcast Channels: Design and Limits"*. Both of these works are coauthored with Professor Bhaskar D. Rao.

Chapter 4

Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models

4.1 Outline

There exists a trade-off between performance and fairness in scheduling algorithms. A greedy scheduling algorithm will schedule the user with the best rate during each epoch to maximize the throughput of the system at the expense of neglecting fairness considerations. In an attempt to balance fairness and performance the proportional fair sharing (PFS) algorithm has been proposed. In this chapter, it is shown that when the rates of each user are i.i.d., the asymptotic performance of the greedy and PFS scheduling algorithms are equivalent. Under the i.i.d. model, the mean asymptotic throughput of the PFS algorithm is characterized and the rate of convergence to this throughput is derived. Additionally, the asymptotic distribution of the distance of the PFS algorithm from the asymptotic throughput is found. Because the rates of each user must be fed back to the base station to make the scheduling decision, a thresholding mechanism is considered to reduce the amount of feedback overhead in the system. The effects of the thresholding mechanisms on the PFS algorithm and the asymptotic performance is shown. This chapter contains material in preparation for submission to *IEEE Transactions on Information Theory* under the title *"Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models"* coauthored by Professor Bhaskar D. Rao.

4.2 Introduction

Scheduling users in a cellular downlink channel is an important question whose solutions balance system throughput and fairness. Suppose there are n users in a given cell and the base station has M transmit antennas and $n > M$. In this case, at most a subset of M of the n users can be scheduled for simultaneous transmission by the base station if all the users are to receive independent signals. The central question addressed by any scheduling algorithm is how to select the $k \leq M$ users to transmit to during any given scheduling epoch.

The scheduling algorithm can often be intimately related to the transmission scheme. For example in the semi-orthogonal user selection (SUS) algorithm proposed in [YG06b], the transmission scheme and scheduling algorithm are one in the same. In the SUS algorithms, the first iteration identifies the user with the best channel conditions and the corresponding transmit direction. In subsequent iterations of the algorithm, the users with the best channel conditions and transmit directions that are nearly orthogonal to the previously selected users are chosen. The data for each selected user is transmitted along their beamforming directions, which by construction are nearly orthogonal. Thus, the SUS algorithm simultaneously considers the transmission and scheduling procedure. Another transmission protocol is the random beamforming scheme considered in [SH05]. This scheme has the base station generate a random M-dimension complex orthonormal basis to use as the transmit beamformers. Each user measures the SINR along the basis vectors and feeds back these measurements to the base station. At this point, a greedy scheduling algorithm is used that schedules the user with the largest SINR on each transmit basis vector.

These algorithms attempt to maximize the sum-rate throughput of the downlink. As such, fairness considers are not explicitly considered. For example, in the the random beamforming scheme, a particular user may experience an arbitrarily long wait time. The Rayleigh fading model assumed in [SH05] however guarantees that asymptotically each user will be scheduled an equal proportion of time. An alternative to greedy scheduling is the round robin scheduler which deterministically schedules a user once every n scheduling epochs. The round robin method guarantees fairness, but the system performance is degraded. Any scheduling algorithm that is agnostic to the channel conditions of the users cannot exploit multi-user diversity. Therefore there is a fundamental trade-off between fairness and performance. The random beamforming method is known to exploit multi-user diversity and exhibit optimal throughput scaling asymptotically in the number of users in the broadcast channel where as the round robin method is fair but the throughput does not scale with the number of users.

Greedy algorithms are very attractive due to their ease of implementation and good performance but ignores any fairness criterion. The proportional fair sharing (PFS) algorithm is a scheduling algorithm that tries to balance performance and fairness. The PFS algorithm forms a ratio of each user's current rate to their average scheduled rate and selects the user with the largest ratio. To see that this provides some notion of fairness, consider a user that has rarely been scheduled in the past. Then that user's average scheduled rate is small, causing the previously mentioned ratio to become large, and increase the likelihood of being scheduled.

This chapter shows that when the rates are i.i.d., under suitable conditions on the distribution, the PFS algorithm performance converges to the performance of the greedy algorithm in time. In the appropriate state space, the convergence is equivalent to showing the two algorithms converge to the same stationary limit point. The rate of convergence of the PFS algorithm to the limit point is characterized as well as the asymptotic normality and covariance of the PFS algorithm around the limit point. These results imply that the PFS algorithm can be used to ensure some degree of fairness, and it is known that asymptotically the same performance as the greedy algorithm will be attained. These results are shown by leveraging the theory of stochastic approximations, as is done in the motivating work [KW04].

Because the PFS algorithm is a function of the current rate of each user, each user must provide channel state information (CSI) feedback to the base station in order to make the scheduling decision. In this chapter it is assumed that the rate information itself is fed back to the base station, and thus schemes where the transmit and scheduling algorithms are decoupled, as in random beamforming, are considered. Motivated by our previous work [PR10b], CSI feedback reduction can be achieved by comparing a function of each user's rate to a threshold, and if the threshold is exceeded the user feeds back its rate information, otherwise the user provides no feedback. CSI reduction is of interest since transmit power can be saved as well as relieving congestion on the radio resource. A thresholding scheme is considered under the PFS scheduling algorithm and is shown to converge to a stationary limit point which is analytically characterized.

The organization of the chapter is as follows: the proportional fair sharing algorithm is described in detail in Section 4.3. The convergence of the PFS algorithm to the greedy algorithm when the rates are bounded is considered in Section 4.4 and the convergence results are extended to the case where the rates are unbounded in Section 4.5. The rate of convergence to the fixed stationary point as well as asymptotic covariance structure around the fixed point are shown in 4.6. Section 4.7 analyzes the effects of a thresholding mechanism applied to the rate information feedback on the fixed point of the PFS algorithm. Section 4.9 summarizes the results and concludes the chapter.

4.3 The Proportional Fair Sharing Algorithm

The general proportional fair sharing problem formulation will follow the set-up established in [KW04]. The results of [KW04] show that under general conditions the PFS algorithm converges, but does not address where it converges to due to the generality of the conditions. The primary goal of this chapter is to show that under an i.i.d. model, the convergence is to the performance of the greedy algorithm.

Let there be N users in the system and let $X_{i,n}$ be the instantaneous rate of user i at time slot n. The $X_{i,n}$ are assumed to be i.i.d. continuous, non-negative random variables, since the rates cannot be negative, for all $1 \leq i \leq N$ and $n \in \mathbb{N}^+$. At time $n+1$, let the rates $X_{i,n+1}$ be known at the scheduler for all i. Then the indicator random variable $I_{i,n+1}$ denotes the event that user i is scheduled for time slot $n+1$. The average throughput of the i^{th} user after n time slots is $\theta_{i,n}$ and is defined as follows:

$$\theta_{i,n} = \sum_{l=1}^{n} \frac{X_{i,l} I_{i,l}}{n}. \tag{4.1}$$

Because $I_{i,l} = 1$ only when user i is scheduled for time slot l, the sum in Equation 4.1 is the average scheduled rate for user i. Since the rates $X_{i,n+1}$ are known at the scheduler at time $n+1$, the PFS algorithm selects the user with the largest ratio $\frac{X_{i,n+1}}{\theta_{i,n}}$, or formally the algorithm performs the following function

$$\arg \max_{1 \leq i \leq N} \left\{ \frac{X_{i,n+1}}{\theta_{i,n}} \right\}. \tag{4.2}$$

When the PFS algorithm starts, all the initial average rates $\theta_{i,n}$ are zero, and the ratio in Equation 4.2 is undefined. This was observed in [KW04] and a solution is to modify the ratio slightly. Let d_i be positive constants for $1 \leq i \leq N$, which can be arbitrarily small. Define the modified PFS algorithm as performing the following function:

$$\arg \max_{1 \leq i \leq N} \left\{ \frac{X_{i,n+1}}{d_i + \theta_{i,n}} \right\}. \tag{4.3}$$

The ratio defined in Equation 4.3 is always well defined since the rates are non-negative. In this chapter it is assumed that $d_i = d$ for all i and that d is incredibly small, and is considered only to be needed so that the ratio is well defined during the initial phases of the algorithm. That is until every user is scheduled at least once so that average throughput term is non-zero for every user. Because the rates $X_{i,n}$ are assumed to be continuous, the probability of there being a tie in either Equation 4.2 or 4.3 is zero.

Having established the PFS algorithm, the next sections will address the convergence under the i.i.d. model to the greedy algorithm. The greedy algorithm selects the user with the largest instantaneous throughput at every time slot n, i.e. the schedule function for time slot $n+1$ is given by

$$\arg\max_{1\leq i\leq N} \{X_{i,n+1}\} \tag{4.4}$$

The approach to prove the convergence is based on the ordinary differential equation (ODE) method of stochastic approximation (SA) as found in [KY03]. First, the convergence is shown when the rate random variables $X_{i,n}$ are bounded and then the results will be extended to rates which are unbounded from above. The extension to unbounded random variables is motivated by the Trajectory-Subsequence (TS) approach to the SA problem found in [Che02].

4.4 Convergence with Bounded Rates

This section is comprised of two parts. The first section discusses how the PFS algorithm can be put into the framework of a stochastic approximation algorithm. Much work has been done to characterize convergence properties of stochastic approximation algorithms and the key points of the ordinary differential equation method of approaching stochastic approximation algorithms are discussed. With the appropriate machinery described, the second part of this section discusses the convergence of the PFS algorithm to the performance of the greedy algorithm asymptotically in time.

4.4.1 The Stochastic Approximation Approach

In Sections 4.4.1 and 4.4.2 the non-negative i.i.d. rate random variables $X_{i,n}$ are assumed to be bounded from above, i.e. there exists some constant K such that $\Pr(X_{i,n} > K) = 0$ for all i and n. The key to showing the convergence result is to put the PFS algorithm into the framework of a stochastic approximation algorithm, and leverage the theory of SA algorithms. The general form of the SA algorithm as defined in [FP96] will now be stated. Let $(\omega^n)_{n\geq 1}$, $(\eta^n)_{n\geq 1}$ be two sequences of $\mathbb{R}^p$- and $\mathbb{R}^d$- valued vectors respectively and let $H : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^d$.

Define the sequence $(\theta^n)_{n\geq 1}$ by the recursive algorithm

$$\begin{cases} \theta^0 \in \mathbb{R}^d \\ \theta^{n+1} = \theta^n + \epsilon_{n+1} H\left(\theta^n, \omega^{n+1}\right) + \epsilon_{n+1}\eta^{n+1}, \quad n \in \mathbb{N}^+, \end{cases} \tag{4.5}$$

where $(\epsilon_n)_{n\geq 1}$ is a sequence of positive real numbers called the gain of the algorithm and η^n is a small residual perturbation.

Having established the general form of the SA algorithm, the PFS algorithm can be put into this framework. Let $\theta^n = [\theta_{1,n}, \ldots, \theta_{N,n}]^T \in \mathbb{R}^N$ be the vector of average throughputs, where it is understood that superscripts are time indices for vector quantities and subscripts indicate elements of a vector at a particular time index. Likewise, let $X^n = [X_{1,n}, \ldots, X_{N,n}]^T \in \mathbb{R}^N$ and let $I^{n+1}(\theta^n, X^{n+1}) \in \mathbb{R}^{N\times N}$ be an $N \times N$ matrix of zeros except for diagonal element $I_{ii,n+1} = 1$ if user i is scheduled for time slot $n+1$. The indicator function $I^{n+1}(\theta^n, X^{n+1})$ is explicitly written as a function of the vector of average throughputs θ^n up to time n and current rates X^{n+1} since these values determine which component is non-zero by Equation 4.3. The vector of average throughputs can be written in the recursive form of Equation 4.5. Let $\omega^{n+1} = X^{n+1}$, $\epsilon_{n+1} = \frac{1}{n+1}$, and the perturbation sequence $\eta^{n+1} = 0$ for all n. Let it also be assumed that the initial average throughputs are zero for every user, i.e. $\theta^0 = \mathbf{0} \in \mathbb{R}^d$.Then the following equation gives the desired recursive relationship:

$$\begin{aligned} \theta^{n+1} &= \theta^n + \epsilon_{n+1} H\left(\theta^n, X^{n+1}\right) \\ &= \theta^n + \epsilon_{n+1}\left[I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n\right] = \theta^n + \epsilon_{n+1} Y^n, \end{aligned} \tag{4.6}$$

where $H\left(\theta^n, X^{n+1}\right) = [I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n] = Y^n$. Equation 4.6 is established in [KW04].

The ODE Method for solving SA algorithms attempts to show that the random SA algorithm generally behaves like a dynamical system, and that asymptotically the SA algorithm converges to a stationary point or stationary set of a set of differential equations. Before putting the SA algorithm of Equation 4.6 into a form suitable for the ODE method, some definitions are required. Let h be called the *mean function*, and in the i.i.d. setting, as mentioned in [FP96], it can be defined as:

$$h(\theta) \triangleq \int H(\theta, x)\mu(dx) \tag{4.7}$$

where $X \in \mathbb{R}^N$ have common distribution μ, $\theta \in \mathbb{R}^N$ and $H(\theta, \cdot)$ is μ-integrable for every $\theta \in \mathbb{R}^N$. The interpretation of the mean function is given that the current location in the θ-space is θ, on average one will move in the direction $h(\theta)$. Let $(\mathcal{F}_n)_{n\geq 0}$ be the filtration generated by θ^0 and $(X^n)_{n\geq 1}$. Then letting the random vector of the current average throughputs θ^n be the argument of the mean function produces

$$h(\theta^n) = \mathbb{E}\left(H(\theta^n, X^{n+1})|\mathcal{F}_n\right). \tag{4.8}$$

The right hand side of Equation 4.8 says where the algorithm is going to move on average in the next time step given the past (the information contained in the filtration $\mathcal{F}_n$). Lastly, define a martingale difference sequence (MDS) as follows:

Definition 3. *Let $(X^n)_{n\geq 0}$ and $(Y^n)_{n\geq 1}$ be two $\mathbb{R}^d$-valued random sequences and let $(\mathcal{F}^n)_{n\geq 1}$ be the filtration generated by $(X^n)_{n\geq 1}$. The sequence $(Y^n)_{n\geq 1}$ is called a martingale difference sequence if*

$$\mathbb{E}(Y^{n+1}|\mathcal{F}^n) = 0 \textit{ for all } n \in \mathbb{N}.$$

Having established the preceding definitions, Equation 4.6 can be written in the following form suitable for analysis by the ODE method:

$$\theta^{n+1} = \theta^n + \epsilon_{n+1}h(\theta^n) + \epsilon_{n+1}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n - h(\theta^n)\right) \tag{4.9}$$

$$= \theta^n + \epsilon_{n+1}h(\theta^n) + \epsilon_{n+1}\delta M^{n+1} \tag{4.10}$$

where $h(\theta^n)$ is the mean function and δM^{n+1} is a MDS. To see that δM^{n+1} is a MDS follows from the definition of the H, mean function h and using the filtration $(\mathcal{F}^n)_{n\geq 0}$ generated by θ^0 and $(X^n)_{n\geq 1}$:

$$\begin{aligned}
&\mathbb{E}\left(\delta M^{n+1}|\mathcal{F}^n\right) = \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n - h(\theta^n)|\mathcal{F}^n\right) \\
&= \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n - \mathbb{E}\left(H(\theta^n, X^{n+1})|\mathcal{F}_n\right)|\mathcal{F}^n\right) \\
&= \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n - \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n|\mathcal{F}_n\right)|\mathcal{F}^n\right) \\
&= \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n|\mathcal{F}^n\right) - \mathbb{E}\left(I^{n+1}(\theta^n, X^{n+1})X^{n+1} - \theta^n|\mathcal{F}_n\right) \\
&= 0 \text{ for all } n \in \mathbb{N}.
\end{aligned}$$

The fact that Equation 4.6 could be written in the form of the sum of a mean function and a MDS is not surprising. As pointed out in Chapter 5 of [KY03], this form naturally arises when Y^n can be written in the form $H(\theta^n, X^n)$ and the X^n are mutually independent, as is the case here. If the conditions of the celebrated Kushner-Clark Theorem ([KY03, FP96]) can be verified, the result of the theorem says that almost surely the limit of sample paths θ^n are trajectories of the differential equation

$$\dot{\theta} = h(\theta) \tag{4.11}$$

The key implication of the Kushner-Clark Theorem is that the SA algorithm can asymptotically be analyzed using the theory of differential equations and dynamical systems. The distribution of the $(\theta^n)_{n\geq 1}$ is not known in closed form, but it is still possible to analyze its performance due to the Kushner-Clark Theorem. Additionally, if it can be shown that there is a unique singleton $\{\theta^*\}$ such that regardless of the initial condition θ^0, the solution to Equation 4.11 is θ^*, then the Kushner-Clark Theorem also says that $(\theta^n)_{n\geq 1}$ converges almost surely to θ^* regardless of the initial condition. From [Ben96], let $\Phi(t, x) = \Phi_t(x)$ be the flow induced by Equation 4.11. A point θ is an *equilibrium point* if $\Phi_t(\theta) = \theta$ for all $t \in \mathbb{R}$. Another way of stating the second conclusion is that if θ^* is the unique global attractor, i.e. $\lim_{t\to\infty} \Phi_t(\theta^0) = \theta^*$ regardless of the initial condition, then the SA algorithm almost surely converges to θ^*.

4.4.2 Convergence of the PFS Algorithm with Bounded Rates

The goal of this section is to show that under bounded non-negative i.i.d. random variables the PFS algorithm converges to greedy algorithm. This result will be shown via the following steps. First, the conditions of the Kushner-Clark Theorem will be shown to hold so the solution of the SA algorithm converges to a solution of the differential equation given in Equation 4.11. Next, a theorem in [KW04] will be stated showing that there exist a unique global attractor, and thus the SA algorithm converges almost surely to θ^*. Lastly, it will be shown that

θ^* corresponds to the solution of the greedy algorithm, so the two algorithms are asymptotically equivalent. It should be pointed out that under the i.i.d. model which yields Equation 4.9, almost sure convergence results are attainable, as opposed to the weak convergence results found in [KW04] which consider more general random sequences $(X^n)_{n\geq 1}$.

The first task is to show that the conditions required of the Kushner-Clark Theorem are satisfied. There are six conditions that must be satisfied and they are listed as follows ([KY03])

1. $\sum_{n=0}^{\infty} \epsilon_n = \infty, \epsilon_n \geq 0, \epsilon_n \to 0$ for $n \geq 0$.
2. $\sup_n \mathbb{E}_n |Y_n|^2 < \infty$.
3. There is a measurable function $\overline{g}(\cdot)$ of θ and random variables β_n such that
$$\mathbb{E}_n Y_n = \mathbb{E}\left[Y_n | \theta_0, Y_i, i < n\right] = \overline{g}(\theta_n) + \beta_n.$$
4. $\overline{g}(\cdot)$ is continuous.
5. $\sum_i \epsilon_i^2 < \infty$.
6. $\sum_i \epsilon_i |\beta_i| < \infty$ w.p.1.

Conditions 1 and 5 are immediate since $\epsilon_{n+1} = \frac{1}{n+1}$ and Condition 3 holds with $\overline{g} = h$, the mean function defined in Equation 4.7. From the mean function, notice that $\beta_n = 0$ for all n, so Condition 6 is trivially true. Condition 2 is true since the sequence of random variables $(X^n)_{n\geq 1}$ are bounded. The last condition to check is that the function $\overline{g} = h$ is continuous. In [KW04], it is shown that $\overline{g}$ is in fact Lipschitz continuous, a condition stronger than continuity, as a consequence of $(X^n)_{n\geq 1}$ being bounded. Therefore, the Kushner-Clark Theorem can be applied to the SA algorithm and the solution converges to a trajectory of the ODE given in Equation 4.11.

The existence of a unique global attractor θ^* for Equation 4.11 is given by Theorem 2.2 in [KW04]. The proof of the theorem is based on dynamical systems theory, more specifically the monotonicity property of the solution to Equation

4.11. An overview of monotone dynamical systems can be found in [Smi95]. Applications of monotone dynamical systems theory in a stochastic approximation setting can be found in [BH99a, BH99b]. Combining the result of the Kushner-Clark Theorem with Theorem 2.2 in [KW04] says that the PFS algorithm converges to a unique equilibrium point θ^* regardless of the initial condition θ^0 almost surely.

Since it is known that the PFS algorithm converges to a unique θ^* almost surely, the last step is to show that θ^* corresponds to the average throughput given by the greedy algorithm. First, the average throughput of the greedy algorithm under the i.i.d. model is established in the following lemma:

Lemma 3. *Let $X_{1,n}, \ldots, X_{N,n}$ be i.i.d. non-negative bounded random variables for all $n \in \mathbb{N}^+$. Define $M_{N,n} = \max\{X_{1,n}, \ldots, X_{N,n}\}$, the maximum order statistic at time n and let $\tilde{\theta}_{i,n}$ be the average throughput of user i at time n under the greedy scheduling algorithm given by Equation 4.4. Then the average throughput of the greedy algorithm $\tilde{\theta}_{i,n}$ converges to $\tilde{\theta}_i^* = \frac{\mathbb{E}[M_N]}{N}$ almost surely for all $i \in \{1, \ldots, N\}$ and M_N does not depend on time.*

Proof. Let $\tilde{I}_{i,n}$ be the indicator random variable for the event that user i was scheduled at time slot n under the greedy algorithm. By Equation 4.4, $\tilde{I}_{i,n}(X^n) = 1$ if and only if $\max\{X_{1,n}, \ldots, X_{N,n}\} = X_{i,n}$, and $\tilde{I}_{i,n}(X^n) = 0$ otherwise. Because the algorithm is greedy, $\tilde{I}_{i,n}(X^n)$ is only a function of the current rates X^n, and for simplicity the the explicit functional dependence will be dropped from the notation. Due to the i.i.d. $X_{i,n}$s, from [Dav70] it is known that $\Pr(X_{i,n} = 1) = \frac{1}{N}$ for all i and all n. The average throughput of user i at time n can be written

$$\tilde{\theta}_{i,n} = \frac{1}{n} \sum_{j=1}^{n} \tilde{I}_{i,j} X_{i,j}. \tag{4.12}$$

The expectation of $\tilde{I}_{i,j} X_{i,j}$ is given by

$$\mathbb{E}[\tilde{I}_{i,j} X_{i,j}] = \mathbb{E}[X_{i,j} | \tilde{I}_{i,j} = 1] \cdot \Pr(\tilde{I}_{i,j} = 1) = \frac{\mathbb{E}[M_{N,n}]}{N}. \tag{4.13}$$

By the underlying i.i.d. assumption, $\mathbb{E}[M_{N,n}]$ does not depend on the time index n, so it will be dropped. Let $Z_{i,n} = \tilde{I}_{i,j} X_{i,j}$, then Equation 4.12 can be rewritten as $\tilde{\theta}_{i,n} = \frac{1}{n}\sum_{j=1}^{n} Z_{i,j}$, and the $Z_{i,j}$ are i.i.d. for each i. By the boundedness and non-negativity of the $X_{i,n}$

$$\mathbb{E}[|Z_{i,n}|] = \mathbb{E}[Z_{i,n}] = \frac{\mathbb{E}[M_N]}{N} < \infty,$$

so the Strong Law of Large Numbers (SLLN) can be applied and yields

$$\tilde{\theta}_{i,n} \to \tilde{\theta}_i^* = \frac{\mathbb{E}[M_N]}{N} \text{ a.s. as } n \to \infty \text{ for all } i \in \{1, \ldots, N\}. \tag{4.14}$$

□

Having established the previous results, the convergence of the PFS algorithm to the greedy algorithm is proven in the following theorem:

Theorem 11. *Let $X_{1,n}, \ldots, X_{N,n}$ be i.i.d. non-negative bounded random variables for all $n \in \mathbb{N}^+$. The PFS algorithm given by Equation 4.3 converges to the same average throughput point $\tilde{\theta}^*$ as the greedy scheduling algorithm given by Equation 4.4. Therefore the asymptotic performance of the two algorithms under i.i.d. rates is identical.*

Proof. Due to the Kushner-Clark Theorem and Theorem 2.2 from [KW04], the PFS algorithm converges to a unique point θ^*. This solution θ^* is the unique zero of the mean function h. Lemma 3 states that the greedy algorithm under i.i.d. rates converges to the average throughput point $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N} \cdot \mathbf{1}_N \in \mathbb{R}^N$, where $\mathbf{1}_N$ is the N-dimensional vector of all ones. The last step to show convergence of the PFS algorithm to the greedy algorithm is to show that the convergence point of the greedy algorithm $\tilde{\theta}^*$ is a zero of the mean function h:

$$\begin{aligned} h\left(\tilde{\theta}^*\right) &= \int H\left(\tilde{\theta}^*, x\right)\mu(dx) &(4.15)\\ &= \int \left[I\left(\tilde{\theta}^*, x\right)x - \tilde{\theta}^*\right]\mu(dx) &(4.16)\\ &= \mathbb{E}_X\left[I\left(\tilde{\theta}^*, X\right)X\right] - \tilde{\theta}^* &(4.17)\\ &= \frac{\mathbb{E}[M_N]}{N} \cdot \mathbf{1}_N - \tilde{\theta}^* &(4.18)\\ &= 0. &(4.19) \end{aligned}$$

Equation 4.15 is the definition of the mean function given in Equation 4.7 with the point $\tilde{\theta}^*$ as the argument and the distribution μ is for the random variable $X \in \mathbb{R}^N$. Equation 4.16 plugs in the definition of $H(\theta, X)$ and since $\tilde{\theta}^*$ is non-random, Equation 4.17 follows. The key step of the proof is going from Equation 4.17 to Equation 4.18. Notice that when the θ vector has equal components, the PFS function given in Equation 4.3 is equivalent to the greedy function in Equation 4.4, i.e.

$$\arg \max_{1 \le i \le N} \left\{ \frac{X_{i,n+1}}{d_i + \theta_{i,n}} \right\} = \arg \max_{1 \le i \le N} \{X_{i,n+1}\} \tag{4.20}$$

when $\theta = c \cdot \mathbf{1}_N$ for some positive constant c. For $\tilde{\theta}^*$, $c = \frac{\mathbb{E}[M_N]}{N}$, so

$$\mathbb{E}_X \left[I\left(\tilde{\theta}^*, X\right) X \right] = \frac{\mathbb{E}[M_N]}{N} = \tilde{\theta}^*,$$

the convergence point of the greedy algorithm. Therefore, $\tilde{\theta}^*$ is a zero of the mean function h, completing the proof. □

An interpretation of the mean function h is as a potential field induced by the PFS algorithm. If one component of the current average throughput vector θ^n is large, then the PFS algorithm will select the large component at the next time slot with decreased probability due to the larger denominator in Equation 4.3. Thus, viewed as a potential field, the potential is trying to push the average throughput vector θ^n back to an equilibrium point, the zeros of h. The monotonicity properties of the PFS algorithm guarantee that there is one unique equilibrium point, which corresponds to the average throughput vector that the greedy algorithm converges to. Figure 4.1 shows the sample paths of the PFS algorithm and the greedy algorithm on the same set of realized i.i.d. uniform random variables. For the two-dimensional uniform case, the mean of the maximum order statistic is 2/3, so the equilibrium point is $[1/3, 1/3]^T$. As is expected by the preceding theory, both the PFS and greedy algorithms converge to the equilibrium point.

4.5 Convergence with Unbounded Rates

In the previous section, the rate random variables $X_{i,n}$ were assumed to be bounded, but in some applications these random variables may be unbounded

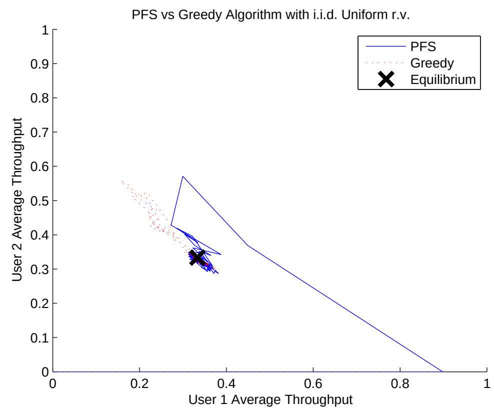

Figure 4.1: Sample Paths of the PFS and Greedy Algorithms for 2 i.i.d. Uniform Rates

from above. For example, in [SH05], the scheduling decision is based on the SINR random variables which are supported on the non-negative real line. Unbounded random variables pose problems for the SA algorithm because at any iteration, the algorithm may move an arbitrarily large distance and prevent convergence. In this section, conditions on the unbounded $X_{i,n}$ are found such the SA algorithm converges to the greedy algorithm.

The possible unboundedness of the state variable θ is often too troublesome to deal with, so in the SA literature, the state variable is constrained to be in some bounded set of the state space. For example, in [KY03], a constraint set $H = \{\theta : a_i \leq \theta_i \leq b_i\}$ is defined, where $[a_i, b_i]$ are the constraints on each individual component of θ. At each iteration, the state space is projected back onto the constraint set, so that θ can never leave this set, i.e.

$$\theta^{n+1} = \Pi_H \left[\theta_n + \epsilon_{n+1} Y^n\right] \tag{4.21}$$

where Π_H is the projection operator onto the constraint set. As in [KY03], Equation 4.21 can be written as

$$\theta^{n+1} = \theta^n + \epsilon_{n+1} Y^n + \epsilon_{n+1} Z^n \tag{4.22}$$

where $Z^n \in \mathbb{R}^N$ is the projection term and $\epsilon_{n+1} Z^n = \theta^{n+1} - \theta^n - \epsilon_{n+1} Y^n$ is the shortest vector in Euclidean distance required to get $\theta^n + \epsilon_{n+1} Y^n$ back into the constraint set H. Another approach to constraining the state variable is taken in [Che02], and the approach is as follows:

$$\theta^{n+1} = I_{[\|\theta^n+\epsilon_{n+1}Y^n\|<b]}\left(\theta^n + \epsilon_{n+1}Y^n\right) + I_{[\|\theta^n+\epsilon_{n+1}Y^n\|\geq b]}\theta^\dagger \tag{4.23}$$

where $\theta^\dagger \in \mathbb{R}^N$ is a fixed point. Here $I_{[\cdot]} \in \mathbb{R}^{N\times N}$ is an indicator matrix of zeros whose diagonal elements are one when the argument in the subscript is true, and zero otherwise. Equation 4.23 states that when the norm of the next iteration $\theta^n + \epsilon_{n+1}Y^n$ is less than some positive constant b, update the state θ^{n+1} as normal, otherwise the state vector gets updated as $\theta^{n+1} = \theta^\dagger$. Both the methods of Equation 4.21 and Equation 4.23 prevent the state variable from leaving some bounded set in $\mathbb{R}^N$.

As discussed in [Che02], the weak convergence method of stochastic approximation analysis can handle the unboundedness of the state variable sequence θ^n in some cases, but this method guarantees convergence in distribution only. The convergence in Section 4.4 is almost sure convergence, and if possible, it should be shown that under certain conditions, the PFS algorithm with unbounded rate random variables also converges almost surely. A two step approach is taken to find the conditions on the $X_{i,n}$ such that the PFS algorithm converges to the greedy algorithm. The first step is to find necessary and sufficient conditions for the convergence of the greedy algorithm. If the greedy algorithm fails to converge, then there is no need to prove the convergence of the PFS algorithm. The second step is to show that under the necessary and sufficient conditions on $X_{i,n}$ for the convergence of the greedy algorithm, the PFS algorithm converges to the greedy algorithm. The method for proving the second step is motivated by the method of stochastic approximation algorithms with expanding truncations found in Chapter 2 of [Che02].

The question addressed by the first step is when does the greedy algorithm converge? Necessary and sufficient conditions on the $X_{i,n}$ random variables are found that guarantee the desired convergence. First, the following theorem from [Dur05] is required:

Theorem 12. *(Thm 7.2, Chapter 1, [Dur05])*
Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i^+ = \infty$ and $\mathbb{E}X_i^- < \infty$ where X_i^+ and X_i^- are the

positive and negative parts of the random variables respectively. If $S_n = X_1 + \cdots + X_n$ *then* $S_n/n \to \infty$ *almost surely.*

Additionally, necessary and sufficient conditions for the existence of the expectation of the maximum order statistic of interest must be found. This conditions are provided in the next lemma:

Lemma 4. *Let* $X_1, \ldots, X_N$ *be i.i.d. non-negative random variables and* $M_N = \max\{X_1, \ldots, X_N\}$*. Then* $\mathbb{E}[M_N] < \infty$ *if and only if* $\mathbb{E}[X_i] < \infty$*.*

Proof. $\Leftarrow$ From Chapter 3 of [Dav70], it is a well known result in order statistics that if $\mathbb{E}[X_i]$ exists, then the mean of all the order statistics exist, including the maximum order statistic. Thus if $\mathbb{E}[X_i] < \infty$, then $\mathbb{E}[M_N] < \infty$.
$\Rightarrow$ To prove that $\mathbb{E}[M_N] < \infty$ implies that $\mathbb{E}[X_i] < \infty$ for all $i \in \{1, \ldots, N\}$ requires the following theorem:

Theorem 13. *(Thm 3.1c, Chapter 1, [Dur05])*
Suppose $X, Y \geq 0$ *or* $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$*. If* $X \geq Y$ *then* $\mathbb{E}X \geq \mathbb{E}Y$*.*

All the $X_i \geq 0$, and hence $M_N \geq 0$ as well. By definition $X_i \leq M_N$, and thus by Theorem 13, $0 \leq \mathbb{E}[X_i] \leq \mathbb{E}[M_N] < \infty$, providing the desired result. □

The result of Lemma 4 may appear obvious, but in general the existence of the mean of an order statistic does not imply the existence of the mean of the underlying random variables. The classical example, as mentioned in Chapter 3 of [Dav70], is when $X_1, \ldots, X_N$ are i.i.d. Cauchy random variables. Let $\mu_{r:N}$ be the mean of the r^{th} order statistic. Then in this case, $\mu_{r:N}$ exists for $2 \leq r \leq N-1$, but the mean of any X_i does not exist.

Having defined Theorem 12 and Lemma 4, the necessary and sufficient conditions for the convergence of the greedy algorithm under the i.i.d. model are now proven.

Theorem 14. *Let* $X_{1,n}, \ldots, X_{N,n}$ *be i.i.d. non-negative random variables for all* $n \in \mathbb{N}^+$*,* $M_{N,n} = \max\{X_{1,n}, \ldots, X_{N,n}\}$ *and let* $\tilde{\theta}_{i,n}$ *be the average throughput of user* i *at time* n *under the greedy scheduling algorithm given by Equation 4.4. A necessary and sufficient condition for the convergence of the greedy algorithm* *is that* $\mathbb{E}[X_{i,n}] < \infty$*, in which case for each user* $i \in \{1, \ldots, N\}$ *the algorithm converges to* $\tilde{\theta}_i^* = \frac{\mathbb{E}[M_N]}{N} < \infty$*.*

Proof. By Lemma 4, a necessary and sufficient condition for the existence of $\mathbb{E}[M_{N,n}]$ is that $\mathbb{E}[X_{i,n}] < \infty$. Thus $\mathbb{E}[X_{i,n}] < \infty$ is a sufficient condition for the convergence of the greedy algorithm because the hypotheses of the SLLN hold and using the same arguments as Lemma 3. Because $\mathbb{E}[X_{i,n}^-] = \mathbb{E}[M_{N,n}^-] = 0$, if $\mathbb{E}[X_{i,n}] = \infty$, then by Lemma 4 $\mathbb{E}[M_{N,n}] = \infty$ and by Theorem 12 the greedy algorithm would diverge to infinity. Therefore the condition $\mathbb{E}[X_{i,n}] < \infty$ is a necessary condition. When the greedy algorithm converges, by the SLLN it converges to $\tilde{\theta}_i^* = \frac{\mathbb{E}[M_N]}{N}$ for each user i, so the convergence point is $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N}\mathbf{1}_N$. □

Theorem 14 provides the sought after necessary and sufficient conditions for convergence of the greedy algorithm. The second step is to show that these same conditions guarantee the convergence of the PFS algorithm to the same point $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N}\mathbf{1}_N$. A method of analysis for the convergence of SA algorithms with possibly unbounded increments is the method of expanding truncations. The basic idea is that the bounded set containing the SA algorithm is expanded at each iteration, and if it can be shown that the SA algorithm leaves these bounded sets a finite number of times, then the performance of the expanding truncation method is equivalent to the SA algorithm as if it were unconstrained. The theorems provided in [Che02] that show convergence of the expanding truncations methods typically depend on properties of the mean function h. In this case, the mean function h is not explicitly known in closed form. Because of the similarity of the PFS algorithm to the SLLN, the proof of the SLLN has provided a method to prove the convergence of the PFS algorithm with unbounded rates under the same conditions as Theorem 14 to the greedy algorithm. The following theorem proves this convergence.

Theorem 15. *Let* $X_{1,n}, \ldots, X_{N,n}$ *be i.i.d. non-negative random variables such that* $\mathbb{E}[X_{i,n}] < \infty$ *for all* $n \in \mathbb{N}^+$*. The PFS algorithm given by Equation 4.3 converges to the same average throughput point* $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N}$ *as the greedy scheduling algorithm given by Equation 4.4 and Theorem 14.*

Proof. Redefine the recursive stochastic approximation algorithm given in Equation 4.6 as

$$\Theta^{n+1} = \Theta^n + \epsilon_{n+1}\left[I^{n+1}(\Theta^n, X^{n+1})X^{n+1}\mathbf{1}_{[\|I^{n+1}(\Theta^n,X^{n+1})X^{n+1}\|\leq n+1]} - \Theta^n\right]$$

where Θ refers to the average throughput vector under the new definition and $\mathbf{1}_{[\|X^{n+1}I^{n+1}(\Theta^n,X^{n+1})\|\leq n+1]}$ is the indicator matrix random variable whose diagonal elements are one when $\|I^{n+1}(\Theta^n, X^{n+1})X^{n+1}\| \leq n+1$ and zero otherwise. Because only one component of $I^{n+1}(\Theta^n, X^{n+1})X^{n+1}$ is non-zero, $\|I^{n+1}(\Theta^n, X^{n+1})X^{n+1}\| = X_{j,n+1}$ if user j is scheduled for time slot $n+1$. By this observation and the Triangle Inequality

$$\begin{aligned}\|\Theta^{n+1}\| &\leq \frac{1}{n+1}\sum_{i=0}^{n}\|I^{i+1}(\Theta^i, X^{i+1})X^{i+1}\mathbf{1}_{[\|I^{i+1}(\Theta^i,X^{i+1})X^{i+1}\|\leq i+1]}\| \\ &\leq \frac{1}{n+1}\sum_{i=1}^{n+1} i = \frac{1}{n+1}\cdot\frac{(n+1)(n+2)}{2} = \frac{n+2}{2}. \end{aligned} \tag{4.24}$$

Therefore, at each iteration, the norm of the state random variable belongs to a bounded set, although the constraint is not the one used in Equation 4.21 or Equation 4.23. If it can be shown that the number of terms that are truncated is finite, then asymptotically the expanding truncations and non-truncated algorithms are equivalent. From Section 4.4, it is known that the in the bounded case, the PFS algorithm converges to $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N}$.

The method used to show the number of truncations is finite is motived by Etemadi's proof of the SLLN ([Ete81]). Because $\|X^{n+1}I^{n+1}(\Theta^n, X^{n+1})\| \leq M_{N,n} = \max\{X_{1,n+1}, \ldots, X_{N,n+1}\}$ and $M_{N,n}$ is a non-negative random variable, the following inequality holds

$$\begin{aligned}\sum_{i=1}^{\infty}\Pr\left(\|X^iI^i(\Theta^{i-1}, X^i)\| > i\right) &\leq \sum_{i=1}^{\infty}\Pr\left(M_{N,i} > i\right) \\ &\leq \int_0^{\infty}\Pr\left(M_N > t\right)dt \\ &= \mathbb{E}M_N < \infty \end{aligned} \tag{4.25}$$

where $\mathbb{E}M_N < \infty$ since $\mathbb{E}[X_{i,n}] < \infty$ by assumption and Lemma 4. Let A_n be the event that $\|X^nI^n(\Theta^{n-1}, X^n)\| > n$. Then by the previous set of inequalities, the

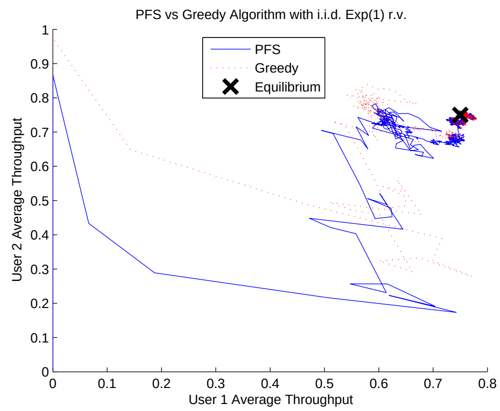

Figure 4.2: Sample Paths of the PFS and Greedy Algorithms for 2 i.i.d. Exp(1) Rates

First Borel-Cantelli Lemma says that $\Pr(A_n \text{ infinitely often}) = 0$. Thus, there are only a finite number of truncations. Therefore, if $\mathbb{E}[X_{i,n}] < \infty$, the PFS algorithm converges to $\tilde{\theta}^* = \frac{\mathbb{E}[M_N]}{N}$, which under the same conditions is the convergence point of the greedy algorithm as proven in Theorem 14. □

This section is concluded with an example. Figure 4.2 shows the sample paths of the PFS and greedy algorithm on the same set of realized i.i.d. exponential-1 random variables. The exponential random variables are unbounded from above, but the convergence still occurs as expected. In the case of two exponential random variables, the mean of the maximum order statistic is 1.5, so the equilibrium point is given by $[3/4, 3/4]^T$.

4.6 Rate of Convergence

4.6.1 Rate of Convergence to the Equilibrium Point

In Sections 4.4 and 4.5, it is shown that the PFS algorithm under the i.i.d. model converges to a single average rate vector θ^*, and that θ^* corresponds to the performance of the greedy scheduler. The goal of this section is to find out how fast $\|\theta^n - \theta^*\| \to 0$. This section will leverage the results in Chapter 3 of [Che02], which are based on four assumptions, which are now stated:

1. $\epsilon_k > 0$, $\epsilon_k \to 0$ as $k \to \infty$, $\sum_{k=1}^{\infty} \epsilon_k = \infty$, and $\frac{\epsilon_k - \epsilon_{k+1}}{\epsilon_k \cdot \epsilon_{k+1}} \to \alpha \geq 0$ as $k \to \infty$.

2. A continuously differentiable function $v(\cdot) : \mathbb{R}^l \to \mathbb{R}$ exists such that

$$\sup_{\delta \leq \|\theta - \theta^*\| \leq \Delta} f^T(\theta) v(\theta) < 0$$

   for any $\Delta > \delta > 0$, and $v(\theta^*) < \inf_{\|\theta\| = c_0}$ for some $c_0 > 0$ with $c_0 > \|\theta^\dagger\|$, where $\theta^\dagger$ is used in Equation 4.23.

3. The noise sequence $\{\delta M^n\}$ can be decomposed into two parts $\delta M^n = \delta M_1^n + \delta M_2^n$ such that

$$\sum_{k=1}^{\infty} \epsilon_k^{1-\delta} \delta M_1^n < \infty \text{ and } \delta M_2^n = O(\epsilon_k^{\delta})$$

   for some $\delta \in (0, 1]$.
   (Note: δM^n is a single quantity, not to be confused with δ.)

4. $h(\cdot)$ is a measurable and locally bounded, and is differentiable at θ^* such that as $\theta \to \theta^*$

$$h(\theta) = F(\theta - \theta^*) + \delta(\theta)$$
$$\delta(\theta^*) = 0$$
$$\delta(\theta) = o\left(\|\theta - \theta^*\|\right).$$

   The matrix F is stable.

Assuming that these four conditions are satisfied, the following theorem is true:

Theorem 16. *(Theorem 3.1.1, [Che02]) Under the four assumptions previously stated, θ^n as given in Equation 4.23 converges to θ^* with the following convergence rate:*

$$\|\theta^n - \theta^*\| = o\left(\epsilon_n^{\delta}\right), \tag{4.26}$$

where δ is the one given in Assumption 3.

To verify the assumptions required of Theorem 16 looks formidable, but for the most part this is not the case. For example, it is shown in [Che02] that when the gain sequence is of the form $\epsilon_k = \frac{a}{k}$ for some constant $a > 0$, then $\frac{\epsilon_k - \epsilon_{k+1}}{\epsilon_k \cdot \epsilon_{k+1}} \to \frac{1}{a}$. Thus, the gain sequence of the PFS algorithm clearly satisfies the first assumption with $a = 1$. The second assumption is only needed to guarantee convergence of the SA algorithm and is not involved in the derivation of the rate of convergence in Theorem 16. The convergence of the PFS algorithm was proven using a different approach in the previous two sections, so Assumption 2 is not required to find the rate of convergence. In the fourth assumption, if h is measurable, locally bounded, and differentiable at θ^*, then the matrix F is stable since θ^* is a the unique zero and global equilibrium. Therefore, to be able to utilize Theorem 16 and compute the rate of convergence, Assumption 3 must be verified, and it must be shown that h is measurable, locally bounded, and differentiable at θ^*. Showing these two results yields the following theorem:

Corollary 5. *Let $X_{i,n}, \ldots, X_{N,n}$ be continuous, non-negative i.i.d. random variables for $n \in \mathbb{N}^+$ with continuous densities f such that $\mathbb{E}[X_{i,n}^2] < \infty$. The convergence rate of the PFS algorithm is given by*

$$\|\theta^n - \theta^*\| = o\left(\frac{1}{n^\delta}\right), \tag{4.27}$$

where $\delta = \frac{1}{2} - \epsilon$ and $\epsilon > 0$.

Proof. As mentioned prior to the statement of the corollary, this results immediately follows from Theorem 16, and is just a matter of verifying all of the assumptions, and all the assumptions are apparent except for Assumption 3 and the differentiability of h at θ^*. To proof that Assumption 3 holds and to compute the value of δ is based on Remark 3.1.2 in [Che02]. In the remark, it is shown that if δM^n is a martingale difference sequence, one can set $\delta M_1^n = \delta M^n$ and $\delta M_2^n = 0$ for all n. It is also shown that in order to have $\sum_{k=1}^{\infty} \left(\frac{1}{k}\right)^{1-\delta} \delta M^k < \infty$ a.s. , it suffices to have

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\frac{1}{k^{2(1-\delta)}} \|\delta M^k\|^2 | \mathcal{F}_k\right] < \infty \tag{4.28}$$

if δM^n is a MDS such that $\sup_k \mathbb{E}\left[\|\delta M^n\|^2 | \mathcal{F}_n\right] < \infty$. Under this condition, Equation 4.28 holds provided $\delta < \frac{1}{2}$. The definition of the MDS δM^n is given in

Equations 4.9 and 4.10 and the condition that $\sup_k \mathbb{E}\left[\|\delta M^n\|^2 | \mathcal{F}_n\right] < \infty$ holds because of the hypothesis on the second moment, $\mathbb{E}[X_{i,n}^2] < \infty$, and the fact all the random variables are i.i.d. Therefore Assumption 3 holds with $\delta < \frac{1}{2}$.

The final part of the proof is to show that h is measurable, locally bounded, and differentiable at θ^*. The function h is clearly measurable from its definition in Equation 4.8. To show that h is locally bounded, consider the first component of h. Let $\mathbf{1}(\theta^n, X_1)$ be the indicator function that the user corresponding to the first component of h is scheduled, and let f denote the density of $X_{i,n}$. Then the following expressions show that the first component of h, $h(\theta^n)_1$ is bounded.

$$\begin{aligned}
h(\theta^n)_1 &= \int x_1 \mathbf{1}(\theta^n, X_1) f(x_1, \ldots, x_n) dx_1 \cdots dx_N - \theta_i^n \\
&= \int x_1 \mathbf{1}(\theta^n, X_1) f(x_1) \cdots f(x_N) dx_x \cdots dx_N - \theta_i^n \\
&\leq \int x_1 f(x_1) \cdots f(x_N) dx_1 \cdots dx_N - \theta_i^n \\
&= \int x_1 f(x_1) dx_1 - \theta_i^n \\
&= \mathbb{E}[X_1] - \theta_i^n \\
&< \infty.
\end{aligned}$$

The last inequality is because $\mathbb{E}[X_{i,n}^2] < \infty$. Since all the random variables are i.i.d., the above analysis holds for all components of h, and thus h is locally bounded. To show that h is differentiable at θ^* leverages an observation on the smoothness of h in [KW04]. In their observation, it is shown that each component of h has continuous derivative if f is bounded and continuous. Here, differentiability is only required at θ^*, and f is bounded in the neighborhood of θ^*, otherwise f would not be continuous. Therefore, since each component of h is differentiable at θ^*, h is differentiable at θ^*, and thus Assumption 4 holds, completing the proof. □

To see if this makes any sense, a simulation was run with $N = 3$ and $X_{i,n} \sim \text{Exp}(1)$. Figure 4.3 shows the results of the simulation. The convergence towards equilibrium of the PFS algorithm will not be monotonic due to the randomness of the rates and sample paths of the algorithm. The results of Theorem 16 and Corollary 5 give the asymptotic scaling rate in terms of o notation. The red curve

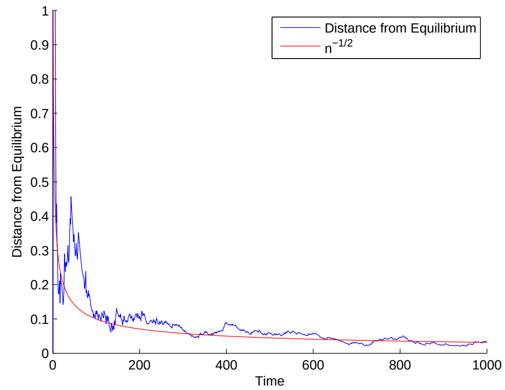

Figure 4.3: Distance from Equilibrium versus Time for $N = 3$, $X_{i,n} \sim \text{Exp}(1)$

in Figure 4.3 shows the function $f(n) = n^{-\frac{1}{2}}$. The long term convergence rate of the PFS algorithm to equilibrium is slightly slower than this, but the curve shows that the results of the theory are reasonable.

4.6.2 Asymptotic Covariance Matrix

In Section 4.6.1 the rate at which the proportional fair sharing algorithm converged to the equilibrium point, or the greedy solution, was found as measured by the Euclidean distance. Because the PFS algorithm is a random algorithm, the vector $\theta^n - \theta^*$ is a random variable. The goal of this section is to that an appropriate normalized sequence $c_n(\theta^n - \theta^*)$ converges to a normal random variable, where c_n is the normalizing sequence, and to to characterize the mean and covariance matrix of this normal random variable.

Work has been done to identify this normal random variable. Two of the more cited references in this area are [Sac58] and [Fab68]. A theorem in [Sac58], adapted to the PFS framework, will be used to find the normal random variable of interest as it applies to the PFS algorithm. The key theorem is now stated.

Theorem 17. *(Theorem 5, [Sac58])*
Let $\mathbb{E}[X_{i,n}^{2+\nu}] < \infty$ *for some* $\nu > 0$. *Let* F *be the Jacobian matrix of* $h(\cdot)$, *i.e.*

$h(\theta) = F(\theta - \theta^*) + \delta(\theta)$ with $\delta(\theta^*) = 0$ and $\delta(\theta) = o(\|\theta - \theta^*\|)$, and let

$$\lim_{\theta \to \theta^*} \mathbb{E}\left[\delta M(\theta) \delta M(\theta)^T\right] = \pi.$$

Finally, let P be the matrix of eigenvectors of $-F$ and Λ be a diagonal matrix whose diagonal values $\lambda_i = \Lambda_{ii}$ are the eigenvalues of $-F$ and $\pi^* = P^{-1}\pi P$. Then $n^{1/2}(\theta^n - \theta^*)$ is asymptotically normal with mean 0 and covariance matrix PQP^{-1} where Q is the matrix whose $(i,j)^{th}$ element is $\frac{\pi^*_{i,j}}{\lambda_i + \lambda_j - 1}$.

To successfully utilize the Theorem 17 requires the explicit computation of F and π, which depends on the underlying distribution of the $X_{i,n}$. It is straight forward to get a general form for π, which is stated in the following corollary.

Corollary 6. *Let $X_{i,n}$ have continuous density f and $\mathbb{E}[X_{i,n}^{2+\nu}] < \infty$ for some $\nu > 0$. Then π is a symmetric matrix with elements $\pi_{ij} = \frac{1}{N}\mathbb{E}[X^2_{max}] - \left(\frac{\mathbb{E}[X_{max}]}{N}\right)^2$ when $i = j$ and $\pi_{ij} = -\left(\frac{\mathbb{E}[X_{max}]}{N}\right)^2$ when $i \neq j$, where N is the number of users.*

Proof. Recall, that parameterized by θ, $\delta M(\theta) = I(\theta, X)X - \theta - h(\theta)$. First, consider the first diagonal element:

$$\pi_{11} = \lim_{\theta \to \theta^*} \mathbb{E}\left[\left(I_{(1)}(\theta, X)X_{(1)} - \theta_{(1)} - h_{(1)}(\theta)\right)^2\right]$$

Notice that as $\theta \to \theta^*$, any product involving $h(\theta)$ goes to zero since h is smooth and is zero at θ^*. Thus, the above equation becomes

$$\pi_{11} = \lim_{\theta \to \theta^*} \mathbb{E}\left[\left(I_{(1)}(\theta, X)X_{(1)} - \theta_{(1)}\right)^2\right] \tag{4.29}$$

$$= \lim_{\theta \to \theta^*} \mathbb{E}\left[I^2_{(1)}(\theta, X)X^2_{(1)} - 2I_{(1)}(\theta, X)X_{(1)} \cdot \theta_{(1)} + \theta^2_{(1)}\right] \tag{4.30}$$

$$= \mathbb{E}\left[X^2_{max}\right] \cdot \Pr(I_{(1)}(\theta^*, X) = 1) - 2\theta^*_{(1)}\mathbb{E}[X_{max}] \cdot \Pr(I_{(1)}(\theta^*, X) + (\theta^*_{(1)})^2 \tag{4.31}$$

$$= \frac{1}{N}\mathbb{E}\left[X^2_{max}\right] - 2\theta^*_{(1)}\frac{1}{N}\mathbb{E}[X_{max}] + (\theta^*_{(1)})^2 \tag{4.32}$$

$$= \frac{1}{N}\mathbb{E}\left[X^2_{max}\right] - \left(\frac{\mathbb{E}[X_{max}]}{N}\right)^2 \tag{4.33}$$

where Equation (4.31) is due to the continuity of the density and Equation (4.33) is due to the definition of $\theta^*_{(1)} = \frac{\mathbb{E}[X_{max}]}{N}$. By the i.i.d. assumption on the $X_{i,n}$,

$\pi_{ij} = \pi_{11}$ when $i = j$. Now consider the off-diagonal element $\pi_{1,2}$, given as follows:

$$\pi_{12} = \lim_{\theta \to \theta^*} \mathbb{E}\left[\left(I_{(1)}(\theta, X) X_{(1)} - \theta_{(1)} - h_{(1)}(\theta) \right) \cdot \left(I_{(2)}(\theta, X) X_{(2)} - \theta_{(2)} - h_{(2)}(\theta) \right) \right].$$

Evaluating this expression in a similar fashion as π_{11} yields

$$\pi_{12} = -\left(\frac{\mathbb{E}[X_{max}]}{N} \right)^2$$

since the terms involving the product $I_{(1)}(\theta, X) \cdot I_{(2)}(\theta, X)$ always equal zero due to the fact only one component in the vector $I(\theta, X)$ can be non-zero. Once again, by the i.i.d. assumption $\pi_{ij} = \pi_{12}$ for $i \neq j$. Note that π can be written in the special form $\pi = \frac{\mathbb{E}[X_{max}]}{N} I_N - \left(\frac{\mathbb{E}[X_{max}]}{N} \right)^2 \mathbf{1}_N \mathbf{1}_N^T$ where I_N is the $N \times N$ identity matrix and $\mathbf{1}_N$ is an N-dimension vector of 1s. This completes the corollary. □

Although a general form is easily obtained for the matrix π to be used in Theorem 17, the matrix F is more complicated. Rather than state a corollary, two examples will be given in the simplest of cases to show how to evaluate F, how the evaluation scales with N, and then the application of F and π to find the asymptotic covariance matrix of $n^{1/2}(\theta^n - \theta^*)$ as promised by Theorem 17.

Example 1.

Consider the case where $N = 2$ and the $X_{i,n}$ are distributed i.i.d. according to $U[0, 1]$. In this case, the distribution function of X_{max} is $F_{max}(x) = x^2$ for $x \in [0, 1]$ and the density is given by $f_{max}(x) = \frac{d}{dx} F_{max}(x) = 2x$ on the same support. Using this, it is straight forward to compute $\mathbb{E}[X_{max}^2] = \frac{1}{2}$ and $\mathbb{E}[X_{max}] = \frac{2}{3}$. Plugging these values into Corollary 6 to get the MDS covariance matrix yields

$$\pi = \begin{bmatrix} \frac{5}{36} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{36} \end{bmatrix}.$$

Next, the Jacobian matrix must be computed. Due to the i.i.d. symmetry of the problem, without loss of generality, consider the first component of $h(\theta)$:

$$\begin{aligned} h_{(1)}(\theta) &= \mathbb{E}_X \left[I_{(1)}(\theta, X) X_{(1)} - \theta_{(1)} \right] = \mathbb{E}_X [I_{(1)}(\theta, X) X_{(1)}] - \theta_{(1)} \\ &= \int_0^1 \int_{\frac{\theta_{(1)}}{\theta_{(2)}} x_2}^1 x_1 dx_1 dx_2 - \theta_{(1)} \end{aligned} \tag{4.34}$$

where the indicator function determines the region of integration which is reflected in the integration limits, i.e. $X_{(1)}$ is selected when $\frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(2)}}{\theta_{(2)}}$ which implies $X_{(1)} \geq \frac{\theta_{(1)}}{\theta_{(2)}} X_{(2)}$. Evaluating $\frac{\partial}{\partial \theta_{(1)}}$ and $\frac{\partial}{\partial \theta_{(2)}}$ of Equation (4.34) gives the first row of the Jacobian matrix F, and by symmetry gives the full matrix F

$$F = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

Using F, π and Theorem 17 gives an asymptotic covariance matrix Σ of

$$\Sigma = \begin{bmatrix} 0.0389 & -0.0111 \\ -0.0111 & 0.0389 \end{bmatrix}$$

A simulation was performed that simulated 10^4 iterations of the PFS algorithm 10^4 times and the empirical covariance matrix was found to be

$$\Sigma_{sim} = \begin{bmatrix} 0.0394 & -0.0117 \\ -0.0117 & 0.0394 \end{bmatrix}$$

which is very close to the theoretical value.

Example 2.

Consider the case where $N = 3$ and the $X_{i,n}$ are distributed i.i.d. according to $U[0, 1]$. This case is identical to the previous case except that there is an additional user. Using the same computations as before, but with $F_{max}(x) = x^3$, gives an MDS covariance matrix of

$$\pi = \begin{bmatrix} \frac{11}{80} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{11}{80} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{11}{80} \end{bmatrix}.$$

The key difference between the $N = 2$ and $N = 3$ case is the evaluation of $\mathbb{E}_X[I_{(1)}(\theta, X) X_{(1)}]$. In the $N = 2$ case, one must only consider when $\frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(2)}}{\theta_{(2)}}$, but in the $N = 3$ case one must consider $\frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(2)}}{\theta_{(2)}}$ and $\frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(3)}}{\theta_{(3)}}$. With this observation, $I_{(1)} X_{(1)}$ can be written as

$$I_{(1)} X_{(1)} = \begin{cases} X_{(1)} & \text{if } \frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(2)}}{\theta_{(2)}} \geq \frac{X_{(3)}}{\theta_{(3)}} \text{ or } \frac{X_{(1)}}{\theta_{(1)}} \geq \frac{X_{(3)}}{\theta_{(3)}} \geq \frac{X_{(2)}}{\theta_{(2)}} \\ 0 & \text{otherwise} \end{cases} \tag{4.35}$$

The conditions for $I_{(1)}X_{(1)} = X_{(1)}$ represent the fact that the ratio $\frac{X_{(1)}}{\theta_{(1)}}$ needs to be larger than $\max\left\{\frac{X_{(2)}}{\theta_{(2)}}, \frac{X_{(3)}}{\theta_{(3)}}\right\}$. Each of the conditions produces a different integration region for the evaluation of $h_{(1)}(\theta)$. Thus, for $N = 3$, the analog of Equation (4.34) is

$$\begin{aligned} h_{(1)}(\theta) &= \mathbb{E}_X[I_{(1)}(\theta, X)X_{(1)}] - \theta_{(1)} \\ &= \int_0^1 \int_{\frac{\theta_{(2)}}{\theta_{(3)}}x_3}^1 \int_{\frac{\theta_{(1)}}{\theta_{(2)}}x_2}^1 x_1 dx_1 dx_2 dx_3 + \int_0^1 \int_{\frac{\theta_{(3)}}{\theta_{(2)}}x_2}^1 \int_{\frac{\theta_{(1)}}{\theta_{(3)}}x_3}^1 x_1 dx_1 dx_3 dx_2 - \theta_{(1)}. \end{aligned} \tag{4.36}$$

As in the $N = 2$ case, evaluating the partial derivatives of Equation (4.36) yields

$$F = \begin{bmatrix} -3 & 1 & 1 \\ 1 & -3 & 1 \\ 1 & 1 & -3 \end{bmatrix}.$$

Although the computation of the Jacobian looks cumbersome, due to the symmetry of the problem $\int_0^1 \int_{\frac{\theta_{(2)}}{\theta_{(3)}}x_3}^1 \int_{\frac{\theta_{(1)}}{\theta_{(2)}}x_2}^1 x_1 dx_1 dx_2 dx_3$ is identical to $\int_0^1 \int_{\frac{\theta_{(3)}}{\theta_{(2)}}x_2}^1 \int_{\frac{\theta_{(1)}}{\theta_{(3)}}x_3}^1 x_1 dx_1 dx_3 dx_2$ if the roles of $\theta_{(2)}$ and $\theta_{(3)}$ are switched. Equation (4.36) also gives the key to the general procedure for arbitrary $N > 2$ by considering all possible orderings of the ratios $\frac{X_{(2)}}{\theta_{(2)}}, \cdots, \frac{X_{(N)}}{\theta_{(N)}}$. Once again, using π, F, and Theorem 17 gives the asymptotic covariance matrix

$$\Sigma = \begin{bmatrix} 0.0232 & -0.0054 & -0.0054 \\ -0.0054 & 0.0232 & -0.0054 \\ -0.0054 & -0.0054 & 0.0232 \end{bmatrix}.$$

As before, a simulation was run that simulated 10^4 iterations of the PFS algorithm 10^4 times and the empirical covariance matrix was found to be

$$\Sigma = \begin{bmatrix} 0.0236 & -0.0053 & -0.0057 \\ -0.0054 & 0.0231 & -0.0057 \\ -0.0057 & -0.0054 & 0.0238 \end{bmatrix}$$

which is once again very close to the theoretically predicated value.

4.7 Thresholding Feedback under the PFS Algorithm

The PFS algorithm considered has each user in the system feedback their instantaneous rate. When there are many users in the system, the overhead of each user feeding back can become large. For the case when the transmitter utilizes the random beamforming algorithm and a greedy scheduling algorithm is applied ([SH05, PR10a]), a thresholding method is proposed in [PR10b] to reduce the feedback overhead as well as save transmit power at each user. The idea behind thresholding in [PR10b] is that if the statistics are known, they can be used to determine the likelihood of a user being selected for scheduling if that user were to feedback and thus if the likelihood is too low, the user can decide not to feed back. For example, when the greedy algorithm is implemented and there are many users in the system, if one user experiences a small instantaneous rate, that user should not feedback their rate information since it will most likely not be selected. A thresholding procedure will be developed in this section to reduce the system feedback overhead under the PFS scheduler.

The PFS algorithm is similar to the greedy algorithm, but is greedy with respect to the ratio of current rate to average rate. A two-stage feedback mechanism is proposed that can be used to reduce feedback on the uplink channel at the expense of introducing an additional layer of feedback in the downlink. The function determining the user scheduled for time slot $n+1$ is given by Equation 4.3. Let the base station, in addition to broadcasting the message of which user is selected, also broadcast the updated average rate vector θ^{n+1}. Notice that to notify all the users of the updated θ^{n+1} only requires sending the index of the scheduled user k and $X_{k,n+1}$ since with this information each user can produce the updated average rate vector θ^{n+1}. Therefore the additional overhead is only the value $X_{k,n+1}$ and the index k. With θ^{n+1} known at all the users, this information can be used to determine the likelihood of scheduling in the next time slot, similar to the ideas introduced in [PR10b]. The two-stage feedback mechanism functions as follows: in the first stage the transmitter feeds back the index of the currently scheduled user k and their instantaneous rate $X_{k,n+1}$, and then in the second stage each user compares some function of their instantaneous rate for the next time slot and θ^{n+1} to some pre-designed threshold and feeds back their rate if it exceeds this threshold.

The threshold that is considered is designed to address the following question: how should the threshold be designed such that the difference between the asymptotic average rate of the thresholded system compared to the asymptotic average rate without thresholding is below some design parameter R_{loss}? Under this metric, an optimal threshold for the previously mentioned two-stage feedback scheme can be found. A slight modification of this scheme will also be considered where in each time slot, the user with the minimum average rate feeds back their rate, even if it falls below the threshold. A priori, it is unclear whether applying a threshold to the PFS algorithm disrupts the algorithm's convergence proven in Sections 4.4 and 4.5. It will be shown that the thresholds considered applied to the PFS algorithm still guarantee convergence, although this convergence will no longer be to the same point as the greedy algorithm.

4.7.1 Pure Threshold

Defining the Pure Threshold

In this subsection, it is assumed that the thresholding decision is a function of the full rate vector θ^n. The goal is to design a threshold $T(N, R_{loss})$ parameterized by the number of users in the system N and the amount of asymptotic performance that is willing to be sacrificed, R_{loss}, in order to reduce the amount of feedback. The threshold will take the following form: having computed θ^n, each user will calculate the following probability

$$P_{i,n+1}^{th} = \Pr\left(\frac{X_{1,n+1}}{\theta_{1,n}} \leq \frac{X_{i,n+1}}{\theta_{i,n}}, \ldots, \frac{X_{i-1,n+1}}{\theta_{i-1,n}} \leq \frac{X_{i,n+1}}{\theta_{i,n}}, \frac{X_{i+1,n+1}}{\theta_{i+1,n}} \leq \frac{X_{i,n+1}}{\theta_{i,n}}, \ldots\right) \tag{4.37}$$

$$= \prod_{j \neq i} \Pr\left(\frac{X_{j,n+1}}{\theta_{j,n}} \leq \frac{X_{i,n+1}}{\theta_{i,n}}\right) \tag{4.38}$$

$$= \prod_{j \neq i} \Pr\left(X_{j,n+1} \leq X_{i,n+1} \cdot \frac{\theta_{j,n}}{\theta_{i,n}}\right) \tag{4.39}$$

$$= \prod_{j \neq i} F\left(X_{i,n+1} \cdot \frac{\theta_{j,n}}{\theta_{i,n}}\right) \tag{4.40}$$

In Equation 4.37, $P_{i,n+1}^{th}$ stands for the threshold probability for user i at time $n+1$, as it will be the quantity that will be compared to some threshold. The quantity on the right of Equation 4.37 is the probability that user i has the largest ratio of current rate to average scheduled rate for the next time slot $n+1$, i.e. the probability that user i will be scheduled for time slot $n+1$. Equation 4.38 is due to the i.i.d. assumption on the $X_{i,j}$, Equation 4.39 is a rearrangement of Equation 4.38, and Equation 4.40 substitutes in the definition of the CDF, where F is the distribution function for the $X_{i,j}$. Notice that the argument of the CDF in Equation 4.40 is known at user i since user i knows its own rate and all the ratios $\frac{\theta_{j,n}}{\theta_{i,n}}$.

Having defined $P_{i,n+1}^{th}$, the thresholding operation is as follows. If $P_{i,n+1}^{th} < T(N, R_{loss})$, then user i does not feed back $X_{i,n+1}$, otherwise user i feeds back $X_{i,n+1}$ to the transmitter. In order to design the threshold to meet the asymptotic rate loss constraint, it must be understood how choosing the threshold $T(N, R_{loss})$ affects the asymptotic performance of the system.

Convergence of the Thresholded Algorithm

In this section, the assumed convergence of the thresholded algorithm is addressed. Under the feedback scheme presented in Section 4.7.1 the proportional fair scheduling algorithm in Equation 4.3 can be rewritten as

$$\arg \max_{1 \le i \le N} \left\{ \frac{X_{i,n+1} \mathbf{1}_{\left[P^{th}_{i,n+1} > T(N, R_{loss})\right]}}{d_i + \theta_{i,n}} \right\}. \tag{4.41}$$

The difference between Equation 4.3 and Equation 4.41 is the inclusion of the indicator random variable $\mathbf{1}_{\left[P^{th}_{i,n+1} > T(N, R_{loss})\right]}$ that determines whether or not a user i exceeded the threshold. The product random variables $X_{i,n+1} \mathbf{1}_{\left[P^{th}_{i,n+1} > T(N, R_{loss})\right]}$ are no longer i.i.d. due to the dependence of the indicator random variable on the entire histories of the past rates captured in the mean rates θ^n. To prove the convergence of the thresholded PFS algorithm to a unique equilibrium, the effects of the threshold on the mean vector field h are analyzed.

Due to the symmetry of the i.i.d. rate random variables, without loss of generality consider the first component of the mean vector field h. From [KW04] and the past analysis in Section 4.6.2, the first component of h is given by

$$\begin{aligned} h_{(1)}(\theta) &= \mathbb{E}_X \left[I_{(1)}(\theta, X) X_{(1)} - \theta_{(1)} \right] \\ &= \int x_1 \mathbf{1}_{\left[\frac{x_1}{\theta_1} \ge \frac{x_2}{\theta_2}, \ldots, \frac{x_1}{\theta_1} \ge \frac{x_N}{\theta_N}\right]} \prod_{i=1}^{N} f(x_i) dx_i - \theta_{(1)} \end{aligned} \tag{4.42}$$

The thresholding mechanism alters the region of integration in Equation 4.42 due to the additional indicator function. Therefore, for the first component of the mean vector field under thresholding $h^{th}_{(1)}$ is given by

$$h^{th}_{(1)}(\theta) = \int x_1 \mathbf{1}_{\left[\frac{x_1}{\theta_1} \ge \frac{x_2}{\theta_2}, \ldots, \frac{x_1}{\theta_1} \ge \frac{x_N}{\theta_N}\right]} \mathbf{1}_{\left[P^{th}_i(\theta) > T(N, R_{loss})\right]} \prod_{i=1}^{N} f(x_i) dx_i - \theta_{(1)} \tag{4.43}$$

$$= \int x_1 \mathbf{1}_{\left[\frac{x_1}{\theta_1} \ge \frac{x_2}{\theta_2}, \ldots, \frac{x_1}{\theta_1} \ge \frac{x_N}{\theta_N}\right]} \mathbf{1}_{\left[\prod_{j \ne 1} F\left(\frac{x_1}{\theta_1}\theta_j\right) > T(N, R_{loss})\right]} \prod_{i=1}^{N} f(x_i) dx_i - \theta_{(1)}. \tag{4.44}$$

In Equation 4.43, $P^{th}_{i,n+1}$ is written as $P^{th}_i(\theta)$ in the indicator variable since the mean vector field is a function of the location θ in the space of average throughputs rather than the past realization of any rate random variables. To prove the uniqueness of the equilibrium point of the thresholded PFS algorithm, it is shown that the thresholded mean vector field satisfies the Kamke condition, or K-condition for short. This condition was used in the proof of a unique equilibrium in [KW04]. Once the K-condition for $h^{th}(\theta)$ is established, the implications of the K-condition will lead to the desired result. Before proceeding with the proof, some definitions are required.

Let $x, y \in \mathbb{R}_+^N$, where $\mathbb{R}_+^N$ is the set of vectors with non-negative components, and define the relation $x \leq y$ if $x_i \leq y_i$ for $i \in \{1, \ldots, N\}$, i.e. every component of x is less than or equal to every component of y. With this relation, the Kamke condition can be defined on $\mathbb{R}_+^N$.

Definition 4. *(Kamke Condition [Smi95]) A function $f : \mathbb{R}_+^N \to \mathbb{R}^N$ satisfies the Kamke condition if for any $x, y \in \mathbb{R}^N$ and $i \in 1, \ldots, N$ such that $x \leq y$ and $x_i = y_i$, then $f_i(x) \leq f_i(y)$.*

From [Smi95], it is known that f satisfies the K-condition if

$$\frac{\partial f_i}{\partial x_j}(x) \geq 0 \text{ when } i \neq j \tag{4.45}$$

Using this fact, the following Lemma is proven.

Lemma 5. *Let the i.i.d. rate random variables have a continuous density f and distribution F on R_+. Then the thresholded mean vector field $h^{th}(\theta)$ satisfies the K-condition.*

Proof. The indicator functions in the formulation of $h^{th}(\theta)$ given by Equation 4.44 determine the region of integration. Therefore as mentioned before, and noticed in [KW04], the region of integration changes continuously with θ, and because f is continuous, $h^{th}(\theta)$ will have continuous derivatives. Without loss of generality, consider the first component of $h^{th}(\theta)$ as given Equation 4.44. Fix θ except for some component θ_j, $j \neq 1$. Also, rewrite the indicator function $\mathbf{1}_{\left[\frac{x_1}{\theta_1} \geq \frac{x_2}{\theta_2}, \ldots, \frac{x_1}{\theta_1} \geq \frac{x_N}{\theta_N}\right]}$ as $\mathbf{1}_{\left[\frac{x_1}{\theta_1} \geq \frac{x_2}{\theta_2}\right]} \cdots \mathbf{1}_{\left[\frac{x_1}{\theta_1} \geq \frac{x_N}{\theta_N}\right]}$. As a function of only θ_j, $h^{th}(\theta_j)$ is monotonically nondecreasing. To see this, notice that more values of x_1 satisfy $\mathbf{1}_{\left[\frac{x_1}{\theta_1} \geq \frac{x_j}{\theta_j}\right]}$ as θ_j increases, the other indicator terms are independent of θ_j and remain unchanged, and more values of x_1 satisfy $\mathbf{1}_{\left[\prod_{j \neq 1} F\left(\frac{x_1}{\theta_1}\theta_j\right) > T(N, R_{loss})\right]}$ as θ_j increases since $F\left(\frac{x_1}{\theta_1}\theta_j\right)$ is nondecreasing in θ_j. Thus, the region of integration becomes larger as θ_j increases, and hence the integral representing $h^{th}_{(1)}$ in Equation 4.44 is nondecreasing since the integrand is nonnegative and the constant term $\theta_{(1)}$ disappears when the derivative is taken w.r.t. θ_j, $j \neq 1$. Because $h^{th}_{(1)}$ is a nondecreasing function of θ_j, the partial derivative w.r.t. θ_j is non-negative. By Equation 4.45, $h^{th}(\theta)$ satisfies the K-condition. □

With the previous Lemma established, the main result of this section regarding the convergence of the thresholded PFS algorithm is now established in two parts. The first result shows that there exists an equilibrium point for the thresholded PFS algorithm of the form $\theta^*_{th} = c \cdot \mathbf{1}_N$ where c is some non-negative constant and $\mathbf{1}_N$ is an N-dimension vector of 1s. The result proves that this equilibrium point is in fact unique and stable. The first result is now stated in the following corollary:

Corollary 7. *Under the same conditions as Theorem 15, the thresholded PFS algorithm of Section 4.7.1 with fixed threshold T has an equilibrium point of the form $\theta^*_{th} = c \cdot \mathbf{1}_N$. Additionally the constant c is given by the expression $\frac{\mathbb{E}[X_{scheduled}]}{N}$ where*

$$\mathbb{E}[X_{scheduled}] = \left(1 - [F(T)]^N\right) \cdot T + \int_T^\infty \left(1 - [F(x)]^N\right) dx.$$

Proof. To show that θ^*_{th} is an equilibrium point is equivalent to showing that θ^*_{th} is a zero of the mean vector field h. Notice when θ^*_{th} has the form $\theta^*_{th} = c \cdot \mathbf{1}_N$ that all the components of the fed back average rate vector are equal, so Equation 4.40 becomes

$$P^{th}_{i,n+1} = \prod_{j \neq i} F\left(X_{i,n+1}\right) = \left[F(X_{i,n+1})\right]^{N-1} \tag{4.46}$$

which is the probability that all other users are experiencing rates smaller than the rate of user i, i.e. the probability that user i has the maximum instantaneous rate. Thus, the i^{th} component of the mean vector field evaluated at θ^*_{th} of the proposed form evaluates to

$$h^{th}_{(i)} = \int x_i \mathbf{1}_{[\arg\max\{x_1,\ldots,x_N\}=i]} \mathbf{1}_{\left[[F(x_i)]^{N-1} > T\right]} \prod_{i=1}^{N} f(x_i) dx_i - \theta_{(i)}. \tag{4.47}$$

The key property of Equation 4.47 is that the integral term no longer depends on θ and by the symmetry of the i.i.d. model, the integral terms are equivalent for each $i \in \{1, \ldots, N\}$. Since all the random variables are non-negative, the integral term in Equation 4.47 evaluates to some non-negative number, so there is some non-negative $\theta_{(i)}$ that causes the equation to evaluate to zero, and all the $\theta_{(i)}$s are equivalent. This shows that there exists some c such that $\theta^*_{th} = c \cdot \mathbf{1}_N$ is a zero of the mean vector field. The next step is to characterize the constant c.

Notice that $P^{th}_{i,n+1}$ is a random variable, and recall that for any random variable X with distribution function F_X that $F_X(X) =_d U$, where U is a uniform random variable and $=_d$ mean equals in distribution. Therefore at $\theta^*_{th} = c \cdot \mathbf{1}_N$, from Equation 4.46, $P^{th}_{i,n+1} =_d U^{N-1}$. Using this fact for the threshold T, the probability at $\theta^*_{th} = c \cdot \mathbf{1}_N$ that $P^{th}_{i,n+1} \leq T$ can be explicitly calculated as follows:

$$\begin{aligned} \Pr\left[P^{th}_{i,n+1} \leq T\right] &= \Pr\left[\left(F(X_{i,n+1})\right)^{N-1} \leq T\right] \\ &= \Pr\left[F(X_{i,n+1}) \leq T^{\frac{1}{N-1}}\right] \\ &= \Pr\left[U \leq T^{\frac{1}{N-1}}\right] \\ &= T^{\frac{1}{N-1}}. \end{aligned} \tag{4.48}$$

Equation 4.48 yields the probability of a user not feeding back at $\theta^*_{th} = c \cdot \mathbf{1}_N$ for a threshold T. Let F be the distribution function for a non-negative random variable X. The distribution function of the thresholded random variable

$$X_{th} = \begin{cases} X & X \geq T \\ 0 & X < T \end{cases} \tag{4.49}$$

is F_{th}, used extensively in [PR10b], and is given by

$$F_{th}(x) = \begin{cases} F(x) & x \geq T \\ F(T) & 0 \leq x < T \\ 0 & x < 0 \end{cases} \tag{4.50}$$

Equation 4.50 can be interpreted as the probability mass less than T concentrates at zero, which is the probability that the user does not feedback. Figure 4.4 gives an example of the truncated distribution function a uniform random variable being truncated at $T = 0.5$.

Recall from Sections 4.4 and 4.5, that when $\theta^*_{th} = c \cdot \mathbf{1}_N$, the scheduling rule given by Equation 4.3 becomes the greedy schedule rule. Thus at $\theta^*_{th} = c \cdot \mathbf{1}_N$, if any

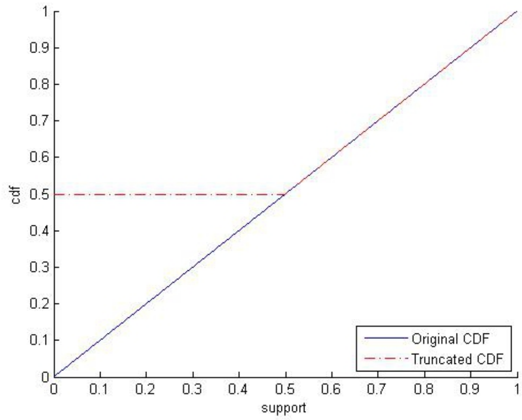

Figure 4.4: Uniform Distribution Truncated at $T = 0.5$

user experiences $P_{i,n+1}^{th} > T$ and feeds back $X_{i,n+1}$, then the maximum ratio will be known at the transmitter. To see this, notice that $P_{i,n+1}^{th}$ at $\theta_{th}^{*} = c \cdot \mathbf{1}_N$ given by Equation 4.46 is monotonic in $X_{i,n+1}$ due to the monotonicity of distribution functions. Thus if at least one user exceeds the threshold at $\theta_{th}^{*} = c \cdot \mathbf{1}_N$, the thresholded algorithm performs the same as if there was no threshold.

With the previous definitions and observations in hand, the performance of the thresholded system can be determined for a fixed threshold T. Equation 4.50 is the distribution function of the fed back rate random variable at equilibrium. At $\theta_{th}^{*} = c \cdot \mathbf{1}_N$, the PFS algorithm selects the largest rate, so the distribution of the scheduled rate is given by

$$F_{th}^{max}(x) = \begin{cases} [F(x)]^N & x \geq T \\ [F(T)]^N & 0 \leq x < T \\ 0 & x < 0 \end{cases} \tag{4.51}$$

Using Equation 4.51, the mean of the scheduled rate at $\theta_{th}^{*} = c \cdot \mathbf{1}_N$ can be computed as follows:

$$\begin{aligned} \mathbb{E}\left[X_{scheduled}\right] &= \int_0^{\infty} \left(1 - F_{th}^{max}(x)\right) dx && (4.52) \\ &= \int_0^{T} \left(1 - [F(T)]^N\right) dx + \int_T^{\infty} \left(1 - [F(x)]^N\right) dx && (4.53) \\ &= \left(1 - [F(T)]^N\right) \cdot T + \int_T^{\infty} \left(1 - [F(x)]^N\right) dx. && (4.54) \end{aligned}$$

Equation 4.54 is the value of c given by the corollary. $\square$

Notice that $\mathbb{E}\left[X_{scheduled}\right]$ is always less than or equal to the performance of the unthresholded PFS algorithm at equilibrium, or equivalently the greedy algorithm, since in Equation 4.53

$$\int_0^T \left(1 - [F(T)]^N\right) dx \leq \int_0^T \left(1 - [F(x)]^N\right) dx$$

for any T since F, by definition, is monotonically nondecreasing.

Corollary 7 shows that there is at least one zero to the mean vector field. It remains to show that this is in fact the only equilibrium point and that the PFS algorithm converges to it. This result is given in the following corollary.

Corollary 8. *Under the same conditions as Theorem 15, the thresholded PFS algorithm of Section 4.7.1 with fixed threshold T converges to* $\frac{\mathbb{E}[X_{scheduled}]}{N}$.

Proof. From the calculations done in Section 4.7.1, there exists at least one zero to the thresholded mean vector field $h^{th}(\theta)$, namely $\theta = \frac{\mathbb{E}[X_{scheduled}]}{N}$. The goal is to show that this is in fact the only equilibrium point. The proof of the unique equilibrium point given in [KW04] is composed of two parts. Suppose θ^* is the equilibrium point and let $\theta(0)$ be the initial condition of the PFS algorithm. The first part of the proof shows that for all starting points $\theta(0) \leq \theta^*$, the algorithm converges to θ^* as $t \to \infty$. The second part of the proof considers starting points $\theta(0) \geq \theta^*, \theta(0) \neq \theta^*$ and shows that these points also converge to θ^* asymptotically in time. The monotonicity theorem given by Theorem 4.1 in [KW04] covers the remaining possibilities, so that all possible initial conditions $\theta(0)$ converge to θ^* as $t \to \infty$. By Lemma 5, the the first part of the proof and Theorem 4.1 in [KW04] hold, so it remains to show that for $\theta(0) \geq \theta^*, \theta(0) \neq \theta^*$ the PFS algorithm converges to θ^*.

Let $\theta^* = \frac{\mathbb{E}[X_{scheduled}]}{N}$ and define the set $Q(\theta) = \{x : x \geq \theta\}$. Consider the set $\widetilde{Q}(\theta^*) = Q(\theta^*) - \theta^*$, i.e. the set $Q(\theta^*)$ with θ^* removed. For every point $\theta \in \widetilde{Q}(\theta^*)$, there is a point $\widehat{\theta} = c \cdot \mathbf{1}_N$, where $\mathbf{1}_N$ is a N-dimensional vector of ones and c a positive scalar, such that $\theta \leq \widehat{\theta}$. For $\widehat{\theta}$, the integral part of $h^{th}_{(1)}$ given in Equation

4.44 equals

$$\int x_1 \mathbf{1}_{\left[\frac{x_1}{\theta_1} \geq \frac{x_2}{\theta_2}, \ldots, \frac{x_1}{\theta_1} \geq \frac{x_N}{\theta_N}\right]} \mathbf{1}_{\left[\prod_{j \neq 1} F\left(\frac{x_1}{\theta_1}\theta_j\right) > T(N, R_{loss})\right]} \prod_{i=1}^{N} f(x_i) dx_i = \frac{\mathbb{E}[X_{scheduled}]}{N}, \tag{4.55}$$

which is true of each component of h^{th}. This is because when all the components of $\widehat{\theta}$ are equal, the integral is the expectation of that component being the maximum and exceeding the threshold, which is given by $\frac{\mathbb{E}[X_{scheduled}]}{N}$. Because $\widehat{\theta} \in \widetilde{Q}(\theta^*)$, $\widehat{\theta}_i \geq \frac{\mathbb{E}[X_{scheduled}]}{N}$ and thus when evaluating $h^{th}(\widehat{\theta})$ given by Equation 4.44 each component of $h^{th}(\widehat{\theta})$ is negative. Thus by the K-condition guaranteed by Lemma 5, for each $\theta \in \widetilde{Q}(\theta^*)$, $h^{th}(\theta)$ has strictly negative components. As in [KW04], this combined with monotonicity implies that all paths with initial condition $\theta(0) \in Q(\theta^*)$ converge to θ^*. Therefore $\theta^* = \frac{\mathbb{E}[X_{scheduled}]}{N}$ is the unique equilibrium of the thresholded PFS algorithm. □

The previous results show that for any threshold T that there is a unique equilibrium that the PFS algorithm converges to. In the next section it is shown how to select T to satisfy the rate loss constraint R_{loss}.

Threshold Design

The definition of $\mathbb{E}[X_{scheduled}]$ given in Equation 4.52 was for a fixed threshold T. The goal is to design a threshold $T(N, R_{loss})$ that is parameterized by the number of users and the acceptable loss. Let $R_{loss} \in [0, \mathbb{E}[X_{max}]]$ represent the amount of asymptotic performance that can be acceptably lost due to thresholding. X_{max} is the maximum order statistic. R_{loss} cannot exceed $\mathbb{E}[X_{max}]$ because this is the asymptotic performance of the system without a threshold as previously proven. The desired threshold $T(N, R_{loss})$ will be found in two steps. First, define a temporary threshold T' that is also parameterized by N and R_{loss} that satisfies the following equation:

$$\begin{aligned} R_{loss} &= \mathbb{E}\left[X_{max}\right] - \mathbb{E}\left[X_{scheduled}\right] \\ &= \int_0^\infty \left(1 - F^{max}(x)\right) dx - \int_0^\infty \left(1 - F^{max}_{th}(x)\right) dx \\ &= \int_0^{T'} \left(1 - F(x)^N\right) dx - \int_0^{T'} \left(1 - F(T)^N\right) dx \\ &= T'F(T')^N - \int_0^{T'} F(x)^N dx \end{aligned} \tag{4.56}$$

An iterative algorithm, or in some case a close form solution, can be used to find the value of T' that satisfies Equation 4.56.

The quantity T' is an intermediary because the random variables themselves are not compared to T' to determine whether or not to feedback; the value that is compared to a threshold is $P^{th}_{i,n+1}$. To achieve the rate loss R_{loss} given by T', $T(N, R_{loss})$ should be chosen such that the probability of a user not feeding back at equilibrium equal T', i.e.

$$\Pr\left[P^{th}_{i,n+1} \leq T(N, R_{loss})\right] = T'. \tag{4.57}$$

Using Equation 4.48, the threshold $T(N, R_{loss})$ can be found as follows:

$$\begin{aligned} \Pr\left[P^{th}_{i,n+1} \leq T(N, R_{loss})\right] &= T(N, R_{loss})^{\frac{1}{N-1}} = T' \\ &\Rightarrow T(N, R_{loss}) = T'^{N-1}. \end{aligned} \tag{4.58}$$

At equilibrium, the probability that a user does not feedback is given by T', thus the number of users that do not feedback is a binomial random variable with parameters N and T'. Thus for this thresholding scheme, the average number of users not feeding back is given by NT', which is the asymptotic average savings in feedback. Equation 4.58 can seem abstract, so the following example will provide a closed form solution for the case of i.i.d. uniform random variables.

Example 3.

Let there be N users, each experiencing i.i.d. uniform random variables and let $R_{loss} \in [0, \mathbb{E}\left[X_{max}\right]]$ be the desired loss. The first task is to find the precursor

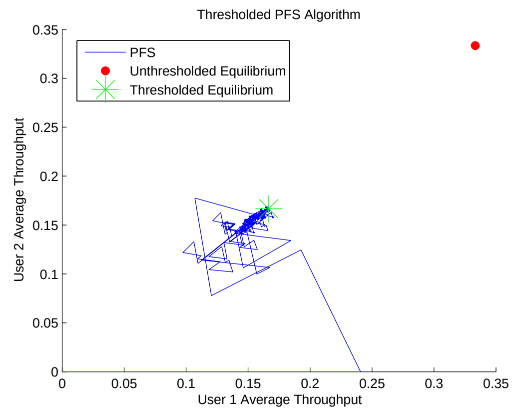

Figure 4.5: Threshold for $N = 2$, $R_{loss} = \frac{1}{3}$

threshold, T' as indicated by Equation 4.56. This can be done as follows:

$$\begin{aligned} R_{loss} &= T'F(T')^N - \int_0^{T'} F(x)^N dx \\ &= T'^{N+1} - \frac{T'^{N+1}}{N+1} \\ &= \frac{N}{N+1} T'^{N+1} \\ \Rightarrow T' &= \left[\frac{N+1}{N} R_{loss}\right]^{\frac{1}{N+1}} \end{aligned} \tag{4.59}$$

Having solved for T', the threshold $T(N, R_{loss})$ can be found using Equation 4.58:

$$T(N, R_{loss}) = T'^{N-1} = \left[\frac{N+1}{N} R_{loss}\right]^{\frac{N-1}{N+1}}. \tag{4.60}$$

Figure 4.5 shows an example where there are $N = 2$ users with i.i.d. uniform random variables and $R_{loss} = \frac{1}{3}$. In this case, The expected asymptotic throughput of the unthresholded PFS algorithm is $\mathbb{E}[X_{max}] = \frac{2}{3}$, and since there are $N = 2$ users, this yields an unthresholded equilibrium point of $[1/3, 1/3]^T$ in Figure 4.5. With $R_{loss} = \frac{1}{3}$, the asymptotic throughput of the thresholded PFS algorithm is expected to be $\frac{1}{3}$, yielding an equilibrium point for the thresholded algorithm of $[1/6, 1/6]^T$, which is indeed the equilibrium point shown in Figure 4.5. A sample path of the thresholded PFS algorithm with the threshold given by Equation 4.60 is also plotted in Figure 4.5.

4.7.2 Modified Threshold

The thresholding scheme presented in Section 4.7.1 can immediately be improved upon by allowing slightly more feedback. In the previous scheme, when no user exceeds the threshold, it is assumed that throughput is zero for that scheduling epoch. The modified thresholding scheme of this section assumes that the user with the smallest average throughput at time period $n+1$ will always feedback their observed rate random variable. The users can determine if they have the smallest average throughput because average rate vector θ^n is known. Therefore, the feedback scheme in this section is identical to that of Section 4.7.1 except for the fact that the user with the smallest average throughput always feeds back its rate. As in Section 4.7.1, the convergence of this thresholding scheme will be proven via two corollaries; one corollary showing the existence of a zero of the mean vector field and one corollary showing that this equilibrium is the only point of convergence of the algorithm.

Corollary 9. *Under the same conditions as Theorem 15, the thresholded PFS algorithm just described with fixed threshold T has an equilibrium point of the form $\theta^*_{th} = c \cdot \mathbf{1}_N$. Additionally the constant c is given by the expression $\frac{\mathbb{E}[X'_{scheduled}]}{N}$ where*

$$\mathbb{E}[X'_{scheduled}] = \mathbb{E}_Y[Y] = \int_0^\infty (1 - F_y(y))\, dy$$

and

$$F_Y(y) = \begin{cases} F_X(y) \cdot F_X(T)^{N-1} & \text{if } y \le T \\ F_{max}(y) & \text{if } y > T \end{cases}$$

Proof. There are two cases to consider when analyzing this scheme: when $P^{th}_{i,n+1}$ exceeds the threshold for at least one user and when no user exceeds the threshold. When no user exceeds the threshold, the central controller will select the rate of the user with the smallest average throughput. In the case when at least one user exceeds the threshold, than as previously mentioned, at $\theta^*_{th} = c \cdot \mathbf{1}_N$ the true maximum rate will have been fed back. Consider a fixed threshold and define the following random variable:

$$Y = \begin{cases} X\mathbf{1}_{X \le T} & \text{if } X_{max} \le T \\ X_{max} & \text{if } X_{max} > T \end{cases} \tag{4.61}$$

where X is the rate random variable, $\mathbf{1}$ is an indicator random variable and X_{max} is the maximum of n i.i.d. copies of X. At $\theta^*_{th} = c \cdot \mathbf{1}_N$, the random variable Y defined above is the random variable that represents the rate selected by the central controller. When $X_{max} \leq T$, then the user with the smallest average throughput will feed back its instantaneous rate, which is distributed as X conditioned on $X \leq T$, and when $X_{max} > T$, the true max is known, conditioned on $X_{max} > T$, and chosen at the controller.

The distribution function of Y is given by $F_Y(y) = \Pr[Y \leq y]$, and this can be broken into two cases depending on if $y \leq T$ or $y > T$. When $y \leq T$, then

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] = \Pr[X \leq y | X \leq T] \cdot \Pr[X_{max} \leq T] \\ &= \frac{F_X(y)}{F_X(T)} \cdot F_{max}(T) = \frac{F_X(y)}{F_X(T)} \cdot F_X(T)^N \\ &= F_X(y) \cdot F_X(T)^{N-1} \end{aligned} \tag{4.62}$$

where $\Pr[X \leq y | X \leq T] = \frac{F_X(y)}{F_X(T)}$ and $\Pr[X_{max} \leq T] = F_X(T)^N$. Likewise, when $y > T$, then the distribution function is given by

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] = \Pr[X_{max} \leq y | X_{max} > T] \cdot \Pr[X_{max} > T] \\ &= \frac{F_X(y)^N}{1 - F_X(T)^N} \cdot (1 - F_X(T)^N) = F_X(y)^N = F_{max}(y). \end{aligned} \tag{4.63}$$

To combine these results yields:

$$F_Y(y) = \begin{cases} F_X(y) \cdot F_X(T)^{N-1} & \text{if } y \leq T \\ F_{max}(y) & \text{if } y > T \end{cases} \tag{4.64}$$

Figure 4.6 gives an example of $F_Y(y)$ and $F_{X_{max}}(x)$ when $X \sim U[0,1]$, $N = 2$, and $T = 0.5$. The linear portion up to $y = 0.5$ corresponds with $F_X(y) = y$ for the uniform distribution and it is scaled by $F_X(T)^{N-1} = 0.5$. After $y = 0.5$, the distribution is equal to the distribution of the maximum order statistic, which in this case is $F_{max}(y) = y^2$.

Let $X'_{scheduled}$ be the random rate selected by the scheduler at $\theta^*_{th} = c \cdot \mathbf{1}_N$ under this modified thresholding scheme. The average throughput is given by

$$\mathbb{E}[X'_{scheduled}] = \mathbb{E}_Y[Y] = \int_0^\infty (1 - F_y(y))\, dy \tag{4.65}$$

This is the analog of Equation 4.52 in Section 4.7.1. □

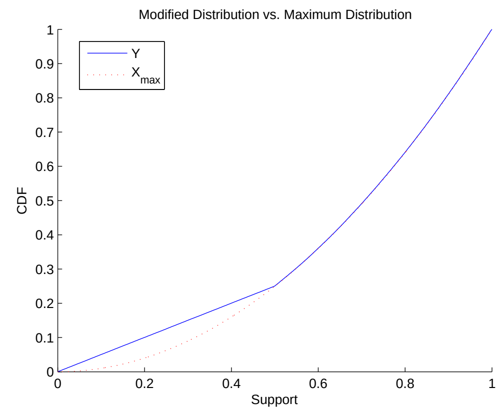

Figure 4.6: Modified CDF vs. Maximum CDF

Aside: The distribution used in the design of $T(N, R_{loss})$ in Section 4.7.1 is given by Equation 4.51. The methods used here in deriving Equation 4.64 are a generalization of Equation 4.51, i.e. Equation 4.51 can be derived from these methods. In Section 4.7.1, if no user exceeds the threshold, then a rate of zero is assigned for that time period. This is equivalent to setting $Y = 0$ if $X_{max} \leq T$ in Equation 4.61. Therefore, computing Equation 4.62 in this case yields

$$F_Y(y) = \Pr[Y \leq y] = \Pr[0 \leq y | X \leq T] \cdot \Pr[X_{max} \leq T] = F_X(T)^N,$$

which is identical to the result given in Equation 4.51 when $X_{max} \leq T$. When $X_{max} > T$, the two equations are clearly identical.

Having addressed the existence of an equilibrium point with Corollary 9, the following corollary says that this is the only equilibrium, and the modified thresholding scheme converges to this point.

Corollary 10. *Under the same conditions as Theorem 15, the modified thresholded PFS algorithm of Section 4.7.2 with fixed threshold T converges to $\frac{\mathbb{E}[X'_{scheduled}]}{N}$ as given by Equation 4.65.*

Proof. The same arguments as the proof of Corollary 8 yield the desired result. □

Corollaries 9 and 10 analyze where the PFS algorithm converges to under the modified thresholding scheme for a fixed threshold T. The goal is to design

T as a function of the number of users N such that a rate loss constraint R_{loss} is satisfied. The solution to the design of $T(N, R_{loss})$ in section 4.7.1 can be applied to this modified feedback scheme using the distribution from Equation 4.64 and the expectation in Equation 4.65. That is to say, for a given N and R_{loss}, first compute the precursor threshold T' using these equations, and then use Equation 4.58 to find $T(N, R_{loss})$. An important note is that for this scheme $R_{loss} \in [0, \mathbb{E}[X]]$, rather than in the previous case when $R_{loss} \in [0, \mathbb{E}[X_{max}]]$. This is because the user with the minimum average throughput always feeds back their observed instantaneous rate and thus the system on average cannot perform worse than $\mathbb{E}[X]$.

For this modified scheme let $N'_{feedback}$ be the number of users feeding back rate information at equilbrium. The average number of users feeding back for this scheme is given by

$$\mathbb{E}[N'_{feedback}] = 1 + (N-1) \cdot (1 - T'). \tag{4.66}$$

The constant term 1 is due to the fact that at least one user is always feeding back (the user with the smallest average rate), and the other $N-1$ users feedback if their rate exceeds the threshold. This probability is given by T' as shown in Equation 4.57. Therefore the average asymptotic savings in the amount of feedback of this scheme is given by

$$N - \mathbb{E}[N'_{feedback}] = N - (1 + (N-1) \cdot (1 - T')) = (N-1)T'. \tag{4.67}$$

For concreteness, Example 3 is revisited under the new modified thresholding scheme.

Example 4.

Under the same set up as Example 3, i.e. $N = 2$ users experiencing i.i.d. uniform random variables, consider applying the same threshold but under the modified framework. From Example 1, the threshold is $T\left(2, \frac{1}{3}\right) = \left(\frac{1}{2}\right)^{\frac{1}{3}}$ and plugging this into Equation 4.65 gives an asymptotic throughput of $\frac{7}{12}$, corresponding to an equilibrium point of $[7/24, 7/24]^T$. Figure 4.7 shows a sample run of the thresholded PFS algorithm, and the equilibrium point matches exactly what the theory predicts. Also notice that for the same threshold, the modified scheme has

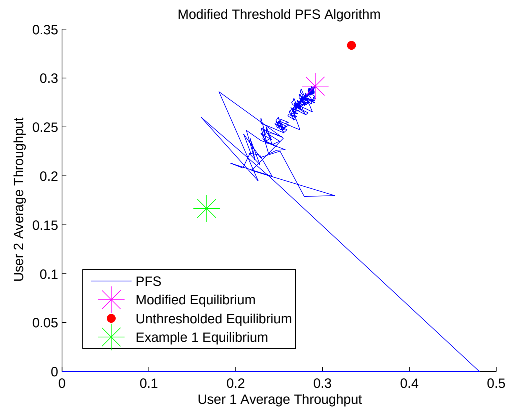

Figure 4.7: Modified Threshold for $N = 2$ and $T = \left(\frac{1}{2}\right)^{\frac{1}{3}}$

a larger rate than the original scheme due to the fact that at least one user always feeds back, which mathematically corresponds to $X_{scheduled} \leq X'_{scheduled}$ for the same threshold.

This comparison is not fair since the threshold was designed for a specific R_{loss} under the feedback scheme in Section 4.7.1. In Example 3, $R_{loss} = \frac{1}{3}$, but in this case $R_{loss} > \mathbb{E}[X]$, which is the worst case scenario as previously described. Choosing $R_{loss} = \frac{1}{12}$ and working through the machinery provides $T(2, \frac{1}{12}) = \left(\frac{1}{2}\right)^{\frac{1}{3}}$ which is the same threshold used in Example 3.

4.8 Additional Thresholding Methods

In Section 4.7 the threshold was based on the idea that only the users who are most likely to be scheduled should be fed back. The threshold was mathematical quantified by the quantity given in Equation 4.40 and the theory was developed from this point of view. In this section, additional thresholding mechanisms are developed and analyzed. The first new method is based on the observation that the metric defined in Equation 4.40 depends on $n-1$ marginal distributions and thus computationally it may be expensive to compute or requires the use of $n-1$ table look ups. A simpler thresholding metric will be based upon a single marginal distribution. The three thresholds considered in Section 3.5 of Chapter 3 will be applied to a system running the proportional fair sharing scheduler.

4.8.1 Minimum Average Rate Feedback

Pure Threshold

In this subsection, the threshold is a function of the minimum average throughput

$$\theta_{\min,n} = \min\{\theta_{1,n}, \ldots, \theta_{N,n}\}$$

rather than the full average rate vector θ^n as in Section 4.7.1. This scheme can reduce computational complexity since accurate evaluation of $P^{th}_{i,n+1}$ given by Equation 4.40 may be difficult. (Note: Assuming a particular distribution on the rates, Equation 4.40 can be implemented as a table look up). Once again, the goal is to design a threshold $\widetilde{T}(N, R_{loss})$ to help reduce the amount of feedback from the users to the transmitter. Having received $\theta_{\min,n}$, each user will compute the following probability:

$$\widetilde{P}^{th}_{i,n+1} = \Pr\left[\frac{X_{\min,n+1}}{\theta_{\min,n}} < \frac{X_{i,n+1}}{\theta_{i,n}}\right] \tag{4.68}$$

where $X_{\min,n+1}$ is the random variable associated with the user with the smallest average throughput. The value of $X_{\min,n+1}$ is of course unknown at all of the users, but under the i.i.d. assumption $X_{\min,n+1} \sim F$, and to avoid confusion $X_{\min,n}$ will be denoted as $X \sim F$. Equation 4.68 is the analog to Equation 4.40 for this scheme.

The computed value of $\widetilde{P}^{th}_{i,n+1}$ will be compared to the threshold $\widetilde{T}(N, R_{loss})$ at each user. If the i^{th} user's $\widetilde{P}^{th}_{i,n+1}$ exceeds the threshold, then $X_{i,n+1}$ is fed back to the central controller, otherwise nothing is fed back. By the monotonicity of Equation 4.68 as a function of θ_{min}, n (and thus monotonically non-decreasing as a function of $\theta_{j,n}, j \neq i$), it is straight forward to show that this thresholding scheme converges using the same arguments as Corollary 7 and Corollary 8. The equilibrium point is once again of the form $\theta^*_{th} = c \cdot \mathbf{1}_N$ and for a fixed parameterization (N, R_{loss}) the constant c is equivalent to c from Section 4.7.1 because in both cases the thresholds are designed to meet some rate loss constraint for a given number of users in the system. Thus knowing that the algorithm converges, and knowing where it converges to, the last challenge the actual design of the threshold $\widetilde{T}(N, R_{loss})$, which will now be addressed.

At equilibrium, all of the components of the average rate vector θ^n are equal, so Equation 4.68 becomes

$$\widetilde{P}^{th}_{i,n+1} = \Pr\left[X < X_{i,n+1}\right], \tag{4.69}$$

which is a random variable. The probability that $\widetilde{P}^{th}_{i,n+1}$ falls below the desired threshold $\widetilde{T}(N, R_{loss})$ is critical in the design of the threshold. To compute this probability, first notice that Equation 4.69 can be rewritten as:

$$\begin{aligned} \widetilde{P}^{th}_{i,n+1} &= 1 - \Pr\left[X_{i,n+1} \leq X\right] \\ &= 1 - F_{X_{i,n+1}}(X) = 1 - F_X(X) \\ &=_d 1 - U =_d U \end{aligned} \tag{4.70}$$

where $F_{X_{i,n+1}} = F_X$ under the i.i.d. assumption and U is a uniform random variable. The probability that $\widetilde{P}^{th}_{i,n+1}$ does not exceed the threshold can now be computed using Equation 4.70 as follows:

$$\begin{aligned} \Pr\left[\widetilde{P}^{th}_{i,n+1} \leq \widetilde{T}(N, R_{loss})\right] &= \Pr\left[1 - U \leq \widetilde{T}(N, R_{loss})\right] \\ &= \Pr\left[1 - \widetilde{T}(N, R_{loss}) \leq U\right] \\ &= 1 - \Pr\left[U \leq 1 - \widetilde{T}(N, R_{loss})\right] \\ &= 1 - \left(1 - \widetilde{T}(N, R_{loss})\right) \\ &= \widetilde{T}(N, R_{loss}). \end{aligned} \tag{4.71}$$

Comparing this to Equation 4.48, the difference in using the entire average rate vector θ^n versus $\theta_{min,n}$ is the exponent of $\frac{1}{N-1}$ versus 1.

Having established Equation 4.71, the computation of $\widetilde{T}(N, R_{loss})$ follows the same procedure as Section 4.7.1. The characterization of the precursor threshold T' for a desired R_{loss} remains unchanged and is given by Equation 4.56. To find $\widetilde{T}(N, R_{loss})$, the analogous version of Equation 4.57 is evaluated:

$$\Pr\left[\widetilde{P}^{th}_{i,n+1} \leq \widetilde{T}(N, R_{loss})\right] = \widetilde{T}(N, R_{loss}) = T', \tag{4.72}$$

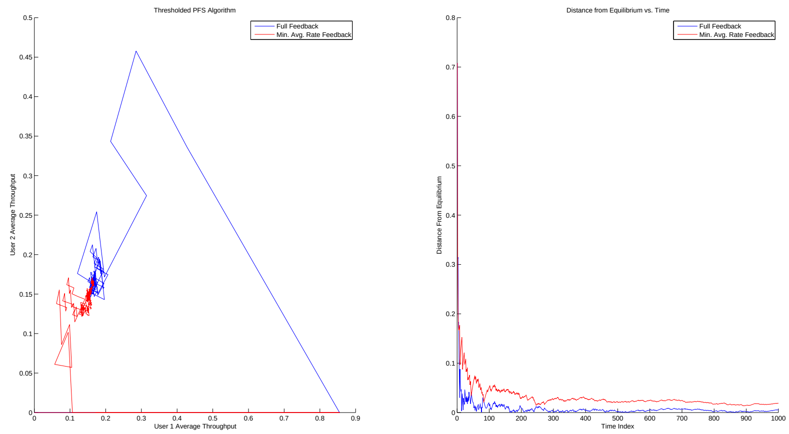

Figure 4.8: Comparing Threshold Schemes for $N = 2$ and $R_{loss} = \frac{1}{3}$

where the first equality comes from Equation 4.71. Equation 4.57 says that the threshold should be set to the solution of T' for a given N and R_{loss}, as opposed to some power of T' as given in Section 4.7.1. For both the schemes considered here and in Section 4.7.1, the average savings in feedback from the users at equilibrium is given by NT'. Therefore, under the metric of designing a threshold as a function of N and R_{loss}, the scheme presented in this section is superior in the sense that it achieve the same performance for less computation. However, while the two schemes considered here perform identically at equilibrium, when the system is starting off, their sample paths and performances will differ. Figure 4.8 shows the two schemes run on the same realizations of the random rates. The plot on the right shows the distance from the equilibrium point. On this realization, the convergence of the scheme utilizing the full rate vector θ^n converges faster than when only $\theta_{\min,n}$ is used. While the relation of the convergence properties of the two schemes depend on the realization, these authors have noticed generally the full feedback scheme converges faster. Because of this, it is possible to use a hybrid scheme that combines the two methods. In the transient initialization phase, the scheme where θ_n is used to compute the $n-1$ marginal distribution is run, and after a significant time, the system switches over to the scheme where only $\theta_{\min,n}$ is used.

Modified Threshold

In this subsection the previous feedback scheme where the threshold is only a function of $\theta_{\min,n}$ is modified in the same way as Section 4.7.2. Because $\theta_{\min,n}$ is computed and known at each user, and each user computes its own average throughput, each user knows if it has the minimum average throughput. If the user's average throughput equals the minimum average throughput $\theta_{\min,n}$, then that user will feed back its instantaneous rate regardless of whether or not $\widetilde{P}_{i,n+1}^{th}$ exceeds the threshold.

The analysis of this modified scheme is the same as Section 4.7.2. The distribution of the selected rates at equilibrium is once again given by Equation 4.64, and thus the computation to find the precursor threshold T' is the same. With T' computed for the parameters N and R_{loss} , Equation 4.72 yields that the threshold should be set to T'. The fact that the threshold $\widetilde{T}(N, R_{loss})$ is set to T' rather than T'^{N-1} as in Section 4.7 is in fact the only formal difference between this section and that of Section 4.7.2. The average savings in feedback from the users at equilibrium is given by Equation 4.67.

4.8.2 Thresholds for Random Beamforming

All the results of this chapter have been developed for any general system utilizing a proportional fair sharing scheduling algorithm with i.i.d. rates. In Chapter 3, three different thresholding metrics were proposed to help reduce feedback in a random beamforming transmit system utilizing the greedy scheduling algorithm. In the random beamforming context, the three thresholding metrics will be considered under the proportional fair sharing algorithm.

The difference between the thresholding methods of Chapter 3 and this chapter is that in Chapter 3 the observation random variables are compared directly to a threshold as opposed to the thresholds presented previously in this chapter where a probability was compared to a threshold. All the thresholds derived in Chapter 3 lead to distribution functions for the random variables at the scheduler of the form given by Equation 4.49. Therefore, under the thresholds considered in Chapter 3, the random variables can be thought of as i.i.d. random variables with a thresholded distribution. Therefore, if the proportional fair sharing algorithm is used instead of the greedy scheduler, the results of Section 4.5 guarantee that the rates converge to same average throughput as the greedy algorithm. Therefore, under the same design criterion as those considered in Chapter 3, the same thresholds can be used under the proportional fair sharing algorithm and yield asymptotically in time the same results. For example, consider restricting the average number of users feeding back, that is on average only $k \leq N$ of the users should provide feedback. Using Equation 3.18 the threshold is designed. This threshold can be used under the PFS scheduling algorithm and the average throughput of each user will converge to the results of Chapter 3.

4.9 Conclusion

This chapter addresses the asymptotic performance of the proportional fair sharing algorithm under i.i.d. rate models. Under this model it is shown in the state space of average throughput and under suitable conditions on the distribution of the rates that the PFS algorithm converges to the same fixed point as the greedy scheduling algorithm. Therefore, from the point of view of average throughput, asymptotically in time there is no difference between using the PFS algorithm and the greedy algorithm.

Under i.i.d. models, it is shown that the rate of convergence of the PFS algorithm to the equilibrium average throughput point is essentially the fastest rate possible for a stochastic approximation procedure. The distance of the PFS algorithm from the equilibrium point at any given time is a random variable. Asymptotically in time, it is known that this distance becomes distributed as a zero mean normal random variable whose covariance matrix is characterized under the i.i.d. model.

To reduce the amount of CSI feedback overhead in the system, a thresholding scheme is analyzed. Under this scheme, the equilibrium point of the PFS algorithm is found. In an effort to increase the average throughput of the thresholded system, a modification is made that allows the user with the smallest average throughput to always provide feedback. Once again, under this modification, the equilibrium point of the system is characterized.

This chapter contains material in preparation for submission to *IEEE Transactions on Information Theory* titled *"Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models"*. This work is coauthored with Professor Bhaskar D. Rao.

Chapter 5

Conclusion

Design of efficient methods to effectively utilize the multi-user diversity and maximize capacity of the broadcast channel is a very challenging problem. One central aspect of any high performance solution to this problem is the requirement that the transmitter have channel state information about the users. Under the random beamforming transmission scheme, the CSI is represented as a single scalar quantity, the signal-to-interference plus noise ratio. With this metric in mind, the question confronted by this work is how to reduce the amount of feedback in the system yet still provide good system performance. Several approaches to this question have been analyzed. One method is to reduce feedback by exploiting the spatial diversity that exists when each user has multiple receive antennas, thus providing multiple SINR samples or the opportunity to implement more sophisticated receive structures. Another approach to reduce feedback is the design and implementation of a threshold such that users whose channel quality does not exceed the threshold do not provide feedback. Lastly, due to the greedy scheduling algorithm in the random beamforming scheme, fairness is considered. The results are summarized in more detail in the following sections.

5.1 Reduced Feedback Schemes using Random Beamforming in MIMO Broadcast Channels

Chapter 2 considers feedback reduction by exploiting the multiple receive antennas at each user. These techniques, while reducing feedback, still achieve the desired asymptotic scaling rate of the sum-rate throughput. The contributions of this chapter are now listed:

- The distribution of the maximum order statistic of the SINR over both receive antennas and transmit beamforming vectors at any user is derived. Because the SINR values on a single receive antenna over the various transmit beams are correlated, the distribution of the maximum order statistic leveraged previous results on the maximum of ratios of exponential random variables.

- Under the feedback scheme where only the largest SINR measured at each user over both transmit beams and receive antennas is fed back achieves the desired sum-rate scaling rate of $\log \log n$.

- Utilizing the multiple receive antennas, each user implements an LMMSE receive filter. The distribution of the post-processed SINR values is found.

- Two feedback schemes with LMMSE receive filtering are considered. The first scheme has each user feed back the post-processed SINR for each transmit beam. Under this scheme it is shown that the sum-rate scaling behaves as desired, i.e. $\log \log n$. The second scheme has each user feed back only the maximum of the M post-processed SINR values. The distribution of this random variable could not be analytically found due to the complicated way in which the random variables become correlated. Bounds are found for this unknown distribution, and the bounds are shown to exhibit the $\log \log n$ sum-rate scaling rate, and thus this reduced feedback scheme also exhibits this rate.

- A fixed finite threshold is considered to reduce the total feedback load in the system. It is shown that for any such threshold, asymptotically as the number of users grows to infinity, the difference between the rate of the system with a threshold and without a threshold goes to zero.

5.2 Reducing Feedback in Broadcast Channels via Thresholding

Chapter 3 extends the idea first mentioned at the end of Chapter 2 regarding reducing feedback via thresholds. A threshold should be designed as a function of the number of users in the system. The asymptotic scaling rate of any successful threshold as a function of the number of users is found. By success, it is meant that the sum-rate throughput of the system under the threshold still achieves a scaling rate of $\log\log n$. Additionally, the explicit design of thresholds under different proposed metrics is undertaken. These contributions are now described in detail.

- A sufficient condition is found for the asymptotic scaling rate as a function of the number of users in the system of any possible thresholding function such that the sum-rate throughput of the thresholded system still grows as $\log\log n$. This result is based on the theory of extreme value distributions and their domain of attraction.

- A corresponding necessary condition is found for the asymptotic scaling rate of any thresholding function. The necessary condition leverages results on the asymptotic stability of order statistics.

- A threshold design metric is proposed that constrains the probability of the event that any user does not exceed the threshold for any transmit beamforming vector. When such an event occurs, the transmitter does not have CSI for at least one transmit direction and thus multi-user diversity cannot be fully utilized. Parameterized by the constraint on this probability, as a function of the number of users in the system, an explicit formulation of the optimal threshold is derived.

- A second threshold design metric is to constrain the average number of users providing feed back for each transmit beam. Under this metric, the optimal threshold is found.

- The last threshold metric considers constraining the difference between the sum-rate throughput of the system without a threshold and the sum-rate throughput of the thresholded system. The optimal threshold is found under the constraint on the difference in these rates.

5.3 Greedy Scheduling and Proportional Fair Sharing Under i.i.d. Models

Chapter 4 confronts the issue of fairness in the random beamforming transmit scheme. The multi-user diversity is utilized by selecting the users with the best channel conditions. The greedy nature of the scheduling algorithm implies that there can be long periods of time in which a particular user may not be scheduled. To address this problem, the use of the proportional fair sharing algorithm is considered. This algorithm attempts to schedule users who have not been scheduled in a long period of time yet still balance high instantaneous throughput. The asymptotic system performance under the PFS scheduling algorithm is found. Additionally, motivated by the results in Chapters 2 and 3, the performance of the PFS algorithm is considered when a thresholding scheme is implemented. The contributions of this chapter are detailed below:

- The appropriate state-space to consider the asymptotic average rate of each user is identified. In this state space the convergence point of the greedy scheduling algorithm is found using the law of large numbers. This point serves as a base line to compare the asymptotic performance of the system under PFS scheduling.

- The PFS algorithm is written in a recursive form suitable for the application of the theory of stochastic approximation. Using this theory, the equilibrium point of the PFS scheduler is found and is identical to that of the greedy algorithm. Therefore, asymptotically in time, the PFS scheduling algorithm converges to the performance of the greedy algorithm.

- The rate of convergence of the PFS algorithm to the equilibrium point is found. This rate of convergence leverages the fact that for the i.i.d. channel model the noise term in the stochastic approximation formulation is a martingale difference sequence.

- At any given finite time, the distance of the PFS algorithm from its convergence point is a random variable. The distribution of this random variable is known to be gaussian from the theory of stochastic approximation. Methods to find the covariance matrix of this distance random variable are discussed and explicitly found in several cases.

- A thresholding scheme is proposed to help reduce the total feedback load in the system. It is shown that the implementation of the proposed threshold does not affect the fact that the PFS algorithm converges to a unique equilibrium. This unique equilibrium is identified and shown to differ from the convergence point of the greedy algorithm. The difference is due to the fact that compared to a system without a threshold, there will be some rate loss due to the event where multi-user diversity is lost when no user exceeds the threshold.

Appendix A

Uzgoren's Theorem and a Corollary of Sharif and Hassibi

Let $x_1, \ldots, x_n$ be a sequence of positive random variables with strictly positive density function $f_X(x)$ on the positive real line and cdf $F_X(x)$. The growth function $g_X(x)$ is defined to be

$$g_X(x) = \frac{1 - F_X(x)}{f_X(x)}. \tag{A.1}$$

Also, define the variable u_n to be the unique solution to

$$1 - F_X(u_n) = \frac{1}{n}. \tag{A.2}$$

With these definitions in hand, the theorem due to Uzgoren is restated.

Theorem 18. *(Uzgoren, [Uzg]) Let $x_1, \ldots, x_n$ be a sequence of i.i.d. positive random variables with continuous and strictly positive pdf $f_X(x)$ for $x > 0$ and cdf of $F_X(x)$. Let also $g_X(x)$ be the growth function. Then if $\lim_{x\to\infty} g_X(x) = c > 0$, then*

$$\log\left\{-\log F^n\left(u_n + u g_X\left(u_n\right)\right)\right\} = -u - \frac{u^2 g_X'\left(u_n\right)}{2!} - \cdots - \frac{u^m g_X^{(m)}\left(u_n\right)}{m!} + O\left(\frac{e^{-u+O\left(u^2 g_X'\left(u_n\right)\right)}}{n}\right).$$

The distributions for the maximum SINR per beam and the optimal combining SINR have support on the non-negative real line and have continuous cdfs that do not attain the value of unity for any finite value of the support, which implies the densities are strictly positive on the support. Thus, it must be shown that $\lim_{x\to\infty} g_X(x) = c > 0$. It is shown more generally that having an asymptotic maximum order statistic distribution of type 3 implies this condition. Equation (2.19) gives a condition for the asymptotic distribution of the maximum order statistic to be of type 3, and it can be rewritten as

$$\lim_{x\to\infty}\left(-1-\left[\frac{1-F_X(x)}{f_X(x)}\right]\left[\frac{f''_X(x)}{f_X(x)}\right]\right) = \lim_{x\to\infty}\left(-1-g_X(x)\frac{f''_X(x)}{f_X(x)}\right) = 0. \qquad \text{(A.3)}$$

It was shown that the distributions of interest satisfy this limit. Thus for the limit to go to zero, we need $\lim_{x\to\infty} g_X(x)\frac{f''_X(x)}{f_X(x)} \to -1$. Because $1-F_X(x)$ and $f_X(x)$ are non-negative for all x in the support of the distribution, $\lim_{x\to\infty} g_X(x) \geq 0$. From basic properties of limits, it is also known that

$$\lim_{x\to\infty} g_X(x)\frac{f''_X(x)}{f_X(x)} = \lim_{x\to\infty} g_X(x) \lim_{x\to\infty}\frac{f''_X(x)}{f_X(x)} = -1. \qquad \text{(A.4)}$$

Therefore, since the limit of the product is finite and non-zero, $\lim_{x\to\infty} g_X(x) = c > 0$, and the conditions of the theorem are satisfied.

Next, it must be shown that the following corollary shown in the appendix of [SH05] holds.

Corollary 11. *Let $x_1,\ldots,x_n$ be a sequence of i.i.d. positive random variables with continuous and strictly positive pdf $f_X(x)$ for $x > 0$ and cdf of $F_X(x)$. If $u_n = O(\log n)$, $\lim_{x\to\infty} g_X(x) = c > 0$, and $g_X^{(m)}(u_n) = O(1/u_n^m)$, then*

$$\Pr\{u_n - c\log\log n \leq \max x_i \leq u_n + c\log\log n\} \geq 1 - O\left(\frac{1}{\log n}\right) \qquad \text{(A.5)}$$

All the conditions of this corollary except for the derivative constraint were previously shown to hold. Suppose that $g_X^{(1)}(x) = \Theta(1/x)$, where $f(x) \in \Theta(h(x))$ means f is asymptotically bounded above and below by h, i.e. asymptotically

$|h(x)|k_1 \leq |f(x)| \leq |h(x)|k_2$ for some k_1, k_2. Then by integration $g_X(x) = \Theta(\log x)$, but $\lim_{x\to\infty} \log x \to \infty$, which contradicts $\lim_{x\to\infty} g_X(x) \to c > 0$, therefore $g_X^{(1)}(x) = o(1/x)$. This implies that $g_X^{(m)}(x) = o(1/x^m)$, and because $o(f(x)) \subset O(f(x))$, the derivative constraint of Corollary (11) is also met.

Appendix B

Post-Processing SINR Distribution

The distribution of the SINR from Equation (2.22) turns out to be a special case of the work in [GS98b], [GS98a]. In [GS98b], Gao and Smith let the random variable Z denote the SINR at the output of the optimal combiner and were interested in the link reliability

$$R(z) = \Pr\left[Z \geq z\right]. \tag{B.1}$$

The quantity of interest in this paper is the cdf of Z as to utilize order statistics, which is given by the quantity $F(z) = 1 - R(z)$. Define $E\left[H_i \phi_j \phi_j^* H_i^*\right] = P_i \mathbf{I}$ for the i^{th} user and j^{th} transmit beam, and the noise covariance $\sigma^2 \mathbf{I}$. Assume the j^{th} beam is the intended signal, then let $\gamma = \frac{\rho P_i}{\sigma^2}$ and $\Gamma_k = \frac{P_k}{P_j}$. For the additive Gaussian noise channel where there are $M-1$ interferers for a given beam (i.e. the other $M-1$ beams), Equations (11) - (13) in [GS98b] define the function $R(z)$:

$$R(z) = \exp\left(-\frac{z}{\gamma}\right) \sum_{i=1}^{N} \frac{A_i(z)}{(i-1)!} \left(\frac{z}{\gamma}\right)^{i-1} \tag{B.2}$$

where

$$A_i(z) = \begin{cases} 1, & N \geq M+i \\ \frac{1+\sum_{j=1}^{N-i} C_j z^j}{\prod_{k=1}^{M-1}(1+\Gamma_k z)}, & N < M+i \end{cases}. \tag{B.3}$$

The coefficient C_j in Equation (B.3) is the coefficient of z^j in $\prod_{k=1}^{M-1}(1+\Gamma_k z)$.

This set-up is more general than is needed. All the channels have the same statistics and all the signals have the same power. Therefore $\Gamma_k = 1$ for all k, and the term $\prod_{k=1}^{M-1}(1+\Gamma_k z)$ becomes $\prod_{k=1}^{M-1}(1+z) = (1+z)^{M-1}$, which is independent of any index, and from the binomial theorem the coefficient $C_j = \begin{pmatrix} M-1 \\ j \end{pmatrix}$. Assuming $M \geq N$ implies $A_i(z)$ never equals unity. These simplifications yield

$$A_i(z) = \frac{1 + \sum_{j=1}^{N-i} \begin{pmatrix} M-1 \\ j \end{pmatrix} z^j}{(1+z)^{M-1}} \tag{B.4}$$

and the cdf of the SINR after optimal combining yields the result of the theorem.

Appendix C

Sum-Rate Scaling with LMMSE Reception

The distribution $F_{MMSE}(z)$ is of the form $F(z) = 1 - R(z)$, with $R(z)$ given by Equation (B.2). Combining this with the condition for the limiting distribution to be of type 3, Equation (2.19), the limiting distribution is of type 3 if the following limit is satisfied:

$$\lim_{z \to \infty} \frac{R(z) \frac{d^2}{dz^2} R(z)}{\left(\frac{d}{dz} R(z) \right)^2} \to 1. \tag{C.1}$$

Similar to previous analysis, consider the terms that dominate the limit and show that their limit goes to unity. First, expanding the expression for $R(z)$ yields

$$R(z) = \frac{e^{-z/\rho}}{(1+z)^{M-1}} \sum_{i=1}^{N} \frac{\left(\frac{z}{\rho}\right)^{i-1}}{(i-1)!} + \frac{e^{-z/\rho}}{(1+z)^{M-1}} \sum_{i=1}^{N} \sum_{j=1}^{N-i} \begin{pmatrix} M-1 \\ j \end{pmatrix} \frac{z^{j+i-1}}{\rho^{i-1}(i-1)!}. \tag{C.2}$$

All the terms in $R(z)$ decay to zero as z tends to infinity monotonically, so the terms that decay to zero the slowest are of interest. Looking at the first term in R(z), of all the terms in the sum, the one that decays the slowest is when z^i has the largest exponent, thus the dominating term in the limit is

$$\frac{e^{-z/\rho} z^{N-1}}{(1+z)^{M-1} \rho^{N-1} (N-1)!}. \tag{C.3}$$

Analyzing the second term in $R(z)$, the term that decays the slowest in the summation as z tends to infinity is when z^{j+i-1} has the largest exponent. The largest value the index j in the exponent can take on is $N-i$, and substituting this back into the expression yields z^{N-1} which is independent of the indices i and j. Therefore, the dominating term is given by

$$\frac{e^{-z/\rho}z^{N-1}}{(1+z)^{M-1}}\sum_{i=1}^{N}\left(\begin{array}{c}M-1\\N-i\end{array}\right)\frac{1}{\rho^{i}(i-1)!} \tag{C.4}$$

Combining Equations (C.3) and (C.4) gives the dominating terms of $R(z)$ in the limit

$$\frac{e^{-z/\rho}z^{N-1}}{(1+z)^{M-1}}C \tag{C.5}$$

where the constant C is defined as

$$C=\frac{1}{\rho^{N-1}(N-1)!}+\sum_{i=1}^{N}\left(\begin{array}{c}M-1\\N-i\end{array}\right)\frac{1}{\rho^{i}(i-1)!} \tag{C.6}$$

From Equation (C.1), the limiting terms of the first and second derivatives of $R(z)$ are needed, and luckily the dominating term of $R(z)$ in Equation (C.5) is readily differentiable. Performing some calculus and ignoring terms that decay too fast, the limiting term of $\frac{d}{dz}R(z)$ is

$$-C\frac{e^{-z/\rho}z^{N-1}}{\rho(1+z)^{N-1}}, \tag{C.7}$$

and the limiting term of $\frac{d^2}{dz^2}R(z)$ is

$$C\frac{e^{-z/\rho}z^{N-1}}{\rho^{2}(1+z)^{M-1}}. \tag{C.8}$$

Combining Equations (C.5),(C.7), and (C.8) yields

$$\lim_{z\to\infty}\frac{R(z)\frac{d^2}{dz^2}R(z)}{\left(\frac{d}{dz}R(z)\right)^2}=\frac{C\frac{e^{-z/\rho}z^{N-1}}{(1+z)^{M-1}}\cdot C\frac{e^{-z/\rho}z^{N-1}}{\rho^{2}(1+z)^{M-1}}}{\left(-C\frac{e^{-z/\rho}z^{N-1}}{\rho(1+z)^{M-1}}\right)^2}=1. \tag{C.9}$$

Therefore the limiting distribution of the maximal order statistic for the SINR after optimal combining is of type 3.

If the scaling rate of the unique solution to $1 - F_{MMSE}(u_n) = \frac{1}{n}$ can be found, then the asymptotic scaling rate is known by Equation (15) of [MPP08]

$$\lim_{n\to\infty} \frac{R}{M \log u_n} = 1 \tag{C.10}$$

For sufficiently large n, $1-F_{MMSE}$ is dominated by Equation (C.5). The solution u_n to $1-F_{MMSE}(u_n) = \frac{1}{n}$ is guaranteed to exist since Equation (C.5) is monotonically decreasing and continuous. Following the analysis of Equation (21) in [SH05],

$$\begin{aligned} 1 - F_{MMSE}(u_n) &= \frac{e^{-u_n/\rho} u_n^{N-1}}{(1+u_n)^{M-1}} C = \frac{1}{n} \\ &\Rightarrow \frac{u_n}{\rho} + (M-1)\log(1+u_n) - (N-1)\log(u_n) \\ &= \log n - \log C. \end{aligned}$$

For fixed $N, M \text{and} \rho$ and sufficiently large n, this yields

$$u_n = \rho \log n + \rho(M-N) \log\log n + O(\log\log\log n)$$

since $\log C$ becomes inconsequential, u_n is monotonically increasing, and

$$\lim_{n\to\infty} (\log(1+u_n) - \log u_n) = 0.$$

Thus, the desired scaling rate is achieved.

Bibliography

[Ben96] Michel Benaim. A dynamical system approach to stochastic approximations. *SIAM Journal on Control and Optimization*, 34(2):437–472, 1996.

[BH99a] Michel Benam and Morris W. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1-2):36 – 72, 1999.

[BH99b] Michel Benam and Morris W. Hirsch. Stochastic approximation algorithms with constant step size whose average is cooperative. *The Annals of Applied Probability*, 9(1):pp. 216–241, 1999.

[BK06] A. Bayesteh and A.K. Khandani. How much feedback is required in mimo broadcast channels? In *Information Theory, 2006 IEEE International Symposium on*, pages 1310 –1314, July 2006.

[CB01] G. Casella and R. L. Berger. *Statistical Inference*. New York: Duxbury Press, 2001.

[Che02] Han-Fu Chen. *Stochastic Approximation and Its Applications*. Kluwer Academic Publishers, 2002.

[CS03] G. Caire and S. Shamai. On the achievable throughput of a multi-antenna gaussian broadcast channel. *Information Theory, IEEE Transactions on*, 49(7):1691 – 1706, July 2003.

[Dav70] H. A. David. *Order Statistics*. New York: Wiley, 1970.

[Dur05] R. Durrett. *Probability: Theory and Examples*. Thomson Brooks/Cole, 2005.

[Ede89] A. Edelman. *Eigenvalues and condition numbers of random matrices*. PhD thesis, Massachusetts Institute of Technology, 1989.

[Ete81] N. Etemadi. An elementary proof of the strong law of large numbers. *Probability Theory and Related Fields*, 55:119–122, 1981.

[Fab68] Vaclav Fabian. On asymptotic normality in stochastic approximation. *The Annals of Mathematical Statistics*, 39(4):pp. 1327–1332, 1968.

[FFM03] O. Edfors F. Floren and B. Molin. The effect of feedback quantization on the throughput of a multiuser diversity scheme. In *Proc. IEEE Glob. Telecom. Conf.*, volume 1, pages 497–501, Dec. 2003.

[FP96] Jean-Claude Fort and Gilles Pags. Convergence of stochastic algorithms: From the kushner-clark theorem to the lyapounov functional method. *Advances in Applied Probability*, 28(4):pp. 1072–1094, 1996.

[GA03] D. Gesbert and M. Alouini. Selective multi-user diversity. In *Proc. IEEE Int. Symp. Signal Proc. Info. Theory*, pages 162–165, Dec. 2003.

[GA04] D. Gesbert and M. Alouini. How much feedback is multi-user diversity really worth? In *Proc. IEEE Int. Conf. on Commun.*, volume 1, pages 234–238, June 2004.

[Gal78] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*. New York: Wiley, 1978.

[GS98a] H. Gao and P. J. Smith. Exact sinr calculations for optimum linear combining in wireless systems. *Prob. in the Eng. and Info. Sciences*, 12:261 – 281, 1998.

[GS98b] H. Gao and P. J. Smith. Theoretical reliability of mmse linear diversity combining in rayleigh-fading additive interference channels. *IEEE Trans. Commun.*, 46(5):666 – 672, May 1998.

[IAR09] Y. Isukapalli, R. Annavajjala, and B. Rao. Performance analysis of transmit beamforming for miso systems with imperfect feedback. *Communications, IEEE Transactions on*, 57(1):222 –231, January 2009.

[IMKT05] S.S. Ghassemzadeh Il-Min Kim, Seung-Chul Hong and V. Tarokh. Opportunistic beamforming based on multiple weighting vectors. *IEEE Trans. Wireless Commun.*, 4(6):2683–2687, Nov. 2005.

[IR07] Y. Isukapalli and B.D. Rao. Finite rate feedback for spatially and temporally correlated miso channels in the presence of estimation errors and feedback delay. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 2791 –2795, Nov. 2007.

[IR10] Y. Isukapalli and B.D. Rao. Packet error probability of a transmit beamforming system with imperfect feedback. *Signal Processing, IEEE Transactions on*, 58(4):2298 –2314, April 2010.

[JCK03] K. Kim J. Chung, C.S. Hwang and Y.K. Kim. A random beamforming technique in mimo systems exploiting multiuser diversity. *IEEE J. Selected Areas Commun.*, 21(5):848–855, June 2003.

[JDP83] K. Joag-Dev and F. Proschan. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286 – 295, Mar 1983.

[Jin06] N. Jindal. Mimo broadcast channels with finite-rate feedback. *Information Theory, IEEE Transactions on*, 52(11):5045 –5060, nov. 2006.

[JWZ08] Y.C. Liang J. Wagner and R. Zhang. On the balance of multiuser diversity and spatial multiplexing gain in random beamforming. *IEEE Trans. Wireless Commun.*, 7(7):2512–2525, July 2008.

[KW04] H.J. Kushner and P.A. Whiting. Convergence of proportional-fair sharing algorithms under general conditions. *Wireless Communications, IEEE Transactions on*, 3(4):1250 – 1259, July 2004.

[KY03] Kushner, H. and Yin, G. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

[LHS03] D.J. Love, Jr. Heath, R.W., and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *Information Theory, IEEE Transactions on*, 49(10):2735 – 2747, Oct. 2003.

[LJL+08] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE Jour. on Sel. Areas in Comm.*, 26(8), Oct. 2008.

[MLS08] Yao Ma, A. Leith, and R. Schober. Predictive feedback for transmit beamforming with delayed feedback and channel estimation errors. In *Communications, 2008. ICC '08. IEEE International Conference on*, pages 4678 –4682, May 2008.

[MMA82] M. Obaidullah M. M. Ali. Distribution of linear combination of exponential variates. *Communications in Statistics-Theory and Methods*, 11(13):1453–1463, 1982.

[MPP07] V. Koivunen M. Pun and H. V. Poor. Opportunistic scheduling and beamforming for mimo-sdma downlink systems with linear combining. In *PIMRC*, Sept. 2007.

[MPP08] V. Koivunen M. Pun and H. V. Poor. Sinr analysis of opportunistic mimo-sdma downlink systems with linear combining. In *ICCC*, May 2008.

[MZ07] Yao Ma and Dongbo Zhang. Error rate of transmit beamforming with delayed and limited feedback. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 4071 –4075, Nov. 2007.

[PR08] M. Pugh and B. D. Rao. On the capacity of mimo broadcast channels with reduced feedback by antenna selection. In *Asilomar Conf. on Signals, Systems and Computers, Asilomar, USA*, Oct. 2008.

[PR10a] M. Pugh and B. D. Rao. Reduced feedback schemes using random beamforming in MIMO broadcast channels. *IEEE Trans. Signal Process.*, 58(3):1821–1832, Mar. 2010.

[PR10b] Matthew Pugh and Bhaskar D. Rao. Reducing feedback in broadcast channels via thresholding, 2010.

[Res87] Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, 1987.

[RT73] Sidney I. Resnick and R. J. Tomkins. Almost sure stability of maxima. *Journal of Applied Probability*, 10(2):387–401, 1973.

[Sac58] Jerome Sacks. Asymptotic distribution of stochastic approximation procedures. *The Annals of Mathematical Statistics*, 29(2):pp. 373–405, 1958.

[SH05] M. Sharif and B. Hassibi. On the capacity of MIMO broadcast channel with partial side information. *IEEE Trans. Inf. Theory*, 51(2):506–522, Feb. 2005.

[Smi95] Hal L Smith. Monotone dynamical systems : an introduction to the theory of competitive and cooperative systems. 1995.

[SN07] S. Sanayei and A. Nosratinia. Opportunistic downlink transmission with limited feedback. *IEEE Trans. Info. Th.*, 53(11):4363–4372, November 2007.

[TV05] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication.* Cambridge: Cambridge University Press, 2005.

[Uzg] N. T. Uzgoren. The asymptotic development of the distribution of the extreme values of a sample. In *Studies in Mathematics and Mechanics Presented to Richard von Mises.*

[VHO07] M. Alouini V. Hassel, D. Gesbert and G.E. Oien. A threshold-based channel state feedback algorithm for modern cellular systems. *IEEE Trans. Wireless Comm.*, 6(7):2422–2426, July 2007.

[VJG03] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels. *Information Theory, IEEE Transactions on*, 49(10):2658 – 2668, Oct. 2003.

[VT03] P. Viswanath and D.N.C. Tse. Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality. *Information Theory, IEEE Transactions on*, 49(8):1912 – 1921, Aug. 2003.

[WCH07] J.G. Andrews W. Choi, A. Forenza and R.W. Heath. Opportunistic space-division multiple access with beam selection. *IEEE Trans. Commun.*, 55(12):2371–2380, Dec. 2007.

[WSS06] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the gaussian multiple-input multiple-output broadcast channel. *Information Theory, IEEE Transactions on*, 52(9):3936 –3964, Sept. 2006.

[YAHA05] A. Tewfik Y. Al-Harthi and M. Alouini. Multiuser diversity for wireless communications. In *Proc. IEEE Glob. Telecom. Conf.*, volume 6, Nov.-Dec. 2005.

[YC04] Wei Yu and J.M. Cioffi. Sum capacity of gaussian vector broadcast channels. *Information Theory, IEEE Transactions on*, 50(9):1875 – 1892, Sept. 2004.

[YG06a] Taesang Yoo and A. Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *Selected Areas in Communications, IEEE Journal on*, 24(3):528 – 541, March 2006.

[YG06b] Taesang Yoo and A. Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *Selected Areas in Communications, IEEE Journal on*, 24(3):528 – 541, March 2006.