Asymptotic Techniques for Space and Multi-User Diversity Analysis in Wireless Communications

by

Adarsh B. Narasimhamurthy

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

December 2010

Asymptotic Techniques for Space and Multi-User Diversity Analysis in Wireless Communications

by

Adarsh B. Narasimhamurthy

has been approved

October 2010

Graduate Supervisory Committee:

Cihan Tepedelenlioğlu, Chair
Tolga M. Duman
Andreas S. Spanias
Martin Reisslein
Antonia Papandreou-Suppappola

ACCEPTED BY THE GRADUATE COLLEGE

ABSTRACT

To establish reliable wireless communication links it is critical to devise schemes to mitigate the effects of the fading channel. In this regard, this dissertation analyzes two types of systems: point-to-point, and multiuser systems.

For point-to-point systems with multiple antennas, switch and stay diversity combining offers a substantial complexity reduction for a modest loss in performance as compared to systems that implement selection diversity. For the first time, the design and performance of space-time coded multiple antenna systems that employ switch and stay combining at the receiver is considered. Novel switching algorithms are proposed and upper bounds on the pairwise error probability are derived for different assumptions on channel availability at the receiver. It is proved that full spatial diversity is achieved when the optimal switching threshold is used. Power distribution between training and data codewords is optimized to minimize the loss suffered due to channel estimation error. Further, code design criteria are developed for differential systems. Also, for the special case of two transmit antennas, new codes are designed for the differential scheme. These proposed codes are shown to perform significantly better than existing codes.

For multiuser systems, unlike the models analyzed in literature, multiuser diversity is studied when the number of users in the system is random. The error rate is proved to be a completely monotone function of the number of

users, while the throughput is shown to have a completely monotone derivative. Using this it is shown that randomization of the number of users always leads to deterioration of performance. Further, using Laplace transform ordering of random variables, a method for comparison of system performance for different user distributions is provided. For Poisson users, the error rates of the fixed and random number of users are shown to asymptotically approach each other for large average number of users. In contrast, for a finite average number of users and high SNR, it is found that randomization of the number of users deteriorates performance significantly.

To

my parents Narasimhamurthy and Radha,

and

my brother Arvind

ACKNOWLEDGMENTS

I would like to take this opportunity to convey my heartfelt gratitude and respect to my advisor Professor Cihan Tepedelenlioglu. He has been an ideal advisor, and a true mentor for me throughout my graduate studies. His enthusiasm to learn and pursue new ideas has been a great motivating force. His friendly demeanor and willingness to go above and beyond himself to ensure that the student understands new concepts has made this journey challenging, yet enjoyable.

I am extremely grateful to Professor Tolga Duman and Professor Andreas Spanias for their insightful courses on wireless communications and signal processing. Their ability to elucidate the intricacies of the topic and provide incisive arguments have helped me in understanding the topics in great detail. I would like to thank them for their invaluable input in the development of this dissertation. I would also like to thank Professor Antonia Papandreou-Suppappola and Professor Martin Reisslein for their helpful suggestions and insights during the course of my degree. This milestone would not be possible without the opportunities provided by the Department of Electrical Engineering. I am also grateful to Ms. Darleen Mandt and Ms. Donna Rosenlof for helping me with all the official documents.

I would like to thank all my friends and colleagues in the Signal processing and Communication group, especially Mahesh Banavar, Kautilya Patel and Harish Krishnamoorthie. I thank Mithila Nagendra for being a true friend,

she injected the required optimism and belief in this journey when I needed it the most. I am eternally indebted to my family for their unconditional love and support, and my friends, who are a big part of my life, for being there when I needed them.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Wireless communication is, by far, the fastest developing section of the communications industry. Fueled by digital and RF fabrication improvements and other miniaturization technologies, mobile devices have become smaller, cheaper and more durable. The advantage of implementing mobile systems is that it offers a great deal of flexibility to its users in terms of connectivity, immaterial of their location. But due to channel characteristics and relative motion between the transmitter and receiver, for reasonable performance at high speeds and high data rates, schemes that can counteract the degrading effects of channels have to be devised. Because of this, there still exists a huge difference between ideal wireless communication systems and the currently existing systems.

The current 3G technology (Third Generation) is associated with services that provide the ability to transfer both voice (eg. Telephone call) and non-voice data (eg. downloading data from internet, exchanging email, instant messaging, video telephony etc.,) simultaneously. The cdma2000 standard was the eventual evolution of the 2G CDMA standard to its 3G equivalent. 3G networks are wide area cellular telephone networks which evolved to incorporate high speed internet access, data transfers and video telephony.

Looking at these trends, it is obvious that the later generations require systems working on technologies capable of transferring huge amounts of data at very high speeds. Recent developments in the physical layer have played a key role in the deployment of new technologies with these standards. The de-

mand for higher data rates, lower error probabilities and reduced interference from other users has led to the development of multiple input multiple output (MIMO) antenna systems. With the growing demand for high speed multimedia applications, efficient use of the available scarce spectrum to satisfy this demand is a key factor. The use of multiple antennas leads to a significant increase in the achievable data rates compared to that achieved by single input single output antenna (SISO) systems [1–3], if the path gains between the different antenna pairs fade independently. Note that these benefits are obtained without any bandwidth expansion or increase in the transmit power. But the significant capacity improvement obtained for such systems are realized only for certain cases; realistic channel models lead to variations in the achieved capacity gains [4].

The high capacity and spectral efficiency results obtained for the MIMO system depends on the communication scheme used. Traditionally, multi-antenna systems have been used to increase the diversity order of the system to combat the effects of fading. Diversity techniques exploit the multiple copies of the transmitted data received over independently fading channels. Multiple copies of data are introduced either in the space, time or frequency domain or a combination of them. By introducing redundancy in the time or frequency domain, the spectral efficiency of the system deteriorates. However, in the spatial diversity scheme, redundancy is added in space which implies that there is no loss in spectral efficiency of the system. For example, in a slow fading

Rayleigh fading channel system with one transmit and N_r receive antennas, the transmitted signal passes through N_r independently faded paths. The average bit error probability of this system can be made to decay as $(SNR)^{-N_r}$ at high SNRs instead of $(SNR)^{-1}$ for a single receive antenna case. This is an example of the spatial diversity scheme [5]. While in the diversity technique, we are trying to combat fading, there are other schemes, like spatial multiplexing, which use fading beneficially to achieve higher capacities. If the path gains between the various antenna pairs fade independently and the channel matrix is well conditioned, by transmitting independent data streams in parallel over the spatial channels, the data rates can be increased. This scheme is called "Spatial Diversity" [6], [7]. It is difficult to compare the performance difference between the Diversity and the Spatial Multiplexing scheme [8].

Space time (ST) codes are used in MIMO flat fading systems to implement spatial diversity schemes. In space time codes, data is coded across both the spatial and temporal domains to achieve diversity and multiplexing gains with a certain trade off between the two as described in [8]. The two main subdivisions in ST codes are the space time block codes (STBC) [9, 10] and space time trellis codes (STTC) [11]. Though the implementation of STTCs is more complex than STBCs because of the presence of the Viterbi decoder, STTCs provide both coding and diversity gains leading to better bit error rate (BER) performance compared to STBCs, which can only achieve full diversity gain. Industry standards like the IEEE 802.16e (also known as Mobile WiMAX),

LTE-UMTS (3GPP Long Term Evolution-Universal Mobile Telecommunication Systems; still in the developmental stage) and IEEE 802.11 (Wi-Fi) use multiple antennas.

Even though most of the literature assumes flat fading channel characteristics, many wireless systems experience frequency selective fading channels. The frequency selective behavior of the channel arises when the symbol duration $T_s << \sigma_{T_m}$, where σ_{T_m} is the rms delay spread of the channel. For frequency selective channels, the channel in the frequency domain is not flat, i.e., the channel response is different at different frequencies. A well known spectrally efficient scheme for data transmission over frequency selective channels is the orthogonal frequency division multiplexing (OFDM). As shown in Fig. 1, using OFDM, a digital multicarrier scheme, the channel can be divided into a number of closely spaced orthogonal sub-carriers to transmit the data in parallel. OFDM is robust against multipath fading and inter symbol interference (ISI) as the symbol duration in time increases for the lower rate parallel subcarriers. Channel equalization is also simplified because OFDM, using the fast Fourier transform (FFT), may be viewed as using many slowly-modulated narrowband signals rather than one rapidly-modulated wideband signal. By design, the subcarriers are orthogonal to each other. Thus there is no cross talk between the various sub-channels and the need for guard bands is eliminated. The orthogonality allows for efficient modulator and demodulator implementation, using the FFT algorithm. Due to the recent advances in dig-

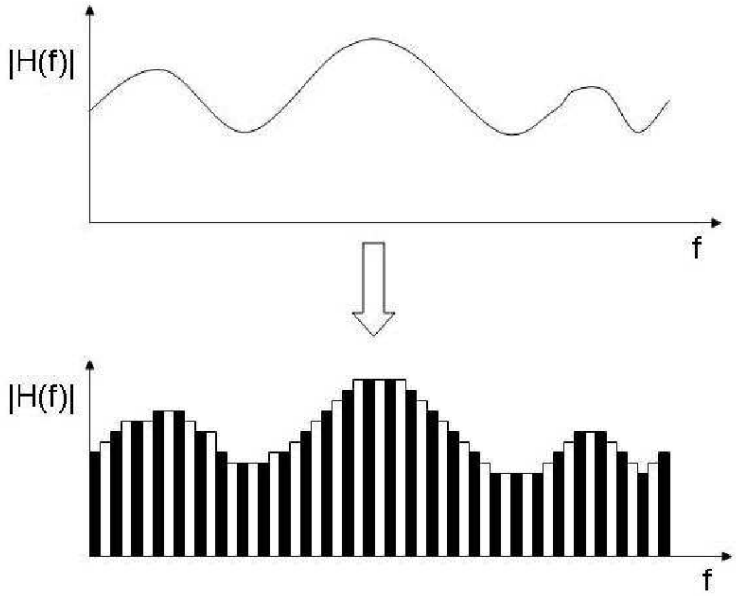

Fig. 1. Frequency Domain representation of a Channel using OFDM

ital signal processing (DSP) hardware design, DSP chips for performing FFT operations have become very fast, small and inexpensive. This is an added incentive for using OFDM. OFDM has already been implemented in IEEE standards for wireless local area networks (WLAN) like the IEEE 802.11a and the IEEE 802.11g standards. OFDM and MIMO are the fundamental building blocks of all future wireless standards, and WiMAX will be the first wide-area implementation to take advantage of these advances.

2. DIVERSITY COMBINING TECHNIQUES

2.1. Need for MIMO

Looking at the trends in the development of wireless communication techniques, it is obvious that there is a requirement for systems based on technologies capable of transferring large amounts of data at very high speeds. The demand for higher data rates, lower error probabilities and reduced interference from other users cannot be satisfied or achieved by having single transmit antennas and thereby has led to the development of multiple input multiple output (MIMO) antenna systems. With the growing demand for high speed multimedia applications, efficient use of the available scarce spectrum to satisfy this demand is a key factor. The use of multiple antennas leads to a significant increase in the achievable data rates compared to that achieved by single input single output antenna (SISO) systems [1–3], only if the path gains between the different antenna pairs fade independently. Note that these benefits are obtained without any bandwidth expansion or increase in the transmit power.

The high capacity and spectral efficiency results obtained for the MIMO system depends on the communication scheme used. In addition to the traditionally used multiple receive antenna systems, implementing diversity schemes to combat the effects of fading, by adding multiple transmit antennas capable of exploiting transmit diversity, the capacity, error performance and the outage characteristics of a system are dramatically improved.

While with diversity techniques the aim is to combat fading at the receiver, at the transmitter by employing space time coding techniques, transmit diversity can be obtained. Beamforming can also be employed at the transmitter to improve the receive SNR of a system. To employ beamforming though, the channel between all transmit and receive antennas need to be known at the transmitter. Using this knowledge the same symbol will be transmitted on each transmit antenna, but with phase and gain values chosen appropriately. For this technique to work it is essential for the transmitter to have accurate channel state information (CSI). Obtaining accurate CSI at the transmitter is very difficult in practice due to feedback delay and estimation errors. Therefore, schemes which do not require the knowledge of the channel at the transmitter are proposed: space-time coding [9–11] and spatial multiplexing [3, 6, 7].

As the name indicates, in space-time coding the information to be transmitted is encoded both spatially and temporally. The encoded sequence is then transmitted over multiple antennas over multiple time slots using the same bandwidth. As a special case of this technique, if independent uncoded streams of symbols are transmitted over different transmit antenna elements, we obtain the spatial multiplexing scheme. The two main subdivisions in ST codes are the space time block codes (STBC) [9, 10] and space time trellis codes (STTC) [11].

For a space time block coded transmit data matrix $\mathbf{X}$ and channel matrix $\mathbf{H}$, which leads to a received matrix $\mathbf{Y}$, the maximum likelihood (ML) decision

$\mathbf{X}$ is based on finding

$$\arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\mathbf{H}\|^2, \tag{2.1}$$

which involves a linear function in the entries of $\mathbf{X}$. For $\mathbf{X}$ belonging to a class of orthogonal space time block codes, the linearity of the likelihood function decouples decisions on the data symbols. This leads to very simple linear receiver implementation with very low complexity. Thus, by implementing orthogonal space-time block codes, while providing both transmit and receive diversity, a very simple receiver can be implemented making this scheme very attractive for implementation. In the other class of space time codes titled Space-Time Trellis Codes, instead of choosing blocks of data sequential coding is performed. The encoding process is performed by using a trellis diagram and at the decoder the Viterbi algorithm needs to be implemented [11]. The Viterbi decoder is also optimal in the ML sense, but its complexity grows exponentially with the number of states in the trellis. Though the implementation of STTCs is more complex than STBCs, STTCs provide both coding and diversity gains leading to a better bit error rate (BER) performance compared to STBCs, which can only achieve full diversity gain. Industry standards like the IEEE 802.16e, LTE-UMTS (3GPP Long Term Evolution-Universal Mobile Telecommunication Systems; still in the developmental stage) and IEEE 802.11 (Wi-Fi) use multiple antennas.

2.2. Diversity Combining

The use of multiple antennas at the receiver to achieve array gain or spatial diversity is a technique that has been known for some time now [5, 12–14]. In this section we concentrate on the techniques that lead to spatial diversity gains. Due to the effects of small scale fading and multi-path propagation, the total signal amplitude received may experience deep fades over time or space. These deep fades lead to system outage. The most popular and efficient technique for combating this phenomenon is to provide multiple, independently faded copies of the same transmitted signal, thereby leading to diversity at the receiver.

The most popular techniques of providing diversity are:

- Space Diversity: Antennas are separated in space
- Angle Diversity: The angle of arrival is different
- Frequency Diversity: Multiple frequencies are used to transmit the same information
- Polarization Diversity: Multiple copies have different field polarization
- Time Diversity: Multiple copies are transmitted over different time slots

To obtain maximum benefit from the above mentioned techniques, the multiple signal copies arriving at the receiver must be uncorrelated (or weakly correlated with correlation coefficient < 0.5). For example, for spatial diversity to ensure maximum gain, rich scattering is required and also the branches

should have sufficient spacing between each other, ($> \lambda/2$), where λ is the wavelength of the received signal, leading to the branches to fade independently.

Depending on the channel characteristics one or more of the above mentioned techniques will be more useful than the others. For example, for a quasi static channel time diversity will not yield any benefits, while for a frequency flat fading frequency diversity is not available. The uncorrelated copies of the transmitted signal can be combined at the receiver by implementing 1) Maximum Ratio Combining (MRC), 2) Equal Gain Combining (EGC), 3) Selection Combining (SC), 4) Threshold Combining (TC), or 5) Hybrid Combining. Selecting n of the available N antennas and implementing MRC on the n antennas is an example of hybrid combining.

2.2.1. Maximal Ratio Combining

In this scheme, as seen in Fig. 2, assuming perfect channel knowledge at the receiver, the received signal on each branch is co-phased and then optimally weighted before summing to maximize the received SNR at the output of the combiner. The optimal weights for each branch are proportional to the branch SNR and hence, the resulting combiner SNR is a sum of the branch SNRs. Thus, for any type of channel fading, MRC is the best combiner in terms of SNR at the receiver, thereby yielding the best performance among all the combining schemes possible. If N indicates the number of receiver antennas,

by analyzing the outage probability it can be shown that a diversity order of N can be achieved for the MRC scheme [15] (i.e., the average bit error probability of this system can be made to decay as $(SNR)^{-N}$ at high SNRs instead of $(SNR)^{-1}$ for a single receive antenna case. This is an example of spatial diversity [5]). Importantly, this result will be achieved only when the branches are uncorrelated. There will be a loss in diversity order if the branches are correlated. The loss in diversity order is proportional to the correlation coefficient between the antennas.

Though this is an optimal combining scheme, the requirement for both the channel phase and gain leads to computation complexity and increase in resource requirement at the receiver, thereby making this a difficult scheme to implement.

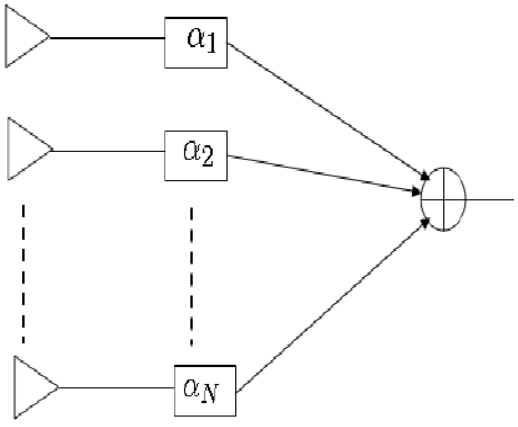

Fig. 2. MRC/EGC Receiver Structure

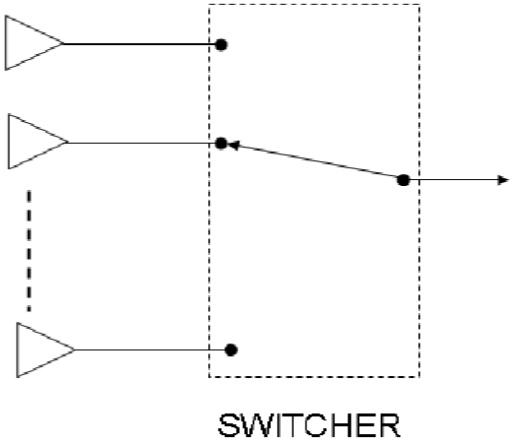

Fig. 3. Antenna Switching Receiver Structure

2.2.2. Equal Gain Combining

The equal gain combining scheme is very similar to the MRC scheme, except for the assumption of equal channel gains at all receiver antennas. Therefore, at each branch the received signal is only co-phased without optimally weighting the branches. When the gains on the branches are equal, EGC performs identical to the MRC, but for unequal gains EGC is suboptimal. The important advantage of EGC over MRC scheme is that only the phase of the channel has to estimated at the receiver thereby leading to lower complexity. An implementation similar to Fig. 2 can be obtained by setting $\alpha_i = \exp(-j\phi)$, where ϕ is the phase rotation induced by the channel.

It can be proved that a diversity order of N can be achieved even for the EGC scheme [15]. Importantly, for both the MRC and EGC schemes, as many RF chains as the number of receive antennas are required to simultaneously receive the signals on all branches. Though antennas are inexpensive and relatively simple devices, RF chains are computationally intensive, and expensive. These are the drawbacks for the above two mentioned schemes when designing mobile terminals with multiple antennas or low complexity devices.

2.2.3. Selection Combining

The key idea in the selection combining scheme is the selection of the branch with the largest channel metric at any given moment of time, for decoding as illustrated in Fig. 3. It is very interesting to note that even though only

one branch is selected at the receiver, because the best branch is selected at any given time, by using ordered statistics, it can be shown that the system achieves full diversity of N. Similar to the schemes described prior to this, channel correlation can significantly degrade the performance and diversity order achievable by implementing SC. It is evident that since only one branch is chosen for decoding, SC is suboptimal compared to MRC because all available resources are not used. SC is also suboptimal to EGC, but the most important reason behind the popularity of the SC is the simplicity in implementation and decrease in resource requirement and complexity at the receiver, while still achieving full diversity.

2.2.4. Threshold Combining

This scheme encompasses a wide range of switching schemes, including the Switch and Stay combining (SSC), switch and examine combining (SEC), post detection combining scheme and so on. In this work we are only interested in the SSC scheme, which can be implemented as shown in Fig. 3. Similar to SC, in the SSC scheme only one receive antenna is used at any given time to decode the received signal. But the key difference is in how the branch is chosen. While in the SC scheme the best branch among all available is chosen, in the SSC scheme a switch is initiated from a particular branch only when the channel metric on that branch drops below a predefined threshold.

Again, it is evident that since the best antenna is not used at any given time the SSC is suboptimal compared to the SC scheme. But the main advantage of this scheme is that the channel metric need not be monitored on all branches to make a decision on the antenna to be used for decoding.

Next, we look at coherent systems which do not have access to perfect channel state information but have to estimate the channel using noise corrupted pilot symbols.

2.3. Channel Estimation

We consider systems employing coherent demodulation of the received data, which necessitates the presence of channel state information (CSI) at the receiver. If CSI is also present at the transmitter end, it can be used beneficially to increase the data throughput by significant margins by implementing the beam forming and water-filling schemes [5]. A majority of work in the literature assumes the presence of perfect CSI at the receiver as this simplifies analysis. But in practice, this assumption can be made only for a very small number of systems, as the channel at the receiver or transmitter has to be estimated. Due to the presence of noise, synchronization errors, approximation errors, time variations in the channel and relative motion between the transmitter and the receiver, the channel estimates are never perfect. The two main techniques of channel estimation are training based and blind estimation methods. In blind estimation methods [16–18], the information symbols

are unknown at the receiver, hence their statistical properties are used to estimate the channel. For the training based scheme, training/pilot symbols already known at the receiver are appropriately inserted into the transmitted signals. The receiver uses these known symbols to estimate the channel. Using training based schemes leads to degradation of the spectral efficiency but these methods simplify receiver design. Also, the channel estimation stage and the data demodulation stage are decoupled in these schemes. Training symbols are also used for carrier synchronization, frequency offset estimation and link recovery from outages. In our work, we shall assume that these are perfectly known and use training only for channel estimation.

Training based channel estimation is a common feature in most of the current communication systems. In GSM (Group Spécial Mobile) [19], a packet contains 148 bits, where 26 training bits are inserted in the middle of each packet, along with 3 more bits at the beginning and the end for training purposes. In the TDMA standard [20], the training symbols are placed at the beginning of each packet. It is important to note that in standards employing CSMA [21,22], training and data symbols are transmitted simultaneously using separate codes. Not only these but standards for broadband LANs [23–26] and wireless broadcast networks [27, 28] depend on training symbols to acquire channel knowledge.

Our work considers the practical receiver using the minimum distance decoder, which is optimal when CSI is perfect, suboptimal otherwise. In general,

the estimated channel is never perfect in any training-based system. Even though sub-optimal, the estimated channel is treated as the perfect channel for detection purposes as this simplifies the receiver complexity and structure. This implies a degradation in the performance of such systems. The training scheme that we consider is optimized in terms of the MSE of channel estimator, the number of training symbols, and error probability.

2.4. Antenna Selection for MIMO-OFDM Systems

As a special case, we next briefly highlight our work in [29], where a diversity combining scheme mentioned before, namely, receive antenna selection, is considered in conjunction with a multiple input multiple output antenna system operating over frequency selective channels by employing Orthogonal Frequency Division Multiplexing (OFDM) technique. Further, the channel is assumed to be unknown at both the transmitter and the receiver. Since, we are interested in analyzing the performance of a coherent system, the channel is estimated at the receiver. We address the problem of imperfect channel knowledge by decoupling the AS problem from channel estimation and proposing a maximum power-based rule for antenna selection and a LMMSE approach for estimating the channel on the selected antenna.

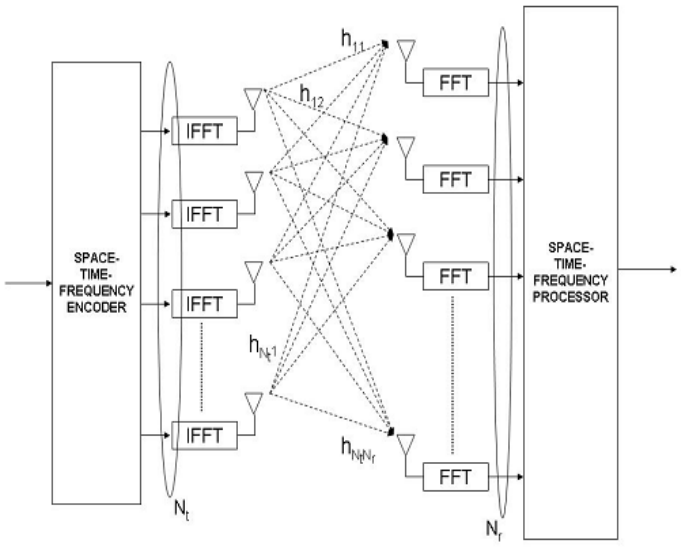

Fig. 4. Block Diagram of the System Model

2.4.1. System Model

A $N_t \times N_r$ MIMO system as illustrated in Fig. 4 is considered. An OFDM system with N_c subcarriers is used to transmit symbols output by the STF encoder. Each STF codeword spans N_x OFDM symbols. The fading channel is assumed to be frequency selective with an order L but time invariant over N_x symbols. The channel is represented by $h_\mu^\nu := \left[h_\mu^\nu, (0), \ldots, h_\mu^\nu (L)\right]^T \epsilon\ \mathbb{C}^{(L+1)\times 1}$, where the elements $h_\mu^\nu (l)$ are i.i.d. $\mathcal{CN}(0,1)$. The received signal by the ν^{th} receive antenna for the q^{th} OFDM symbol, on the p^{th} subcarrier is given by,

$$y_q^\nu(p) = \sqrt{\frac{\rho}{N_t}} \sum_{\mu=1}^{N_t} H_\mu^\nu(p) \cdot x_q^\mu(p) + w_q^\nu(p) \tag{2.2}$$

where $p\,\epsilon\,\{0,\ldots,N_c-1\}$, $\nu\,\epsilon\,\{1,\ldots,N_r\}$, $\mu\,\epsilon\,\{1,\ldots,N_t\}$ and $q\,\epsilon\,\{1,\ldots,N_x\}$. The noise $w_q^\nu(p)$, is i.i.d $\mathcal{CN}(0,1)$. The channel coefficients are defined as

$H_\mu^\nu(p)=(1/\sqrt{L+1})\sum_{l=0}^{L} h_\mu^\nu(l)\cdot e^{-j2\pi lp/N_c}$. The transmitted matrix on the p^{th} subcarrier is represented by $X(p)\,\epsilon\,\mathbb{C}^{N_x\times N_t}$ with $[X(p)]_{q\mu}=x_q^\mu(p)$ and the received matrix $Y(p)\epsilon\ \mathbb{C}^{N_x\times N_r}$ with $[Y(p)]_{q\nu}=y_q^\nu(p)$. The MIMO channel matrix $H(p)\epsilon\ \mathbb{C}^{N_t\times N_r}$ is defined as $\left[H_\mu^\nu(p)\right]=\frac{1}{\sqrt{L+1}}[h_\mu^\nu]^T\omega(p)$ where, we have defined $\omega(p)=\left[1,e^{-j2\pi p/N_c},\ldots,e^{-j2\pi Lp/N_c}\right]^T$. The noise matrix $W(p)\epsilon\ \mathbb{C}^{N_x\times N_r}$ is defined similar to $Y(p)$. For mathematical convenience, we can represent the channel matrix $H\,(p)$ in terms of the time domain channel coefficients $h_\mu^\nu(l)$ as $H(p)=\Omega(p)\cdot h$, where $h=\left[h^1,\ldots,h^{N_r}\right]\epsilon\ \mathbb{C}^{N_t(L+1)\times N_r}$ and $h^\nu=\left[(h_1^\nu)^T,\ldots,(h_{N_t}^\nu)^T\right]^T$, and $\Omega(p)=\frac{1}{\sqrt{L+1}}I_{N_t}\bigotimes\omega^T(p)$. Here $\bigotimes$ denotes the Kronecker product, and I_n is the $n\times n$ identity matrix. So, (2.2) can be equivalently expressed in matrix form as,

$$Y(p)=\sqrt{\frac{\rho}{N_t}}X(p)\cdot\Omega(p)\cdot h+W(p) \tag{2.3}$$

We rely on subcarrier grouping [30] which divides the OFDM symbol with N_c subcarriers into N_g groups each having N_c/N_g subcarriers which are decorrelated within each group. Each group contains $(L+1)$ subcarriers, i.e., $N_c=(L+1)\cdot N_g$.

Defining $Y_g(l):=Y(N_gl+g)$ and $X_g(l)$, $\Omega_g(l)$ and $W_g(l)$ similarly for $g=0,\ldots,N_g-1$ and $l=0,\ldots,L$, the codeword for the g^{th} group is $X_g=\ \mathrm{diag}\,[X_g(0),\ldots,X_g(L)]$, a block diagonal matrix with each block $X_g(l)\epsilon\ \mathbb{C}^{N_x\times N_t}$. The received signal on the g^{th} group $Y_g=\left[Y_g(0)^T,\ldots,Y_g(L)^T\right]^T$, the reduced DFT matrix $\Omega_g=\left[\Omega_g(0)^T,\ldots,\Omega_g(L)^T\right]^T$

and the noise matrix $W_g = \left[W_g(0)^T, \ldots, W_g(L)^T\right]^T$. The input-output relation for each group can now be expressed as,

$$Y_g = \sqrt{\frac{\rho}{N_t}} X_g \cdot \Omega_g \cdot h + W_g \tag{2.4}$$

Let each transmitted codeword on the group, X_g, be chosen from a code book $\mathcal{X}$. Defining $|\mathcal{X}|$ as the cardinality of the code book, the data rate, in bits per subcarrier, for the GSTF code can be expressed as, $R = \{\log_2 |\mathcal{X}|/(L+1)\}$ bits/subcarrier.

The power constraint on the transmitted codeword is $\mathrm{E}\left[\mathrm{tr}\left(X^H X\right)\right] = N_x N_c N_t$, and the truncated Fourier matrix satisfies $\Omega^H \Omega = N_g I_{N_t(L+1)}$. The latter is because we design each group to contain exactly $(L+1)$ subcarriers, making Ω_g unitary for $g = 0, \ldots, N_g - 1$, i.e., $\Omega_g^H \Omega_g = I_{N_t(L+1)}$.

2.4.2. Training and Data Transmission

Training sequences are inserted into transmission frames to estimate the channel. In every symbol there is a training group, $g = \tau$, and $N_g - 1$ data groups, $g \neq \tau$ and the following power constraint is implemented,

$$\sigma_\tau^2 (L+1) N_x N_t + \sigma_D^2 (L+1) N_x \left(N_g - 1\right) N_t = N_c N_x N_t \tag{2.5}$$

where σ_τ^2 and σ_D^2 are the transmit power allocated for each training subcarrier in $g = \tau$ and for each subcarrier in data groups, $g \neq \tau$, respectively. Incorporating the power constraint, the received signal on the g^{th} group can be

expressed as,

$$Y_g = \sqrt{\frac{\rho\sigma_g^2}{N_t}} X_g \cdot \Omega_g \cdot h + W_g \tag{2.6}$$

where $\sigma_g^2 = \sigma_D^2$ or $\sigma_g^2 = \sigma_\tau^2$ depending on whether $g = \tau$ or not. Defining γ as the power allocation ratio for training with respect to the total transmit power, $\gamma = \frac{\sigma_\tau^2 (L+1)}{N_c}$, we can express σ_τ^2 and σ_D^2 in terms of γ as,

$$\sigma_\tau^2 = \gamma N_g \qquad \sigma_D^2 = \frac{(1-\gamma)\, N_g}{N_g - 1} \tag{2.7}$$

2.4.3. Channel Estimation

We use the training group to estimate the channel on the ν^{th} antenna. For the LMMSE estimator that we adopt, it is well known that when X_τ is unitary, i.e., $X_\tau^H X_\tau = I_{N_t(L+1)}$, the mean squared estimation error is minimized [31]. The channel estimate on the ν^{th} antenna can be expressed as,

$$\hat{h}^\nu = \sqrt{\frac{N_t}{\sigma_\tau^2 \rho}} \left(\frac{N_t}{\sigma_\tau^2 \rho} + 1 \right)^{-1} \Omega_\tau^H X_\tau^H Y_\tau^\nu \tag{2.8}$$

and it is Gaussian with $\mathrm{E}\left[\left(\hat{h}^\nu \right) \left(\hat{h}^\nu \right)^H \right] = \sigma_{\hat{h}^\nu}^2 I_{N_t(L+1)}$ where, $\sigma_{\hat{h}^\nu}^2 = \left(\frac{\rho\sigma_\tau^2}{N_t + \rho\sigma_\tau^2} \right)$. The estimation error covariance is given by, $\mathrm{E}\left[(e^\nu) \left(e^\nu \right)^H \right] = \sigma_e^2 I_{N_t(L+1)}$, where $\sigma_e^2 = \left(\frac{\rho\sigma_\tau^2}{N_t} + 1 \right)^{-1}$. The MMSE estimate, $\hat{h}^\nu$, is uncorrelated with the corresponding error vector e^ν due to the orthogonality principle.

Note that the number of parameters to be estimated in $\hat{h}^\nu$ is $N_t N_r (L+1)$. Therefore, at least $N_t N_r (L+1)$ measurements are needed. Looking at the dimensions of Y_τ, $N_x (L+1) \cdot N_r \geq N_t (L+1) \cdot N_r$, which implies that $N_x \geq N_t$.

2.4.4. Antenna Selection

For the sake of exposition we study the selection of a single antenna only, which can be extended to the case of selection of multiple antennas. The selection rule we implement is given by,

$$d = \arg \max_{\nu=1,\ldots,N_r} \|Y_\tau^\nu\|^2. \tag{2.9}$$

This is the maximum signal power selection rule, which also corresponds to selecting the antenna with the best MMSE channel estimate.

Analog power estimators can be implemented using a bandpass filter followed by an envelope detector and a squarer, all of which can be realized using passive circuits.

Note that the analog power estimators cannot differentiate between the various subgroups, so the power has to be estimated based on all the subcarriers in each of the N_x OFDM symbols. Thus, the AS rule for our proposed system is defined as,

$$\arg \max_{\nu} \|Y^\nu\|^2 = \arg \max_{\nu} \left\|\hat{h}^\nu\right\|^2$$

2.4.5. Decoder

The minimum distance decoder, defined as,

$$\hat{X}_g = \arg \min_{X_g} \|Y_g^s - m_g^s\|^2 \tag{2.10}$$

is chosen, where, the estimated channel on the selected antenna $\hat{h}^s$ is used for decoding. Though sub-optimal, we use the minimum distance decoder because of its relative simplicity. It can be seen that the minimum distance decoder in (2.10) is equivalent to the ML decoder when the channel estimates are perfect or when unitary codes are used.

2.4.6. Performance Analysis

The received signal at the selected antenna on the g^{th} group can be expressed as,

$$Y_g^s = \sqrt{\frac{\rho\sigma_D^2}{N_t}} X_g \cdot \Omega_g \cdot \hat{h}^s + \eta \tag{2.11}$$

where we have defined, $\eta := \sqrt{\frac{\rho\sigma_D^2}{N_t}} X_g \cdot \Omega_g \cdot e^s + W_g^s$.

Therefore, the received signal on the g^{th} group and the selected antenna can be expressed in terms of the (known) estimated channel and colored Gaussian noise η with covariance $R_\eta = \frac{\rho\sigma_D^2}{\rho\sigma_\tau^2 + N_t} X_g X_g^H + I_{N_x(L+1)}$. Using this, the Chernoff bound on the PEP is given by

$$\Pr\left(X_g \to X_g'\right) \leq \left[\frac{N_r}{2\left(M!\right)^{N_r-1}(\pi)^M} \frac{1}{\prod_{i=1}^M \lambda_i} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN_r-M}=1}^{M} \frac{l_1! \ldots l_M!}{\lambda_1^{l_1} \ldots \lambda_M^{l_M}}\right] \cdot \left(\frac{\sigma_D^2 \rho}{4\alpha^2 N_t \lambda_m}\right)^{-MN_r} \tag{2.12}$$

where $\lambda_m = (c_1\beta + 1)$, $c_1 = \frac{\rho\sigma_D^2}{\rho\sigma_\tau^2 + N_t}$, $\beta := \max_{X_g \epsilon \mathcal{X}} \left\{\lambda_{max}\{X_g(l) X_g(l)^H\}\right\}$, $M = N_t(L+1)$, and the underlying codes are assumed to achieve full transmit diversity. Thus it can be seen that a diversity advantage of

$MN_r{=}N_t(L+1)N_r$ is achieved by this system, which is the same as that achieved by a full complexity system.

For the system with perfect CSI available at the receiver, the channel at each antenna (h^ν) has the distribution $\mathcal{CN}(0, I_{N_t(L+1)})$ instead of $\hat{h}^\nu{\sim}\mathcal{CN}(0, R_{\hat{h}^\nu})$, implying $\alpha^2{=}1$ in (2.12) when CSI is perfect. Moreover, due to the absence of estimation error, the additive gaussian noise is zero mean Gaussian distributed with covariance $R_\eta{=}I_{N_x(L+1)}$. Thus, $\lambda_{max}(R_\eta){=}1$. As there are no training symbols, the total available power is completely allocated to data symbols, i.e., $\sigma_D^2{=}1$. So, the PEP for the perfect CSI case is obtained by substituting $\alpha^2{=}1$, $\sigma_D^2{=}1$ and $\lambda_m{=}c_1\beta+1{=}1$ in (2.12), i.e.,

$$\Pr\left(X_g \to X_g'\right) \leq \left[\frac{N_r}{2\left(M!\right)^{N_r-1}\pi^M} \cdot \frac{1}{\prod_{i=1}^{M}\lambda_i} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN_r-M}=1}^{M} \frac{l_1! \ldots l_M!}{\lambda_1^{l_1} \ldots \lambda_M^{l_M}}\right] \left(\frac{\rho}{4N_t}\right)^{-MN_r} \tag{2.13}$$

This expression will be compared with (2.12) to quantify the loss in performance due to CEE.

2.4.7. Optimal Power Allocation

The effective loss due to estimation is obtained by taking the ratio of the average PEPs in (2.12) and (2.13) as,

$$\delta{:=}\frac{\rho\gamma\left(1-\gamma\right)N_g^2}{\left(N_g-1\right)\left(\frac{\rho\beta(1-\gamma)N_g}{(N_g-1)}+\rho\gamma N_g+N_t\right)} \tag{2.14}$$

To improve the performance of the system, δ has to be maximized. For a fixed value of N_g, N_t and SNR ρ, we find the optimal value of γ (i.e., γ_{opt}),

that maximizes δ to be,

$$\gamma_{opt}=\frac{\left(\beta+\frac{\mathcal{A}}{\rho}\right)-\sqrt{\left(\beta+\frac{\mathcal{A}}{\rho}\right)\left((N_g-1)+\frac{\mathcal{A}}{\rho}\right)}}{(\beta+1-N_g)} \tag{2.15}$$

where $\mathcal{A}=\frac{N_t(N_g-1)}{N_g}$.

Further analytical and simulation results to corroborate the analytical results derived above can be found in [29]. In the following section, we highlight the novel contributions in this dissertation following.

2.5. Contributions

Here we list the novel contributions in this dissertation:

- Switch and stay diversity combining for multiple input multiple output has been proposed and analyzed for the first time in the literature
- For when perfect channel knowledge is available at the receiver only, a novel switching algorithm is proposed based on instantaneous received SNR
- For a fixed switching threshold that does not depend on the SNR, it is proved that receive spatial diversity cannot be achieved
- Optimal switching threshold is derived to be a logarithmic function of SNR, and when implemented is shown to help achieve full spatial diversity

- When channel is unknown at the coherent receiver, linear minimum mean square error (LMMSE) estimation is proposed and a low complexity transmission frame structure is developed
- Receive power based switching algorithm is proposed, and an optimal switching threshold is derived to be a logarithmic function of SNR again, and full spatial diversity is shown to be achievable
- Power allocation between training and data codewords is optimized to minimize the loss in performance due to channel estimation errors
- Switching rates are calculated and for high SNR it is shown that the switch and stay combining has significantly lower switching rates than antenna selection diversity
- When the channel is unknown at the receiver, differential space time systems are considered
- Lower and upper bounds on the error rate is derived and once again the dependence of the achievable diversity on the switching threshold is illustrated
- Code design criteria is proposed and for the special case of two transmit antenna systems, parametric unitary codes are designed and are shown to outperform existing codes
- Multiuser system implementing multiuser diversity (MUD) when the number of users N is random is analyzed

- Error rate of the MUD system with a fixed number of users N is shown to be a completely monotonic function while the corresponding throughput has a completely monotonic derivative with respect to N
- Error rate is also shown to be a log-concave function with respect to N which implies that the error rate is a non-increasing function of N
- Using the completely monotone property of the error rate and the completely monotone derivative property of the throughput, it is shown that randomness in N always hurts the performance of the MUD system
- Further, using Laplace transform ordering, the distributions of the random variable N can be ordered in terms of the performance of the averaged system
- When N is assumed to be Poisson distributed and the user channel is assumed to be Rayleigh faded, the SNR of the best user chosen from this random set is shown to be Gumbel distributed even for a finite Poisson parameter
- Finally, the error rate, throughput and the ϵ-outage capacity is derived for this system and compared against their corresponding values for a MUD system with a fixed N

2.6. Organization of Dissertation

In the next chapter we present switch and stay combining (SSC) for a multiple input multiple output (MIMO) systems with perfect channel knowledge at the

receiver. Following this, in the Chapter 4, we analyze SSC for MIMO systems for which the channel is unavailable at the receiver but is estimated. We present training, channel estimation and switching schemes and then analyze the performance of such systems. For non-coherent MIMO systems implementing SSC, differential space time coding is proposed in Chapter 5. In this chapter, based on the error rate novel code design criteria are proposed.

In Chapter 6 the second area of focus, multiuser systems, are introduced. Following this in Chapter 7 the problem of multiuser diversity in multiuser systems with a random number of active users is described and analyzed in depth.

3. SWITCH AND STAY FOR MIMO SYSTEMS WITH PERFECT CHANNEL KNOWLEDGE

3.1. Introduction

For communication over wireless fading channels, diversity combining techniques can lead to significant improvement in the system performance [5]. However, the receiver complexity required for implementing diversity combining in multiple input multiple output (MIMO) systems can be significant since as many radio frequency (RF) chains as the number of receive antennas are needed to estimate the channel amplitude and phase on each antenna. Antenna Selection (AS), where only the antenna with the largest channel gain is chosen for decoding, is one possible technique that can reduce MIMO system complexity [32, 33]. AS can be implemented using a single RF chain but the channel gain has to be monitored on all antennas. To further reduce the complexity, while still maintaining the diversity advantage, the switched combining technique for single input multiple output (SIMO) systems has been proposed, where the antenna used for reception is switched only when the gain on the current antenna falls below a predetermined threshold. This has the advantage over AS that the gain needs to be monitored only on the antenna in use [34].

We now summarize the literature on SIMO switch and stay combining. In [34], the performance of a continuous-time model of switch and stay combining for SIMO systems is analyzed over independent Rayleigh fading channels, where a switch occurs if and only if the SNR downward crosses a pre-specified threshold. In [35], a discrete-time switch and stay system model is intro-

duced and its performance for a downward threshold crossing switching rule is analyzed. The performance of a discrete-time system with binary NCFSK modulation over Nakagami-m and Rician fading channels is analyzed for a switching rule based only on the current estimate of the channel on the current antenna in [36,37]. References [38] and [39] analyze the effect of correlated and non identical branches for different fading distributions on the performance of systems implementing the switching algorithm proposed in [36].

Existing literature on switch and stay combining only considers single input multiple output (SIMO) systems for analysis. MIMO systems have been shown to offer tremendous gains in capacity and achievable data rates [1–3] compared to their single input counterparts without any increase in bandwidth or power. However, this gain is obtained at the cost of higher complexity and implementation cost. Switched diversity at the receiver end promises a significant reduction in the system complexity while suffering a modest performance loss compared to antenna selection.

In this work, for the first time, ST coded MIMO systems employing switch and stay combining (SSC) at the receiver, with two antennas, is proposed. A bound on the PEP is derived and the optimal switching threshold that minimizes the bound in derived. Using this bound we illustrate a $(\log \rho)^N/\rho^{2N}$ behavior of the error rate at high SNR. Also, we shown that when a fixed switching threshold is used a maximum diversity order of N only is achievable whereas when the optimal switching threshold is implemented the full spatial

diversity of $2N$ can be achieved. To help with the hardware design we also derive the switching rate achieved by this system.

3.2. System Model

The system under consideration employs N transmit antennas and following the SSC literature, two receive antennas. At the receiver, only one RF chain is assumed to be present. Therefore, at any given time data can be received only on one antenna. The received signal on the i^{th} antenna is given by,

$$\mathbf{y}_i = \sqrt{\frac{\rho}{N}}\mathbf{X}\mathbf{h}_i + \mathbf{w}_i, \tag{3.1}$$

where ρ is the average receive SNR, $i \in \{1, 2\}$ is the receive antenna index, and $\mathbf{y}_i \in \mathbb{C}^{T_c \times 1}$ is the received vector at the i^{th} receive antenna; $\mathbf{X} \in \mathbb{C}^{T_c \mathrm{x} N}$ is the transmitted ST codeword spanning N transmit antennas and T_c time instants, where T_c is the coherence time of the channel, with $T_c \geq N$; $\mathbf{h}_i = [h_{i1}, h_{i2}, \ldots, h_{iN}]^T \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_N)$ contains the channel coefficients between the transmit antennas and the i^{th} receive antenna. Here, the channel h_{in} is assumed to be frequency flat Rayleigh block faded, i.e., the channel fades independently between adjacent blocks but is constant across each block of data spanning T_c samples. The noise vector for the i^{th} receive antenna satisfies $\mathbf{w}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{T_c})$. For the case of switch and stay combining with independent and identically distributed (i.i.d.) branches, due to symmetry, having more than two receive antennas does not provide any improvement

in performance since switching occurs without examining the other antennas' gain [40]. The channel is unknown at the transmitter and a power constraint, $\mathrm{E}\left[\mathrm{tr}(\mathbf{X}^H\mathbf{X})\right]=NT_c$, where $\mathrm{tr}(\cdot)$ is the trace operator, is imposed on the transmitted symbols.

3.2.1. Switching Algorithm and Receive SNR Distribution

We now describe the algorithm used to switch between the two receive antennas. Define $s_{i,t} := (\rho/N)\|\mathbf{h}_i\|^2$ as the SNR at receive antenna i, where $i \in \{1,2\}$, and time $t \in \mathbb{Z}$ and z_t as the SNR at the output of the switch and stay combiner, which is a function of the antenna currently in use. The switching rule adopted in this work is based on the discrete-time algorithm proposed in [36] and is defined as below,

$$z_t{=}s_{1,t} \quad \text{iff} \quad \begin{cases} z_{t-1}{=}s_{1,t-1} \quad \text{and} \quad s_{1,t} \geq \Theta \\ \\ z_{t-1}{=}s_{2,t-1} \quad \text{and} \quad s_{2,t} < \Theta \end{cases}, \tag{3.2}$$

where Θ is the switching threshold common to both the branches. The case when $z_t{=}s_{2,t}$ is the same as (3.2) with $s_{1,t}$ interchanged with $s_{2,t}$.

The cumulative distribution function (CDF), $F_z(u){=}\Pr\{z_t \leq u\}$, of the instantaneous SNR at the output of the combiner is [36],

$$\begin{aligned} F_z(u) &= \Pr\{(z_t{=}s_{1,t} \text{ and } s_{1,t} \leq u) \text{ or } (z_t{=}s_{2,t} \text{ and } s_{2,t} \leq u)\} \quad (3.3) \\ &= \Pr\{\Theta \leq s_{1,t} \leq u\} + \Pr\{s_{2,t} < \Theta\}\Pr\{s_{1,t} \leq u\}, \\ &= \Pr\{\Theta \leq s_{1,t} \leq u\} + \Pr\{s_{1,t} < \Theta\}\Pr\{s_{2,t} \leq u\}, \quad (3.4) \end{aligned}$$

where (3.4) is obtained from (3.3) based on the assumption that the channel is both spatially and temporally (across blocks) i.i.d. Since the channel at the i^{th} receive antenna is $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, the instantaneous received SNR is chi-square distributed with $2N$ degrees of freedom, $s_{i,t} \sim \chi^2(2N)$. Thus, from (3.4), we have $F_z(u)=(F_s(u) - F_s(\Theta))I(u \geq \Theta) + F_s(\Theta)F_s(u)$, where, $F_s(\cdot)$ is the common CDF of $s_{1,t}$ and $s_{2,t}$ and the indicator function $I(u \geq \Theta)=1$ if $u \geq \Theta$, and 0, else. Thus, the CDF of z_t can be expressed as,

$$F_z(u)=\begin{cases} A(\Theta)\left[1-\exp\left(-\frac{Nu}{\rho}\right)\sum_{j=0}^{N-1}\frac{(Nu/\rho)^j}{j!}\right] & u<\Theta \\ 1-[1+A(\Theta)]\left[1-\exp\left(-\frac{Nu}{\rho}\right)\sum_{j=0}^{N-1}\frac{(Nu/\rho)^j}{j!}\right] & u\geq\Theta \end{cases}, \quad (3.5)$$

where $A(\Theta)=1-\exp\left(-N\Theta/\rho\right)\sum_{j=0}^{N-1}[(N\Theta/\rho)^j/j!]$. The probability density function (PDF) can be expressed from (3.5) as,

$$f_z(u)=[I(u \geq \Theta) + A(\Theta)]\left[\left(\frac{N}{\rho}\right)^N \exp\left(-\frac{Nu}{\rho}\right)\left(\frac{u^{N-1}}{(N-1)!}\right)\right]. \quad (3.6)$$

3.3. Performance Analysis

We will consider the maximum likelihood (ML) decoder given by,

$$\widehat{\mathbf{X}}=\arg\min_{\mathbf{X}\in\mathcal{X}}\left\|\mathbf{y}^*-\sqrt{\frac{\rho}{N}}\mathbf{X}\mathbf{h}^*\right\|^2, \quad (3.7)$$

where for simplicity $\mathbf{y}^*:=\mathbf{y}_{i^*}$ and $\mathbf{h}^*:=\mathbf{h}_{i^*}$ are the received signal and the channel vectors at the current antenna i^*, selected according to the switching rule in (3.2), so that $z_t=(\rho/N)\|\boldsymbol{h}^*\|^2$, and $\mathcal{X}$ is the code book with $T_c \times N$ matrix elements.

The pairwise error probability (PEP) is the probability of detecting one codeword while the other is transmitted from a code book containing only a pair of codewords. The PEP of a space-time coded MIMO system implementing SSC at the receiver has not yet been analyzed prior to this work. Defining, $\mathbf{X}_1$ and $\mathbf{X}_2$ as the two chosen codewords from $\mathcal{X}$, the instantaneous PEP conditioned on the channel, can be upper bounded as [32],

$$\mathrm{PEP}(\rho|\mathbf{h}^*) \leq \frac{1}{2}\exp\left(\frac{-\rho}{4N}\|(\mathbf{X}_1-\mathbf{X}_2)\mathbf{h}^*\|^2\right). \tag{3.8}$$

Let the eigenvalue decomposition $(\mathbf{X}_1-\mathbf{X}_2)^H(\mathbf{X}_1-\mathbf{X}_2)=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H$, define $\boldsymbol{\Lambda}$ as the diagonal matrix containing the eigenvalues λ_n, for $n \in \{1,\ldots,N\}$, and $\boldsymbol{V}$ as the corresponding unitary matrix. We assume, without loss of generality, that $\lambda_N \geq \lambda_{N-1} \geq \ldots \geq \lambda_1$ and that $\lambda_1 > 0$ so that the codeword difference matrix, $(\mathbf{X}_1-\mathbf{X}_2)$, is full rank. Further, since $\|\boldsymbol{V}^H\mathbf{h}^*\|^2=\|\mathbf{h}^*\|^2$, the instantaneous PEP in (3.8) can be further upper bounded as,

$$\mathrm{PEP}(\rho|\mathbf{h}^*) \leq \frac{1}{2}\exp\left(-\frac{\rho\lambda_1}{4N}\|\mathbf{h}^*\|^2\right). \tag{3.9}$$

Thus, recalling that $z_t=(\rho/N)\|\boldsymbol{h}^*\|^2$, the average PEP given by $\mathrm{E}[\mathrm{PEP}(\rho|\boldsymbol{h}^*)]$, can be expressed as,

$$\overline{\mathrm{PEP}}(\rho)\leq\frac{1}{2}\int_0^\infty \exp\left(-\frac{\lambda_1}{4}u\right)f_z(u)du, \tag{3.10}$$

where, $f_z(u)$ is given in (3.6). Substituting (3.6) into (3.10) and defining, $C{:=}(\lambda_1/4) + (N/\rho)$, we express the average PEP as,

$$\overline{\mathrm{PEP}}(\rho) \leq \frac{1}{2}\left(\frac{N}{C\rho}\right)^N \left[1 - \exp\left(-\Theta\frac{N}{\rho}\right)\sum_{j=0}^{N-1}\frac{\Theta^j}{j!}\left(\frac{N}{\rho}\right)^j + \exp\left(-C\Theta\right)\sum_{j=0}^{N-1}\frac{(C\Theta)^j}{j!}\right] . \tag{3.11}$$

Since the upper bound on the average PEP expression is not readily expressible in the form $\overline{\mathrm{PEP}}(\rho) \leq (G_c\rho)^{-d}$, with a coding gain of G_c and a diversity order of d, it is not possible to determine the coding gain or the achievable diversity order directly from (3.11). In Section 3.3.2, the achievable diversity order is determined by further analyzing (3.11). Before we embark on the diversity and coding gain analysis, we address the issue of the optimal selection of the switching threshold.

3.3.1. Optimal Switching Threshold

If the switching threshold Θ is either too large (switching too often) or too small (switching rarely) compared to the average SNR ρ, the performance of the system resembles that of a multiple input single output (MISO) system because in either case the information about the channel quality is not fully exploited before switching and thereby the available diversity benefit is not exploited. This motivates optimizing the threshold Θ as a function of ρ. By taking the derivative of the average PEP upper bound (3.11) with respect to Θ and setting it to zero, the optimal switching threshold can be derived for a

given ρ as,

$$\Theta_o = \left(\frac{4N}{\lambda_1}\right) \ln\left(\frac{C\rho}{N}\right), \tag{3.12}$$

where we recall that $C := (\lambda_1/4) + (N/\rho)$. Note that the choice in (3.12) depends on the codeword pair in the PEP through λ_1, and does not necessarily minimize the *bit error probability*. Hence, it should be viewed as a guideline for choosing the optimal threshold with its logarithmic dependence on ρ. Nevertheless, in the following, we show that for this choice of the switching threshold the system yields full spatial diversity.

3.3.2. Diversity Order

By using the optimal switching threshold (3.12), the PEP upper bound in (3.11) can be expressed as,

$$\overline{\mathrm{PEP}}(\rho) \leq \frac{1}{2}\left(\frac{N}{C\rho}\right)^N \left[1 - \left(\frac{C\rho}{N}\right)^{-\frac{\varphi}{\rho}} \left\{ \sum_{j=0}^{N-1} \frac{\left(\frac{\varphi}{\rho}\ln\left(\frac{C\rho}{N}\right)\right)^j}{j!} + \left(\frac{C\rho}{N}\right)^{-N} \sum_{j=0}^{N-1} \frac{\left(\left(N+\frac{\varphi}{\rho}\right)\ln\left(\frac{C\rho}{N}\right)\right)^j}{j!} \right\} \right], \tag{3.13}$$

where, for convenience we define $\varphi := 4N^2/\lambda_1$.

Using (3.13), we establish that $\overline{\mathrm{PEP}}(\rho) = O\left((\ln\rho)^N/\rho^{2N}\right)$ as $\rho \to \infty$ in the following theorem, which will also be helpful in determining the achievable diversity order.

Theorem 1. *The average PEP scales as,*

$$\lim_{\rho\to\infty} \frac{\overline{\mathrm{PEP}}(\rho)}{(\ln\rho)^N/\rho^{2N}} \leq \frac{\varphi^{N+1}}{2N^N N!} \quad . \tag{3.14}$$

Proof. Please see the *Appendix*. □

Using the conventional definition of diversity order [41], i.e., $d = \lim_{\rho\to\infty} -\ln(\overline{\mathrm{PEP}}(\rho))/\ln(\rho)$, we have the following corollary,

Corollary 1. *The diversity order achieved by this system, d=$2N$.*

Proof. Taking the natural log of both sides of (3.14) we have for ρ sufficiently large,

$$\ln(\overline{\mathrm{PEP}}(\rho)) - N\ln(\ln(\rho)) + 2N\ln(\rho) \leq \ln\left(\frac{\varphi^{N+1}}{2N^N N!}\right). \tag{3.15}$$

Dividing both sides by $-\ln\rho$ and taking the limit we have

$$d = \lim_{\rho\to\infty} \frac{-\ln(\overline{\mathrm{PEP}}(\rho))}{\ln\rho} \geq 2N. \tag{3.16}$$

Since the diversity order cannot exceed $2N$ in a $N\times 2$ system, together with (5.28) we have d=$2N$ and the corollary is proved. □

Remark: For the case of full complexity systems and systems employing antenna selection, the average PEP is expressed in the form $\overline{\mathrm{PEP}}(\rho)$=$O(\rho^{-d})$, where d=$2N$. In the SSC case, however, we have $\overline{\mathrm{PEP}}(\rho)$=$O((\ln(\rho))^N\rho^{-d})$. The $(\ln(\rho))^N$ term can be viewed as a penalty in the average PEP for using SSC over antenna selection. Nevertheless the diversity order as defined in (5.28) is still $2N$.

Note that, in achieving full diversity, we used the optimal threshold in (3.12). In fact, following a similar line of reasoning, it can be shown that when Θ is independent of ρ and not optimized, only $d{=}N$ can be achieved, leading to the loss of receive antenna diversity.

3.4. Switching Rate

Though switching between antennas has certain benefits, a high switching rate requires hardware capable of faster synchronization and noise immunity which leads to larger power consumption [42]. This is of particular concern for mobile terminals with limited resources and battery power. The switching rate, $S_r(\rho)$, for a ST coded MIMO system employing SSC at the receiver is given by the expression,

$$\begin{aligned} S_r(\rho) &= \frac{1}{T_c}\left[\Pr\{\mathcal{S}|z_t{=}s_{1,t}\}\Pr\{z_t{=}s_{1,t}\} + \Pr\{\mathcal{S}|z_t{=}s_{2,t}\}\Pr\{z_t{=}s_{2,t}\}\right] && (3.17)\\ &= \frac{2}{T_c}\left[\Pr\{\mathcal{S}|z_t{=}s_{1,t}\}\Pr\{z_t{=}s_{1,t}\}\right] = \frac{1}{T_c}\Pr\{s_{1,t} < \Theta\}, && (3.18)\end{aligned}$$

where $\mathcal{S}$ is the switching event as described in (3.2). The last equality is obtained since the channel on the two branches are assumed to be i.i.d. Here, $\Pr\{s_{1,t} < \Theta\}{=}F_s(\cdot)$ is the common normalized CDF of $s_{i,t}$, for $i \in \{1,2\}$. The instantaneous received SNR $s_{i,t}{=}(\rho/N)\|\mathbf{h}_i\|^2$. Thus the switching rate can be expressed as,

$$S_r(\rho) = \frac{1}{T_c}\left(1 - \exp\left(-\frac{N\Theta_o}{\rho}\right)\sum_{j=0}^{N-1}\frac{(N\Theta_o/\rho)^j}{j!}\right), \qquad (3.19)$$

where T_c is the channel coherence time and Θ_o is the appropriate optimal switching threshold given by (3.12).

Using the properties of the χ^2 CDF, it can be seen that $S_r(\rho)$ behaves like $O((\ln(\rho)/\rho)^N)$. If instead, antenna selection (AS) is used at the receiver with i.i.d. antennas, then the switching rate is $S_r(\rho)=1/(2T_c)$. At high SNR values, it can be seen that this is much higher than the switching rate for SSC.

3.5. Simulations

In this section, simulation results are provided for the ST coded MIMO system employing the switch and stay diversity combining at the receiver. For all simulations, we choose a 2×2 MIMO system implementing the switching algorithm described in (3.2). For the above described system, Alamouti codes, with symbols chosen from a QPSK constellation are used at the transmitter

In Fig. 15, we illustrate the effect of optimization of the switching threshold. The analytical upper bound derived in (4.11) is also plotted against the simulated PEP curve to validate the tightness of the bound. It can be seen that the analytical upper bound is about 3 dB loose for the fixed threshold case. But by optimizing the switching threshold the analytical curve is only about 1 dB loose.

In Fig. 17, we plot the BER performance of the system. The effect of choosing a fixed switching threshold independent of ρ, compared to using the

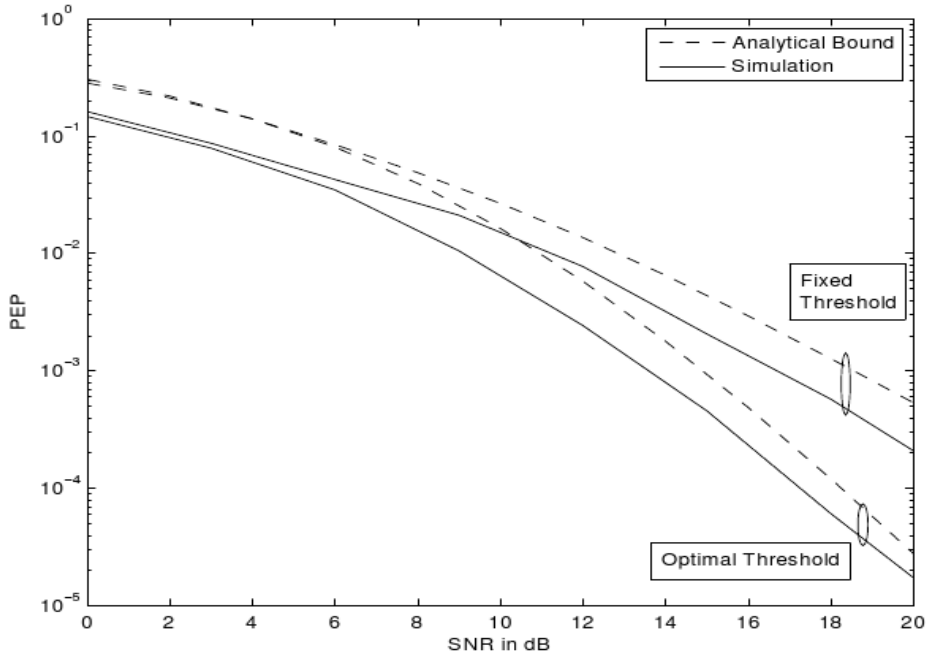

Fig. 5. Pairwise Error Probability: Simulation vs. Analytical

optimal switching threshold, (4.12), on the BER performance, is shown. It can be seen that the performance and diversity order of the system is significantly degraded when a fixed threshold is employed instead of the optimal switching threshold. The SSC scheme with perfect CSI at the receiver is also illustrated in Fig. 17. Due to the extra $(\log(\rho))^N$ penalty term in SSC performance compared to the AS case, the achievable diversity gain is observed at higher SNRs for the SSC case. Both the systems employing SSC and antenna selection achieve a diversity order of $2N$ at high SNR values which cannot be illustrated in the plots.

To verify the validity of the optimal switching threshold in (4.12), we

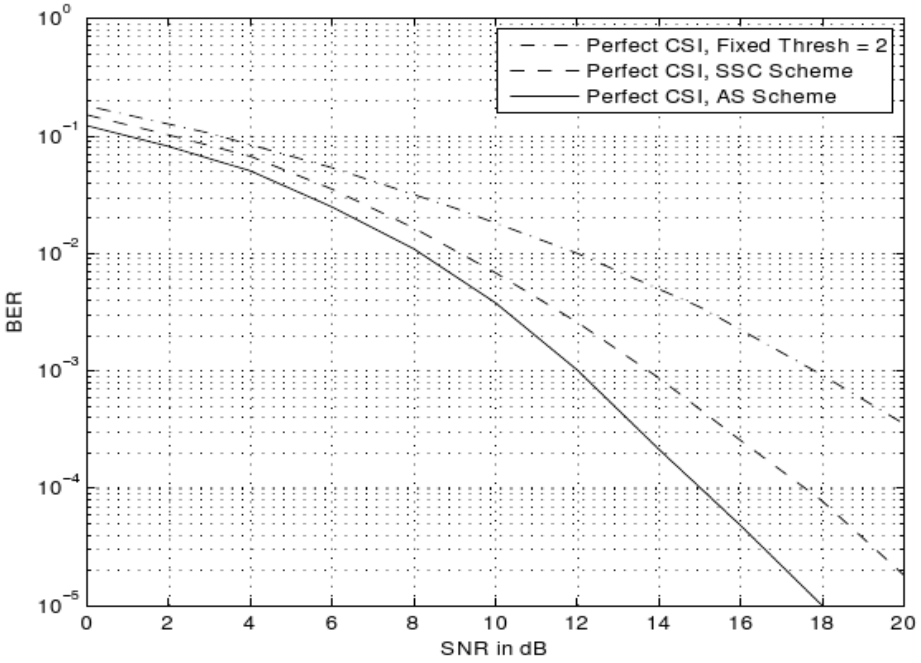

Fig. 6. BER: Alamouti Code, QPSK Symbols, Receive End SSC

simulate the BER of the system in Fig. 7, for fixed SNR values, over a range of switching thresholds to find the optimum value. The BER for average SNR values of $\rho = 2$, 5, 8, 12 and 15 dB is illustrated to show the logarithmic growth of the switching threshold that minimizes the BER. Further, in Table I, we compare the analytically derived optimal switching threshold in (4.12), against the switching threshold values obtained numerically from Fig. 7 for these average SNR values. It can be seen that the analytical and the simulation values are very close to each other which validates our analytical results.

Finally, in Fig. 8, we illustrate the behavior of the switching rate with SNR and number of transmit antennas as discussed in Section 3.4, for both the

Table I. Optimal Switching Threshold Θ_o, Analytical vs Simulation

SNR (dB)	Analytical (dB)	Simulation (dB)
2	1.33	0.5
5	2.33	3.00
8	3.82	5.50
12	6.407	7.50
15	8.84	9.00

perfect and the estimated channel case. We assume a normalized coherence time T_c=1 sec. It can be seen that the optimal switching threshold scales as $O\left((\ln(\rho)/\rho)^N\right)$. When the switching threshold is fixed at a low value compared to the SNR ρ, it can be seen that the switching rates are the lowest, as compared to the case when the switching threshold is fixed at Θ=30 dB, leading to very high switching rates. Implementing the optimal switching threshold, it can be seen that switching rates in-between the above two cases can be achieved along with significant performance improvement as seen from Fig. 15, and 17. Additionally, the switching rate for systems employing antenna selection is also shown and it can be seen that for i.i.d. block fading channels, the SSC scheme has a smaller switching rate.

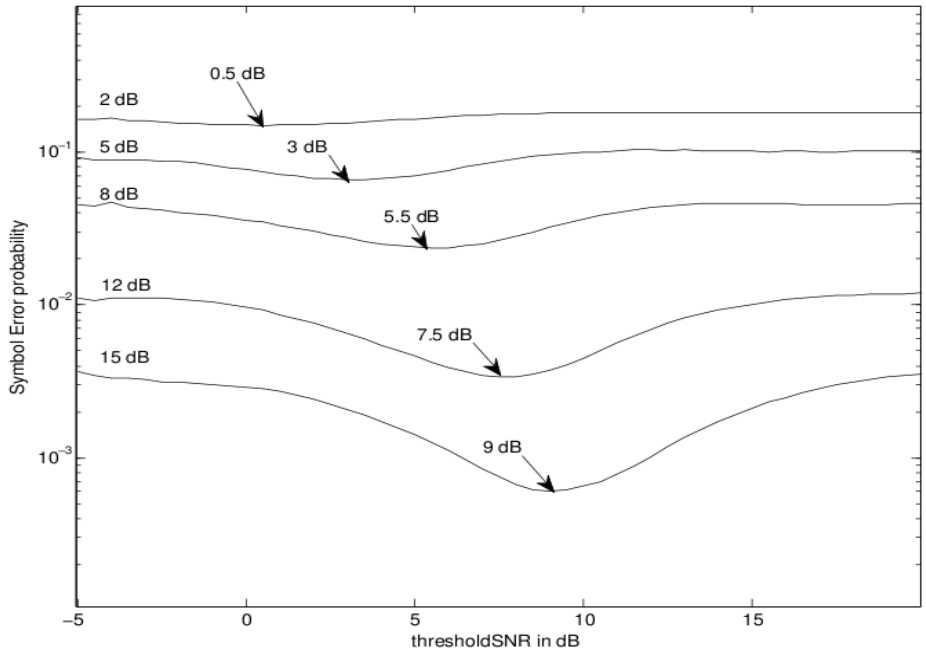

Fig. 7. BER for orthogonal ST codes vs Switching Threshold Θ

3.6. Appendix: Proof of Theorem 4

In this Appendix, we establish (3.14). Using $\sum_{j=0}^{\infty}(x^j/j!)=e^x$, in (3.13), with $x=(\varphi/\rho)\ln(C\rho/N)$, we can write,

$$\overline{\mathrm{PEP}}(\rho) \leq \frac{1}{2}\left(\frac{N}{C\rho}\right)^N \left[1-\left(\frac{C\rho}{N}\right)^{-\frac{\varphi}{\rho}}\left[\exp\left(\frac{\varphi}{\rho}\ln\left(\frac{C\rho}{N}\right)\right)-\sum_{j=N}^{\infty}\frac{\left(\frac{\varphi}{\rho}\ln\left(\frac{C\rho}{N}\right)\right)^j}{j!}\right]\right.$$
$$\left.+\left(\frac{C\rho}{N}\right)^{-(N+\frac{\varphi}{\rho})}\sum_{j=0}^{N-1}\frac{\left(\left(N+\frac{\varphi}{\rho}\right)\ln\left(\frac{C\rho}{N}\right)\right)^j}{j!}\right] \quad (3.20)$$

Substituting (3.20) on the LHS of (3.14), we encounter three limits: (i) $\lim_{\rho\to\infty}(C\rho/N)^{(\varphi/\rho)} = 1$, which can be shown by taking logarithm of both

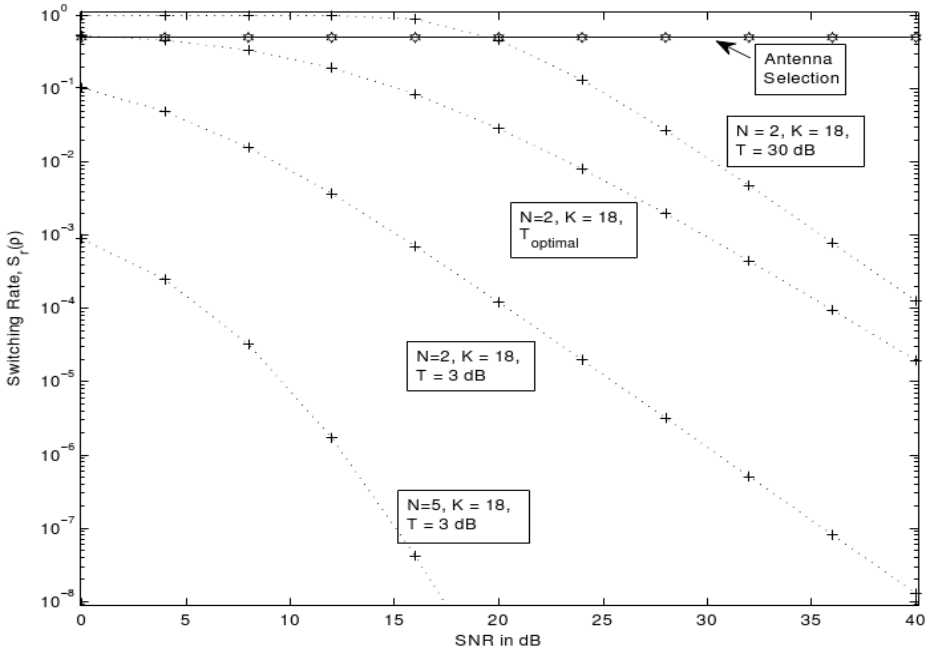

Fig. 8. Switching Rates $S_r(\rho)$: Perfect CSI vs Estimated CSI

sides and using the L'Hospital's rule; (ii)

$$\lim_{\rho\to\infty} \frac{\rho^N}{(\ln \rho)^N} \sum_{j=N}^{\infty} \frac{(\frac{\varphi}{\rho} \ln(\frac{C\rho}{N}))^j}{j!} = \frac{\varphi}{N!}, \tag{3.21}$$

where we interchanged the limit and infinite sum by using the Dominated Convergence Theorem and (iii)

$$\lim_{\rho\to\infty} \frac{\rho^N}{(\ln \rho)^N} \sum_{j=0}^{N-1} \frac{\left(\left(N + \frac{\varphi}{\rho}\right) \ln\left(\frac{C\rho}{N}\right)\right)^j}{(\frac{C\rho}{N})^{(N+\frac{\varphi}{\rho})} j!} = 0 \quad . \tag{3.22}$$

which is straightforward to show since even for $j = N - 1$, the largest term of the summation, the limit is zero. Using these we arrive at

$$\lim_{\rho\to\infty} \frac{\overline{\mathrm{PEP}}(\rho)}{(\ln \rho)^N / \rho^{2N}} \leq \frac{\varphi^{N+1}}{2N^N N!} \quad . \tag{3.23}$$

which is what we needed to show.

4. SWITCH AND STAY FOR MIMO SYSTEMS WITH IMPERFECT CHANNEL KNOWLEDGE

4.1. Introduction

In the previous chapter switch and stay combining for space time coded MIMO systems with perfect CSI was studied. Again, we would like to note that in all previous work on switch and stay combining, the channel is assumed to be perfectly known at the receiver or differential/non-coherent schemes are adopted. It is well known that coherent systems yield better performance compared to differential/non-coherent schemes at the additional cost of channel estimation complexity.

So in this chapter we assume the channel is unknown at both the transmitter and the receiver. To coherently decode the received signal we design a MMSE channel estimation scheme. In this work, for the first time, ST coded MIMO systems employing switch and stay combining (SSC) at the receiver, with two antennas, is proposed. When the channel has to estimated, the training scheme required for MMSE channel estimation is described, based on which, a received-power based switching algorithm, which makes the channel estimation process independent of the switching algorithm, is proposed for the first time, to the best of our knowledge. After upper bounding the PEP, the optimal switching threshold for the imperfect channel case is derived. Further, the power distribution between training and data is optimized to minimize the loss suffered due to channel estimation and is shown to go to zero when the

block length is increased. The switching rate of the system is also calculated as this is important in design of the system hardware.

4.2. System Model

The system under consideration employs N transmit antennas and following the SSC literature, two receive antennas. At the receiver, only one RF chain is assumed to be present. Therefore, at any given time data can be received only on one antenna. The received signal on the i^{th} antenna is given by,

$$\mathbf{y}_i = \sqrt{\frac{\rho}{N}}\mathbf{X}\mathbf{h}_i + \mathbf{w}_i, \tag{4.1}$$

where ρ is the average receive SNR, $i \in \{1, 2\}$ is the receive antenna index, and $\mathbf{y}_i \in \mathbb{C}^{T_c \times 1}$ is the received vector at the i^{th} receive antenna; $\mathbf{X} \in \mathbb{C}^{T_c \mathrm{x} N}$ is the transmitted ST codeword spanning N transmit antennas and T_c time instants, where T_c is the coherence time of the channel, with $T_c \geq N$; $\mathbf{h}_i = [h_{i1}, h_{i2}, \ldots, h_{iN}]^T \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_N)$ contains the channel coefficients between the transmit antennas and the i^{th} receive antenna. Here, the channel h_{in} is assumed to be frequency flat Rayleigh block faded, i.e., the channel fades independently between adjacent blocks but is constant across each block of data spanning T_c samples. The noise vector for the i^{th} receive antenna satisfies $\mathbf{w}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{T_c})$. For the case of switch and stay combining with independent and identically distributed (i.i.d.) branches, due to symmetry, having more than two receive antennas does not provide any improvement

in performance since switching occurs without examining the other antennas' gain [40]. The channel is unknown at the transmitter and a power constraint, $\mathrm{E}\left[\mathrm{tr}(\mathbf{X}^H\mathbf{X})\right] = NT_c$, where $\mathrm{tr}(\cdot)$ is the trace operator, is imposed on the transmitted symbols.

In practice, the channel has to be estimated on each antenna by using pilot symbols. Due to the presence of noise, the estimated channel at the receiver is imperfect, leading to degradation in performance compared to the perfect channel case. Channel estimation for MIMO systems implementing SSC at the receiver and performance analysis of MIMO-SSC systems with imperfect channel has never been considered prior to this work. In this section we address the above stated issues.

When only one RF chain is available at the receiver, the channel has to be estimated sequentially at each antenna which requires multiplexed training [43]. This technique is inefficient because the system cannot estimate the channel on both antennas and make a decision to switch simultaneously. For MIMO systems with only one RF chain at the receiver, we now propose a training scheme to make a switching decision, estimate the channel at the receiver for coherent detection, and characterize the loss in performance suffered due to channel estimation error, and suggest how the loss can be minimized.

To implement the coherent detector at the receiver, the channel on the current antenna is estimated using a MMSE estimator. To aid this estimation process, pilot codewords are inserted at the beginning of each block followed

by K data codewords. The total length of the block is the coherence time T_c, and each block is assumed to have a training duration $T_\tau \geq N$, and K data codewords with duration T_D each.

Assuming that adequate timing synchronization exists at the receiver, the received codewords can be decoded individually. The received signal for the training codeword can be expressed as,

$$\widetilde{\mathbf{y}}_i = \sqrt{\frac{\rho\sigma_\tau^2}{N}}\widetilde{\mathbf{X}}\mathbf{h}_i + \widetilde{\mathbf{w}}_i, \tag{4.2}$$

where, $\widetilde{\mathbf{X}} : T_\tau \times N$ is the training codeword, and for the data codewords as,

$$\mathbf{y}_i^{(k)} = \sqrt{\frac{\rho\sigma_D^2}{N}}\mathbf{X}^{(k)}\mathbf{h}_i + \mathbf{w}_i, \tag{4.3}$$

where σ_τ^2 is the portion of the total transmit power allocated per symbol during the training part of the block and σ_D^2 is the portion allocated per symbol during the data transmission, $\mathbf{X}^{(k)}$ and $\mathbf{y}_i^{(k)}$ are the k^{th} data codeword and received vector, $k \in \{1, \ldots, K\}$, where K is the total number of data codewords in each block. Additionally, the total training and data power satisfies:

$$\sigma_\tau^2 T_\tau N + \sigma_D^2 K T_D N = (T_\tau + K T_D)N. \tag{4.4}$$

4.3. Channel Estimation and Switching with a Single RF Chain

Based on the assumption that the channel on each antenna is constant over the coherence time, it is sufficient to estimate the channel using the training codeword placed at the beginning of the block. The MMSE estimator is chosen

to be implemented here. Choosing the training matrix, $\widetilde{\mathbf{X}}$, to be unitary, $\left(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}}=\boldsymbol{I}_N\right)$, which is optimal over quasi-static channels [43], the channel estimate on the i^{th} antenna is given by,

$$\widehat{\mathbf{h}}_i=\frac{\sqrt{\rho\sigma_\tau^2 N}}{\rho\sigma_\tau^2+N}\widetilde{\mathbf{X}}^H\widetilde{\mathbf{y}}_i. \tag{4.5}$$

Using (4.5) it can be shown that $\widehat{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0},\alpha^2\boldsymbol{I}_N)$, where $\alpha^2{:=}\rho\sigma_\tau^2/(\rho\sigma_\tau^2+N)$. The channel estimation error $\mathbf{e}_i{:=}\mathbf{h}_i-\widehat{\mathbf{h}}_i$ is uncorrelated with $\widehat{\boldsymbol{h}}_i$ due to the orthogonality principle [44]. Further, the error $\mathbf{e}_i \sim \mathcal{CN}\left(\mathbf{0},(N/(N+\rho\sigma_\tau^2))\,\mathbf{I}_N\right)$. We reiterate that to get more equations than unknowns for estimation we assume $T_\tau \geq N$.

4.3.1. Switching without Perfect Channel Knowledge

For receivers with only estimated channel knowledge, the switching algorithm in (3.2) will beimplemented by defining $s_{i,t}{:=}(\rho/N)\|\widehat{\mathbf{h}}_i\|^2$ as the SNR at the i^{th} receive antenna, $i \in \{1,2\}$. But, because only one RF chain is available at the receiver, the system cannot estimate the channel on both antennas and make a decision to switch simultaneously. Further, if the estimated channel norm on the current antenna is larger than the predefined threshold Θ, then the current antenna will be used again for demodulating the present block of data and the channel information on the other antenna is unnecessary. So, if the channel on both receive antennas is estimated every time, the resource requirement at the receiver is higher. Also, the switching necessary for estimating the channel

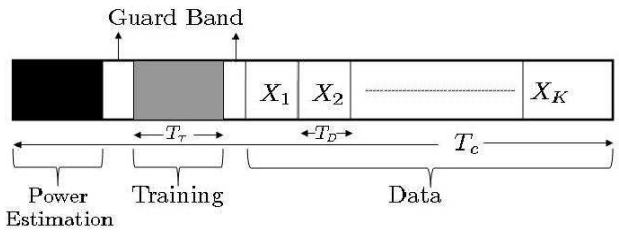

Fig. 9. Transmitted Block

on the other antenna increases the switching overhead. To overcome these problems, the receive power, instead of the estimated channel power can be used for switching. In fact, since $\widetilde{\mathbf{X}}$ in (4.5) is unitary, the received power is a known scalar multiple of the estimated channel power:

$$s_{i,t}=\frac{\rho}{N}\|\widehat{\mathbf{h}}_i\|^2=\left[\frac{1}{\sigma_\tau}\left(\frac{1}{\frac{N}{\rho\sigma_\tau^2}+1}\right)\right]^2\|\widetilde{\mathbf{y}}_i\|^2, \tag{4.6}$$

where $\|\widetilde{\mathbf{y}}_i\|^2$ is the received power at the i^{th} antenna over the training codeword $\widetilde{\mathbf{X}}$. Therefore, for switching, one can use only the received power by computing the right hand side of (4.6) to obtain $s_{i,t}$.

Based on this, each transmitted block can be designed as shown in Fig. 9, where it can be seen that the power on the current antenna is estimated at the beginning of the block, prior to the training period or A/D conversion. If a switch is necessary, it is assumed that it takes place during the guard band following the power estimation phase. Following this, the channel on the chosen antenna is then estimated using the training codeword, which is used for coherently decoding the K data codewords. Note that the received power can be estimated using simple analog circuitry which avoids the need for channel estimation before switching. The decision to switch is made prior to channel

estimation thereby leading to significant reduction in system complexity compared to the alternative of estimating the channel before switching. In what follows, we characterize the performance in the presence of channel estimation error where we assume the guard bands are negligible, for simplicity.

4.4. Performance Analysis for the Imperfect Channel Case

In each received block, since the channel and the noise statistics are assumed to be the same for all $k \in \{1, \ldots, K\}$ data code words, for simplicity, the index k is dropped. Therefore, the received signal on any of the data codewords and the current antenna i^*, after channel estimation can be expressed as:

$$\mathbf{y}^* = \sqrt{\frac{\rho\sigma_D^2}{N}}\mathbf{X}\widehat{\mathbf{h}}^* + \eta, \tag{4.7}$$

where, $\mathbf{y}^* := \mathbf{y}_{i^*}$ and $\widehat{\boldsymbol{h}}^* := \widehat{\mathbf{h}}_{i^*}$ for convenience, $\eta := \sqrt{\rho\sigma_D^2/N}\mathbf{X}\boldsymbol{e} + \boldsymbol{w}$ which can be used to show that $\eta \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\eta)$ with $\mathbf{R}_\eta = a^2\mathbf{X}\mathbf{X}^H + \boldsymbol{I}_{T_D}$, and $a^2 := \rho\sigma_D^2/(N + \rho\sigma_\tau^2)$. Here, $\widehat{\mathbf{h}}^*$ is the estimated channel on the current antenna and $\boldsymbol{e}$ is the estimation error.

Though the switching algorithm uses the received power instead of the estimated channel on the antenna, we analyze the performance for the case when the estimated channel is used, since from (4.6) we know that they yield the same switching rule by appropriately normalizing the threshold. Since the noise η containing the channel estimation error is not white, implementing the ML decoder would be computationally complex. Hence, the minimum

distance decoder in (3.7) is implemented by replacing the true channel $\mathbf{h}^*$ with the estimated channel $\widehat{\mathbf{h}}^*$, which though suboptimal, is simpler. However, the minimum distance decoder is equivalent to the ML decoder when the channel is known perfectly at the receiver, or if the transmitted data codewords are square unitary.

For this imperfect channel case, conditioned on the estimated channel, using (4.7) and the decoder in (3.7), the instantaneous PEP can be upper bounded as,

$$\mathrm{PEP}(\rho|\widehat{\mathbf{h}}^*) \leq \frac{1}{2}\exp\left(\frac{-\rho\sigma_D^2}{4N}\frac{\|(\mathbf{X}_1-\mathbf{X}_2)\widehat{\mathbf{h}}^*\|^2}{\varepsilon^H\mathbf{R}_\eta\varepsilon}\right), \tag{4.8}$$

where ε=$(\mathbf{X}_2-\mathbf{X}_1)\widehat{\mathbf{h}}^*/\|(\mathbf{X}_1-\mathbf{X}_2)\widehat{\mathbf{h}}^*\|$. To get a PEP expression that is similar to the perfect channel case in (3.11), we would like to upper bound $\varepsilon^H\mathbf{R}_\eta\varepsilon$. Now, defining $\lambda_N(\cdot)$ as the N^{th} eigenvalue function, and using $\|\varepsilon\|^2$=1, we have $\varepsilon^H\mathbf{R}_\eta\varepsilon \leq \lambda_N(\mathbf{R}_\eta) \leq 1 + a^2\lambda_N(\mathbf{X}_1{\mathbf{X}_1}^H)$. Since the bound depends on the transmitted codeword, we can further upper bound it by defining β:=$\max_{\mathbf{X}_1\in\mathcal{X}}\lambda_N(\mathbf{X}_1{\mathbf{X}_1}^H)$, so that λ_{max}:=$1+a^2\beta$ to get a bound independent of the transmitted codeword. Substituting into (4.8), we continue to have an upper bound on the PEP:

$$\begin{aligned}\mathrm{PEP}(\rho|\widehat{\mathbf{h}}^*) &\leq \frac{1}{2}\exp\left(\frac{-\rho\sigma_D^2}{4N}\frac{\|(\mathbf{X}_1-\mathbf{X}_2)\widehat{\mathbf{h}}^*\|^2}{\lambda_{max}}\right) &(4.9)\\ &\leq \frac{1}{2}\exp\left(\frac{-\rho\sigma_D^2\lambda_1\|\widehat{\mathbf{h}}^*\|^2}{4N\lambda_{max}}\right), &(4.10)\end{aligned}$$

where λ_1 is as defined in (3.9), and (4.10) is obtained with an argument identical to that of (3.9).

Since the estimated channel at the i^{th} receive antenna is $\widehat{\mathbf{h}}_i \sim \mathcal{CN}(\mathbf{0}, \alpha^2\mathbf{I}_N)$, the normalized instantaneous received SNR is still $\chi^2(2N)$. Thus, the density function of the SNR at the output of the combiner for this case can be expressed as in (3.6), but with the substitution $\rho \rightarrow \rho\alpha^2$. Using similar upper bounding techniques as in Section 3.3 for the perfect channel knowledge case, the average PEP for imperfect channel knowledge case can be expressed as,

$$\overline{\mathrm{PEP}}(\rho) \leq \frac{1}{2}\left(\frac{N}{C_e\rho\alpha^2}\right)^N \left[1 - \exp\left(-\Theta\frac{N}{\rho\alpha^2}\right)\sum_{j=0}^{N-1}\frac{\Theta^j}{j!}\left(\frac{N}{\rho\alpha^2}\right)^j + \exp\left(-C_e\Theta\right)\sum_{j=0}^{N-1}\frac{(C_e\Theta)^j}{j!}\right], \tag{4.11}$$

where $C_e := (L_r + N/\rho\alpha^2)$ and $L_r := \sigma_D^2\lambda_1/(4\lambda_{max})$. Optimizing (4.11) with respect to Θ, we have,

$$\Theta_o = \left(\frac{N}{L_r}\right)\ln\left(\frac{C_e\rho\alpha^2}{N}\right), \tag{4.12}$$

which is the threshold that $(\rho/N)\|\widehat{\mathbf{h}}^*\|^2$ should be compared with for switching. If instead, we threshold the received signal $\|\mathbf{y}_{i^*}\|^2$ for switching, using (4.6), we see that (4.12) should be scaled with $\left[\sigma_\tau\left(N/(\rho\sigma_\tau^2)+1\right)\right]^2$. Like the perfect channel case, (4.12) should be viewed as a guideline as to how to choose the threshold, rather than an exact value for optimum BER performance.

For the system with perfect channel knowledge, the channel at the receiver $\boldsymbol{h}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_N)$, making α^2=1 and σ_D^2=1 since all the available power is allocated to the data payload. Additionally, since the noise is $\boldsymbol{w}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{T_D})$,

λ_{max}=1. Thus, not surprisingly, the average PEP upper bound for the perfect channel case can be obtained from (4.11).

Since, the PEP in the imperfect channel case in (4.11) has a similar form as the perfect case in (3.11) with the substitutions $\rho \to \rho\alpha^2$, $\lambda_1/4 \to L_r$, after substituting the optimal switching threshold, the following follows from Theorem 4,

Theorem 2. *The average PEP, with the optimal switching threshold (4.12), can be upper bounded as,*

$$\lim_{\rho\to\infty} \frac{\overline{\mathrm{PEP}}(\rho)}{(\ln \rho)^N/\rho^{2N}} \leq \frac{\varphi_e^{N+1}}{2N^N N!}. \tag{4.13}$$

where $\varphi_e = \lim_{\rho\to\infty} N^2/(\alpha^2 L_r)$.

Proof. Identical to the proof of Theorem 4 in the *Appendix*, by making the following substitutions: $\rho \to \rho\alpha^2$; $\varphi \to \varphi_e$; and $C \to C_e$. □

Using similar arguments as described in Corollary 1, it is straightforward to show that a diversity order of $2N$ is achieved when the optimal switching threshold is used, which is the full spatial diversity.

From Theorems 4 and 5, we observe that the average PEP for both the perfect channel and the estimated channel case asymptotically behave as $O\left((\ln(\rho))^N/\rho^{2N}\right)$. So, it would be insightful to characterize the loss in performance suffered due to channel estimation error. Taking the ratio of (4.13) and (3.14), we see that the loss in performance at high SNR is proportional to

$(\varphi_e/\varphi)^{N+1} = \lim_{\rho\to\infty}((a^2\beta+1)/(\sigma_D^2\alpha^2))^{N+1}$. This is an approximation to the loss in performance because both (4.13) and (3.14) are upper bounds. This loss can be minimized by optimizing the power distribution between training and data code words, i.e., σ_τ^2 and σ_D^2, which we pursue without resorting to approximations in section 4.4.1, not only for high SNR, but also for any average SNR.

4.4.1. Optimal Power Allocation

We now calculate the loss induced due to imperfect channel estimates as this would indicate when differential systems, with less complexity, yield better performance than coherent systems with imperfect channel for the SSC system under consideration.

Defining $\gamma{:=}(\sigma_\tau^2 T_\tau/(T_\tau + KT_D))$ as the power allocation ratio between the training power and the total transmitted power, we can express σ_τ^2 and σ_D^2 as,

$$\sigma_\tau^2{=}\frac{\gamma(T_\tau + KT_D)}{T_\tau} \qquad \sigma_D^2{=}\frac{(T_\tau + KT_D)(1-\gamma)}{KT_D}. \tag{4.14}$$

Recalling the power constraint in (4.4), if $\sigma_\tau^2 \gg \sigma_D^2$, the quality of the estimates improve but there is a degradation in the performance due to a decrease in the SNR for the data symbols. Conversely, if $\sigma_D^2 \gg \sigma_\tau^2$, again there is a loss in performance owing to the fact that the channel estimates are degraded. So, the power allocated to the training and data codewords, σ_τ^2 and σ_D^2, has

to be optimized to reduce the degradation suffered due to imperfect channel estimation.

Recalling from the average PEP upper bound derivation that $\varepsilon^H \mathbf{R}_\eta \varepsilon \leq \lambda_{\max}=a^2\beta+1$, we can express the instantaneous PEP for the imperfect channel case as,

$$\begin{aligned}\mathrm{PEP}(\rho|\widehat{\mathbf{h}}^*)=Q\left(\sqrt{\frac{\rho\sigma_D^2}{2N}\frac{\|(\mathbf{X}_1-\mathbf{X}_2)\widehat{\mathbf{h}}^*\|^2}{\varepsilon^H\mathbf{R}_\eta\varepsilon}}\right)&\leq Q\left(\sqrt{\frac{\rho\sigma_D^2}{2N}\frac{\|(\mathbf{X}_1-\mathbf{X}_2)\widehat{\mathbf{h}}^*\|^2}{(a^2\beta+1)}}\right)\\&=Q\left(\sqrt{\frac{\rho\delta}{2N}\|(\mathbf{X}_1-\mathbf{X}_2)\overline{\mathbf{h}}^*\|^2}\right),\end{aligned} \tag{4.15}$$

where the normalized channel of the selected antenna in the imperfect channel case $\overline{\mathbf{h}}^*:=(1/\alpha)\widehat{\mathbf{h}}^*$, has the same distribution with $\mathbf{h}^*$, which is its counterpart for the perfect channel case, and

$$\delta:=\frac{\sigma_D^2\alpha^2}{a^2\beta+1}. \tag{4.16}$$

By substituting for σ_D^2, α^2 and a^2, we can also express δ in terms of the power allocation ratio γ as,

$$\delta=\frac{\rho\gamma(1-\gamma)(T_\tau+KT_D)^2}{\rho((1-\gamma)T_\tau\beta+\gamma KT_D)(T_\tau+KT_D)+N}. \tag{4.17}$$

When the channel is perfectly known at the receiver, the noise covariance $\mathbf{R}_\eta=\mathbf{I}_{T_D}$, and $\sigma_D^2=1$, yielding the instantaneous PEP as,

$$\mathrm{PEP}(\rho|\mathbf{h}^*)=Q\left(\sqrt{\frac{\rho}{2N}\|(\mathbf{X}_1-\mathbf{X}_2)\mathbf{h}^*\|^2}\right). \tag{4.18}$$

Since in (4.15) and (4.18), $\overline{\mathbf{h}}^*$ and $\mathbf{h}^*$ are identically distributed, comparing (4.15) with (4.18), the effective loss in SNR due to channel estimation error is lower bounded by δ, which is a lower bound on the actual loss in SNR due to the fact that (4.15) is an upper bound. Note that since there is a loss due to channel estimation, the loss in average SNR is always less than one, and therefore the lower bound to it satisfies $\delta \leq 1$. To minimize the loss suffered due to estimation, δ has to be maximized (i.e., made as close as possible to one) with respect to γ (equivalently, σ_D^2 and σ_τ^2). Further, it can be seen that there is an equivalence between the high SNR quantity $\varphi_e/\varphi = \lim_{\rho\to\infty}(a^2\beta + 1)/(\sigma_D^2\alpha^2)$ of the previous section and $\lim_{\rho\to\infty}\delta^{-1}$. Therefore, maximizing δ at high SNR with respect to the power allocation ratio γ implies minimizing the loss in performance mentioned right before Section 4.4.1. Different than that discussion, we will consider maximizing δ for any average SNR ρ. For the special case when square unitary codes are used for transmitting data, $\beta = 1$ thereby making (4.15) an equality, and δ is no longer a bound on the loss, but is an equality.

The optimal value of γ that maximizes δ can be found by setting the derivative of δ in (4.17) with respect to γ to zero, as

$$\gamma_{opt} = \frac{-\zeta + \sqrt{\zeta(\zeta - \beta + \frac{KT_D}{T_\tau})}}{\frac{KT_D}{T_\tau} - \beta}, \tag{4.19}$$

where $\zeta := \beta + NKT_D/(\rho(T_\tau + KT_D))$.

Using (4.19) we can show the following for any ρ:

Theorem 3. *As the number of data codewords increases ($K \to \infty$) the pairwise error probability for the imperfect channel case approaches that of the perfect channel case, for any ρ.*

Proof. To prove the theorem, we will need to unpack the dependence of each term in (4.16) on K. Since σ_D^2 in (4.14) depends on γ, we begin by using (4.19) to see that $\gamma_{opt} = O(1/\sqrt{K})$. Using this in (4.14), it can be seen that $\sigma_D^2 \to 1$ as $K \to \infty$. Also, from (4.14) $\sigma_\tau^2 = O(\sqrt{K})$. Recalling that $\alpha^2 = \rho\sigma_\tau^2/(\rho\sigma_\tau^2 + N)$, and that $\sigma_\tau^2 = O(\sqrt{K})$, it is clear that $\alpha^2 \to 1$ as $K \to \infty$. Using the definition of $a^2 = \rho\sigma_D^2/(\rho\alpha_\tau^2 + N)$, it is clear that $a^2 \to 0$ as $K \to \infty$, since $\sigma_\tau^2 \to \infty$ and $\sigma_D^2 \to 1$ as $K \to \infty$. Putting it all together we have $\delta \to 1$ as $K \to \infty$ for all ρ. Since δ is a lower bound on the actual loss, when $\delta \to 1$ so does the loss in average SNR. □

Theorem 3 indicates that when the coherence time is large and one can afford to increase the number of data blocks K, the effect of channel estimation can be made arbitrarily small when the power allocation is optimized. This further underlines the need for optimization of the power allocation.

4.5. Switching Rate

Though switching between antennas has certain benefits, a high switching rate requires hardware capable of faster synchronization and noise immunity which leads to larger power consumption [42]. This is of particular concern for mobile

terminals with limited resources and battery power. The switching rate, $S_r(\rho)$, for a ST coded MIMO system employing SSC at the receiver is given by the expression,

$$S_r(\rho)=\frac{1}{T_c}\Pr\{s_{1,t} < \Theta\} = S_r(\rho)=\frac{1}{T_c}\left(1-\exp\left(-\frac{N\Theta_o}{\alpha^2\rho}\right)\sum_{j=0}^{N-1}\frac{(N\Theta_o/\alpha^2\rho)^j}{j!}\right) \quad (4.20)$$

where T_c is the channel coherence time and Θ_o is given by (4.12). Recall from (4.6), using $s_{i,t} = (\rho/N)\|\widehat{\mathbf{h}}_i\|^2$ with Θ_o defined in (4.12) would result in the same switching rate as using $s_{i,t} = \|\mathbf{y}_i\|^2$ with Θ_o in (4.12) scaled by $\left[\sigma_\tau\left(N/(\rho\sigma_\tau^2)+1\right)\right]^2$. Therefore, the switching rate for the received power switching rule is captured by the estimated channel switching.

Using the properties of the χ^2 CDF, it can be seen that $S_r(\rho)$ behaves like $O((\ln(\rho)/\rho)^N)$. If instead, antenna selection (AS) is used at the receiver with i.i.d. antennas, then the switching rate is $S_r(\rho)=1/(2T_c)$. At high SNR values, it can be seen that this is much higher than the switching rate for SSC. If the fading is assumed to be correlated across time on each antenna but i.i.d. across the two antennas, it is straightforward to show that the switching rates for both the case of SSC and AS are scaled down by an identical factor proportional to the increase in the coherence time T_c. But even for this case, for the channel model described, the SSC scheme yields a lower switching rate than the AS scheme. This shows that the switching overhead needed to implement the SSC algorithm is significantly lower compared to the AS algorithm.

4.6. Correlated Fading

For the switching model assumed in our work, assuming the fading to be correlated across blocks will not change the average probability of error performance. However the outage duration changes thereby leading to changes in the switching rate $S_r(\rho)$. Further, for the channel model assumed, even for the case of time correlated block fading channels, the switching rates of both the SSC and the antenna selection scheme are scaled down identically proportional to the coherence time. But the SSC scheme still has a lower switching rate than antenna selection. We explain these issues in more detail below.

Consider the CDF expression (3) in the manuscript which can be expressed as,

$$F_{z_t}(u) = [\Pr\{z_t = s_{1,t} \text{ and } s_{1,t} \leq u\} + \Pr\{z_t = s_{2,t} \text{ and } s_{2,t} \leq u\}] \tag{4.21}$$

$$= 2\,[\Pr\{z_t = s_{1,t} \text{ and } s_{1,t} \leq u\}] \tag{4.22}$$

$$= 2\,[\Pr\{(z_{t-1} = s_{1,t-1} \text{ and } s_{1,t} \geq \Theta) \text{ or } (z_{t-1} = s_{2,t-1} \text{ and } s_{2,t} < \Theta) \text{ and } s_{1,t} \leq u\}] \tag{4.23}$$

$$= 2\,[\Pr\{z_{t-1} = s_{1,t-1} \text{ and } s_{1,t} \geq \Theta \text{ and } s_{1,t} \leq u\}]$$
$$+\Pr\,[\{z_{t-1} = s_{2,t-1} \text{ and } s_{2,t} < \Theta \text{ and } s_{1,t} \leq u\}] \tag{4.24}$$

Here, $z_{t-1} = s_{1,t-1}$ or $z_{t-1} = s_{2,t-1}$ indicates which antenna was used for the $(t-1)$ block. Though the fading is assumed to be correlated, conditioning on the antenna used for the $t-1$ block does not give any information about either $s_{1,t}$ or $s_{2,t}$. Thus, the result in eqn. (4) in the manuscript still holds.

This implies that the PDF of the SNR at the output of the combiner is the same as that for the i.i.d. case. Therefore, there is no change in performance when the fading is correlated compared to the i.i.d. case.

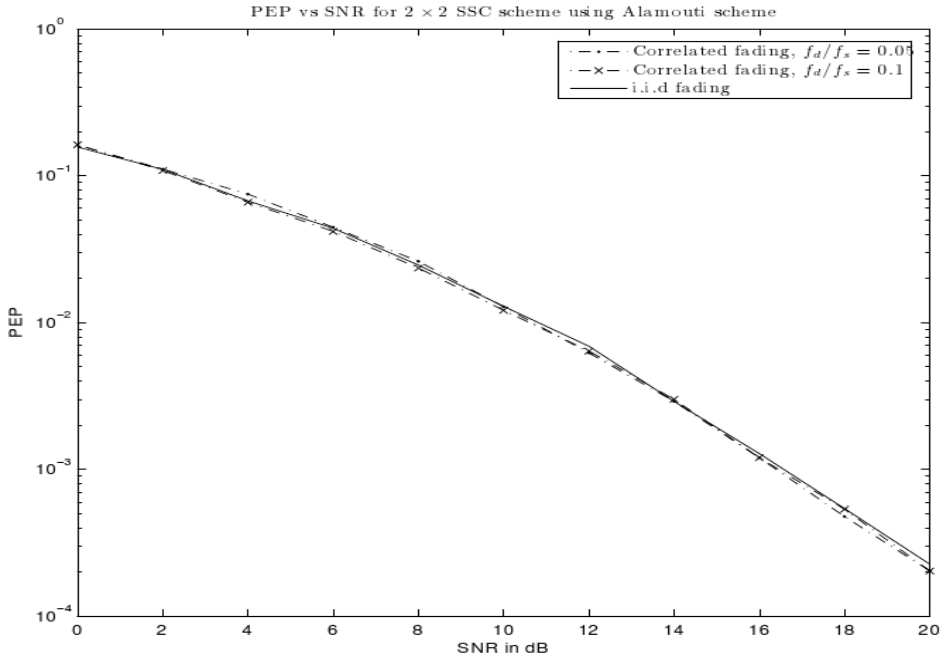

Fig. 10. PEP, Correlated vs. i.i.d. temporal fading

Assuming the fading to be correlated also assumes a larger coherence time T_c. Though larger T_c implies longer fade duration for each antenna, it also implies that the average rate of the SNR falling below the switching threshold is also smaller. Therefore, assuming time correlation implies longer outage duration. Therefore errors instead of occurring independently are now bursty in nature. Fig. 10 shows the PEP plots for the i.i.d vs. correlated fading case. Jakes' channel model is used for generating the correlated channels and for this

a sampling frequency $f_s = 1000$ samples/sec is used and Doppler shifts of 50 Hz and 100 Hz are considered. It can be seen from Fig. 2 that the performance of the correlated block fading is identical to the i.i.d. block fading case. To conclude, assuming the fading to be correlated across the blocks does not affect the average performance of the system, but changes the distribution of errors as explained previously.

Switching Rate for the SSC scheme considered is given by the expression:

$$S_r(\rho) = \frac{1}{T_c}\Pr\left[s_{1,t} < \Theta\right] \tag{4.25}$$

Since the two branches at the receiver are assumed to be i.i.d. block fading, even when the channel on each antenna is correlated across time (i.e., across blocks), the above expression holds. For larger coherence time (T_c), the switching rate decreases for the SSC schemes. In comparison, the switching rate for an antenna selection scheme with two receive antennas can be expressed as,

$$S_r = \frac{1}{T_c}\left[\Pr\left\{s_{1,t} < s_{2,t} | z_t = s_{1,t}\right\} + \Pr\left\{s_{2,t} < s_{1,t} | z_t = s_{2,t}\right\}\right] \tag{4.26}$$

Since the two antennas are spatially i.i.d,

$$S_r = \frac{1}{T_c}\Pr\left\{s_{1,t} < s_{2,t}\right\}. \tag{4.27}$$

Since $s_{1,t}$ and $s_{2,t}$ are i.i.d., as the coherence time of the channel increases it can be seen from (4.27) that the switching rate decreases but it does not depend on ρ. So, for the SSC scheme, from (4.25), it can be seen that as ρ increases the switching rate decreases, while for antenna selection scheme

(4.27) the switching rate does not decrease with SNR. Thus to conclude, for the channel model assumed, even for the case of time correlated channels, since the switching rates of both schemes are scaled down identically proportional to the coherence time, the SSC scheme still yields lower switching rates than the antenna selection scheme. We have added these points to Section VI in the manuscript.

Table II. Optimal Switching Threshold Θ_o, Analytical vs Simulation

SNR (dB)	Analytical (dB)	Simulation (dB)
5	2.77	2.50
8	5.82	5.50
12	9.02	9.00
15	10.84	11.00

4.7. Simulations

In this section, simulation results are provided for the ST coded MIMO system employing the switch and stay diversity combining at the receiver. For all simulations, we choose a 2×2 MIMO system implementing the switching algorithm described in (3.2) for the perfect channel case or the received power based switching algorithm for the estimated channel case. For the above described system, Alamouti codes, with symbols chosen from a QPSK constellation are used at the transmitter, yielding β=1 for all simulations considered. For a fixed value of T_c, the data transmission efficiency, is defined as $\mathcal{E}$:=$K/(K+1)$,

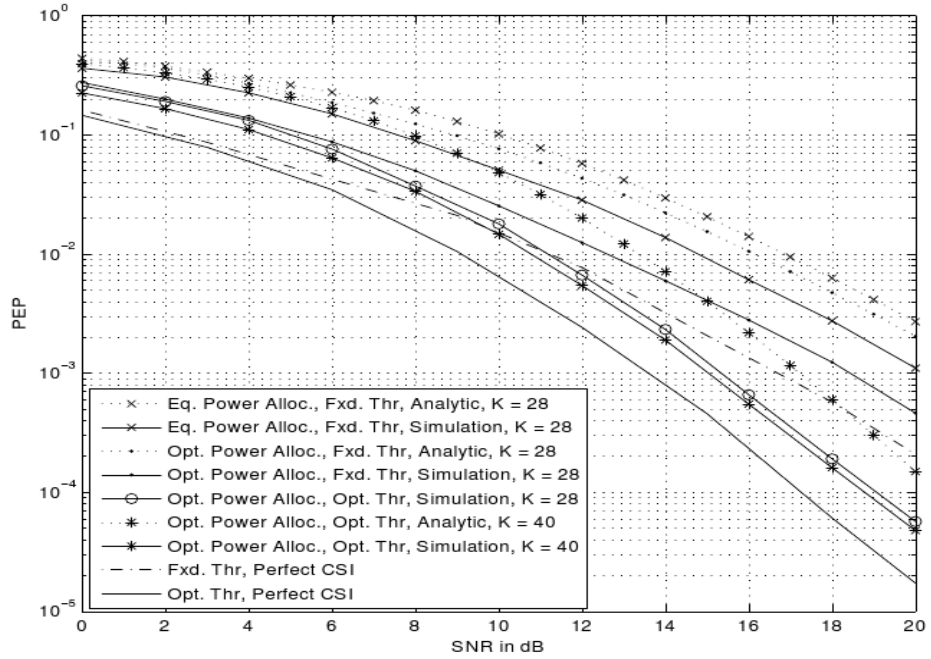

Fig. 11. Pairwise Error Probability: Simulation vs. Analytical

where it is assumed that T_τ=T_D. When the power allocation between training and data is not optimized, i.e., when equal power allocation is employed, we have σ_τ^2=σ_D^2=1 to satisfy the power constraint defined in (4.4).

The value of K implicitly defines the channel coherence time and vice-versa. Since we assume that the power estimation duration, and the training duration are fixed, to study the behavior of the system for different values of the coherence time T_c, we assign different values for K. The value of $K = 5$ is chosen to indicate very short coherence time, while $K = 40$ is chosen for large coherence time and $K = 28$ is for a moderate value. For example, in an urban environment, a slow fading channel is known to have a doppler of about

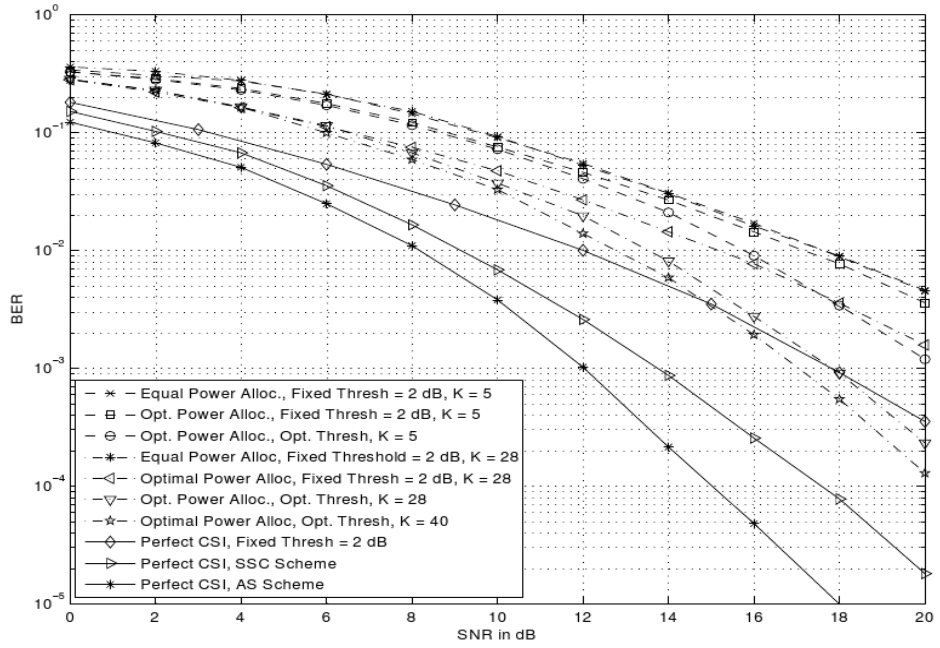

Fig. 12. BER: Alamouti Code, QPSK Symbols, Receive End SSC

5 Hz, while a fast fading channel has a doppler of about 50 Hz. So, if the training, power estimation and each data codeword duration is assumed to be 5 ms each, then, ignoring the guard bands, a coherence time of about 35 ms is represented when $K = 5$, a coherence time of about 150 ms when $K = 28$, and a coherence time of about 210 ms when $K = 40$.

In Fig. 11, we illustrate the effect of optimization of power allocation and switching threshold for K=28 and 40 on the PEP performance. The analytical upper bound derived in (4.11) is also plotted against the simulated PEP curve to validate the tightness of the bound. It can be seen that the analytical upper bound is about 3 dB loose for the fixed threshold case. But

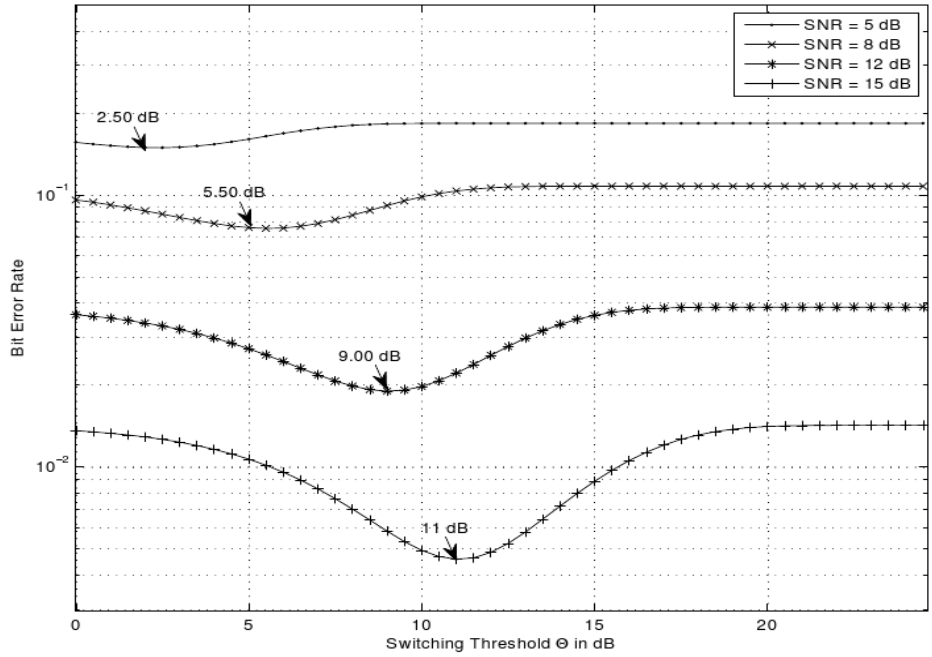

Fig. 13. BER for orthogonal ST codes vs Switching Threshold Θ

by optimizing the switching threshold, for K=40, the analytical curve is only about 1 dB loose. Additionally, we see that for $K = 28$, by optimizing the power allocation between training and data for the fixed threshold case, a gain of about 2 dB is obtained at 10^{-3} PEP which reduces the loss in performance to only 2 dB compared to the perfect CSI case with fixed threshold. By using the optimal switching threshold in addition to optimal power allocation 4 dB improvement in the PEP can be observed with respect to the fixed threshold PEP simulation curve. The loss compared to the perfect CSI case using the optimal switching threshold is only about 1.8 dB at 10^{-3} PEP.

In Fig. 12, we plot the BER performance of the system for K=5, 28 and

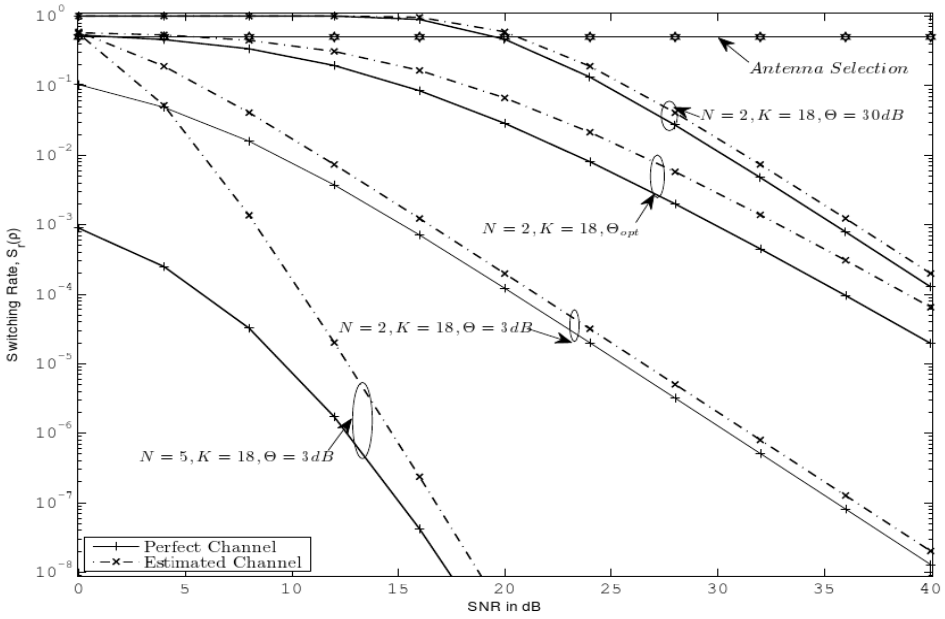

Fig. 14. Switching Rates $S_r(\rho)$: Perfect CSI vs Estimated CSI

40, with $\mathcal{E}$=0.83, 0.97, and 0.98 respectively. The effect of choosing a fixed switching threshold independent of ρ, compared to using the optimal switching threshold, (4.12), on the BER performance, is shown. Also, for the estimated channel case, in Fig. 12, the effect of increasing K on the BER performance is shown to validate Theorem 3. It can be seen that the performance and diversity order of the system is significantly degraded when a fixed threshold is employed instead of the optimal switching threshold. The system performs the worst when neither the power allocation nor the switching threshold is optimized. For the fixed threshold and equal power allocation case the BER performance is identical for both $K = 5$ and 28. For $K = 5$, for a fixed

switching threshold, but optimal power allocation, a marginal improvement of about 0.4 dB can be observed while an improvement of about 3 dB can be seen for $K = 28$. By optimizing both the power allocation and switching threshold, an improvement of 2 dB for $K = 5$ and about 6 dB for $K = 28$ can be seen at about 10^{-3} BER. Also, a significant improvement in the diversity order can be observed. To corroborate Theorem 3, we also plot the BER curve for K=40 with optimal power allocation and optimal switching threshold. It can be seen that for $K = 40$, the loss in performance due to channel estimation is now less than 3 dB at a BER of 10^{-3} which indicates the gain in performance by increasing K when γ is optimized. The SSC scheme with perfect CSI at the receiver is also illustrated in Fig. 12. Due to the extra $(\log(\rho))^N$ penalty term in SSC performance compared to the AS case, the achievable diversity gain is observed at higher SNRs for the SSC case. Both the systems employing SSC and antenna selection achieve a diversity order of $2N$ at high SNR values which cannot be illustrated in the plots.

To verify the validity of the optimal switching threshold in (4.12), we simulate the BER of the system in Fig. 13, for fixed SNR values, over a range of switching thresholds to find the optimum value. The BER for average SNR values of $\rho = 5$, 8, 12 and 15 dB is illustrated to show the logarithmic growth of the switching threshold that minimizes the BER. Further, in Table II, we compare the analytically derived optimal switching threshold in (4.12), against the switching threshold values obtained numerically from Fig. 13 for

these average SNR values. It can be seen that the analytical and the simulation values are very close to each other which validates our analytical results.

Finally, in Fig. 14, we illustrate the behavior of the switching rate with SNR and number of transmit antennas as discussed in Section 4.5, for both the perfect and the estimated channel case. We assume a normalized coherence time T_c=1 sec. It can be seen that the optimal switching threshold scales as $O\left((\ln(\rho)/\rho)^N\right)$. When the switching threshold is fixed at a low value compared to the SNR ρ, it can be seen that the switching rates are the lowest, as compared to the case when the switching threshold is fixed at Θ=30 dB, leading to very high switching rates. Implementing the optimal switching threshold, it can be seen that switching rates in-between the above two cases can be achieved along with significant performance improvement as seen from Fig. 11, and 12. It can also be seen that in all three cases, the estimated channel switching rate is marginally lower compared to the perfect channel case. Additionally, the switching rate for systems employing antenna selection is also shown and it can be seen that for i.i.d. block fading channels, the SSC scheme has a smaller switching rate.

5. DIFFERENTIAL MIMO SYSTEMS WITH RECEIVE SWITCH AND STAY DIVERSITY COMBINING

5.1. Introduction

Communication with multiple antennas have been of prime interest for over a decade [3, 45]. Results capturing the significant improvements in error rate performance, and achievable capacity due to multiple antennas have been well established [1–3]. Unfortunately, due to the limitations in implementing multiple antenna schemes posed mainly by space constraints and hardware shortcomings, many of the well established multiple antenna results are yet to be realized.

To address hardware complexity in single input multiple output (SIMO) systems, antenna selection at the receiver is one of the proposed approaches [29, 33, 46, 47], where only the best antenna, in the sense of a largest instantaneous received signal metric (SNR or power), is used for decoding the received signal. However, to select the best antenna, the metric has to be monitored on all antennas. To reduce complexity, in [34], the switch and stay combining technique was proposed where the antenna used for reception is switched only when the metric on the current antenna falls below a predetermined threshold. The advantage of this scheme over selection of the best antenna is that the metric needs to be monitored only on the antenna in use.

The authors in [34] analyze the performance of a SIMO system implementing the continuous time model of switch and stay combining over independent Rayleigh fading channels, where a switch occurs if and only if a downward

crossing of the threshold is detected. A discrete time version was first proposed and analyzed in [35, 48]. In [36, 37], the performance of binary NCFSK modulation over Nakagami-m and Rician fading channels is analyzed for a simplified version of the switching algorithm based only on the current channel estimate on the current antenna. In [38, 39], the effect of correlated and non identical branches for different fading distributions on the performance of systems implementing the switching algorithm proposed in [36] is studied.

To reap the benefits of a multiple input multiple output (MIMO) system, in [49] an extension of the switch and stay combining scheme to space-time coded MIMO systems was proposed. It is shown that when full rank space-time codes are used at the transmitter and an optimally designed switching threshold is used at the receiver, full diversity can be achieved both when perfect or imperfect channel knowledge is available at the receiver. However, as the channel fading rapidity increases or the number of antenna pairs whose channel needs to be estimated increases, a significant challenge is posed to accurately track the channels at the receiver. Channel estimation also has the drawbacks of increasing the training overhead and system complexity. This problem can be alleviated by implementing non-coherent schemes which do not need channel information, similar to those proposed in [50, 51].

In this work, differential space-time modulation for a MIMO system employing SSC at the receiver is proposed and studied for the first time in the literature. Assuming the receiver does not have access to the channel state

information, a receive-power based switching algorithm is put forth. With no need for channel estimation and a low-complexity switching algorithm at the receiver, the proposed scheme is arguably the least complex multiple input multiple output (MIMO) system which can still achieve the maximum available spatial diversity. For the proposed system, the Chernoff bound on the pairwise error probability (PEP) is derived and the optimal switching threshold that minimizes the bound is obtained. It is shown that even when full transmit diversity is attained by using appropriately designed codes, receive diversity can be achieved only by using the optimal switching threshold. It is proved analytically and corroborated by simulations that the optimal switching threshold for the proposed system is a logarithmic function of the average SNR. Further, the rank criteria for achieving full transmit diversity is proved to be identical to that of the full complexity systems but novel performance design criteria are developed to select the best codes satisfying the rank criterion. For the special case of 2 transmit antennas new unitary codes are designed which outperform existing codes designed for full complexity systems.

5.2. System Model

The system under consideration employs N transmit and 2 receive antennas because for the case of independent and identically distributed (i.i.d.) branches, having more than 2 receive antennas does not yield any additional performance improvement for switch and stay combining [40]. Since, the chan-

nel is assumed to be unknown at both the transmitter and the receiver end, differential space-time modulation with N transmit antennas spanning T time instants is employed. At the receiver, only one RF chain is present, so that at any given time data can be received only on one receive antenna. Therefore, the received signal on the i^{th} receive antenna, after demodulation, matched filtering and sampling is given by,

$$y_i^t = \sqrt{\frac{\rho}{N}} \sum_{n=1}^{N} x_n^t h_i^n + w_i^t, \qquad i = 1, 2, \qquad t = 1, \ldots, T \tag{5.1}$$

where ρ is the average SNR at the receiver, x_n^t is the transmitted symbol on the n^{th} antenna at time t, w_i^t and y_i^t are the additive noise and the received signal on the i^{th} receive antenna at time t, respectively, and h_i^n is the channel between the n^{th} transmit and the i^{th} receive antenna. The channel at the two receive branches are spatially i.i.d. and on each branch is assumed to change in an i.i.d. manner every T instants of time. The channel is neither known at the transmitter nor the receiver and has the distribution $h_i^n \sim \mathcal{CN}(0, 1)$. The additive noise is $w_i^t \sim \mathcal{CN}(0, 1)$.

Defining $\mathbf{X} \in \mathbb{C}^{T \times N}$ with its t^{th} row and n^{th} column containing $[\mathbf{X}]_{t,n} = x_n^t$, and channel matrix as $\mathbf{H} \in \mathbb{C}^{N \times 2}$ with $[\mathbf{H}]_{n,i} = h_i^n$, one can compactly represent the received signal using matrices as,

$$\mathbf{Y} = \sqrt{\frac{\rho}{N}} \mathbf{X}\mathbf{H} + \mathbf{W}, \tag{5.2}$$

where $\mathbf{Y} = [\mathbf{y}_1,\ \mathbf{y}_2]$, and $\mathbf{W}$, with dimensions $T \times 2$ denote the received signal and the additive white Gaussian noise, respectively.

5.2.1. Data Transmission

Adopting the approach in [51] for differential space-time coding, the transmitted codeword can be represented as $\mathbf{X} := \sqrt{T/N}\mathbf{\Phi}_l$, with $\mathbf{\Phi}_l^H\mathbf{\Phi}_l = \mathbf{I}_N$. The constant $\sqrt{T/N}$ is selected to ensure the transmit power constraint $\sum_{n=1}^{N} \mathrm{E}|x_n^t|^2 = 1$. The $T \times N$ differential transmission can be represented in canonical form by,

$$\mathbf{\Phi}_l = \frac{1}{\sqrt{2}}\left[\mathbf{I}_N \;\; \mathbf{V}_l^T\right]^T, \quad l = 1, \ldots, L. \tag{5.3}$$

where $\mathbf{V}_l$ is a $N \times N$ unitary matrix drawn from the constellation $\boldsymbol{\mathcal{V}} := \{\mathbf{V}_1, \ldots, \mathbf{V}_L\}$ of size $L = 2^{RN}$, and R is the transmission rate in bits per channel use. Though the first half of the canonical representation of $\mathbf{\Phi}_l$ is $\mathbf{I}_N$, only the second half is used to form the transmitted codeword across N channel uses [51], thereby preserving a rate of R bits per channel use. Since $\mathbf{\Phi}_l$ is a unitary matrix, for the rest of the work the unitary space-time modulation structure is assumed.

5.2.2. Switching Algorithm

The switching algorithm determines the trade-off between performance and the switching rate of the system, which is important from a hardware complexity perspective. Given that channel state information is neither available at the receiver nor will it be estimated, the decision to switch has to be based on the channel power. Channel power can be estimated using a bandpass filter

followed by an envelope detector and squarer, which are simple passive analog circuitry. Since A/D conversion is avoided, this further follows our goal of designing a minimum complexity MIMO system. Thereby, define $s_{i,c} := \|\mathbf{y}_i\|^2$ as received signal power currently at the i^{th} antenna, $i \in \{1, 2\}$, z_c as the currently selected antenna index and z_p as the previously selected antenna index. The switching rule is based on the discrete time algorithm proposed in [49]:

$$z_c = 1 \quad \text{iff} \quad \begin{cases} z_p = 1 \quad \text{and} \quad s_{1,c} \geq \Theta \\ \\ z_p = 2 \quad \text{and} \quad s_{2,c} < \Theta \end{cases} \tag{5.4}$$

where Θ is the switching threshold common to both branches. Since the two branches at the receiver are assumed to be identically distributed, the case when $z_c = 2$ would be similar as (5.4) with $s_{1,c}$, $z_p = 1$ interchanged with $s_{2,c}$, and $z_p = 2$ respectively.

For comparison purposes, a genie-aided switching algorithm, which has access to perfect channel SNR (unavailable in a differential receiver), is implemented to switch between the antennas. The improvement in performance when the genie-aided switcher is used instead of the above proposed algorithm is shown in Fig. 17 in the Simulations section. Since the improvement in performance is seen only at very low SNR values, and at high SNR the performance is identical to that of the receive-power based switcher, for the rest of this work the above proposed receive-power based switching algorithm is used.

5.3. Performance Analysis

5.3.1. Decoder

Let $\mathbf{y}^* := \mathbf{y}_{z_c}$ be the $T \times 1$ received signal on the selected antenna. Consider the maximum likelihood (ML) decoder:

$$\widehat{\mathbf{\Phi}} = \arg \max_{\mathbf{\Phi}_l} p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l) \tag{5.5}$$

where $p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l)$ is the distribution of the received signal on the selected antenna, conditioned on the transmitted codeword $\mathbf{\Phi}_l$.

Since $\mathbf{y}_1$ and $\mathbf{y}_2$ are identically distributed, and $\Pr(z_c = i|\mathbf{\Phi}_l) = 1/2$,

$$p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l) = p_{\mathbf{y}_i}(\mathbf{y}|z_c = i, \mathbf{\Phi}_l). \tag{5.6}$$

Using Bayes' rule, (5.6) can be rewritten as,

$$p_{\mathbf{y}_i}(\mathbf{y}|z_c = i, \mathbf{\Phi}_l) = \frac{\Pr(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l)\, p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l)}{\Pr(z_c = i|\mathbf{\Phi}_l)} \tag{5.7}$$

$$= 2\Pr(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l)\, p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l). \tag{5.8}$$

From (5.8) it is seen that the distribution of the received signal on the selected antenna can be obtained by appropriately scaling the distribution of the received signal on the i^{th} antenna. The scaling factor depends on the switching/selection algorithm used. Here, $\Pr(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l)$ is the probability that the i^{th} antenna is currently selected given that the received signal on the i^{th} antenna is $\mathbf{y}$, and the transmitted codeword is $\mathbf{\Phi}_l$. Since the switching rule in (5.4) depends on the received power and the transmitted codewords

are unitary, the distribution of the received power does not depend on the transmitted codeword $\mathbf{\Phi}_l$. Therefore, $\Pr\left(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l\right)$ does not depend on the transmitted codeword. Thus, the ML decoder in (5.5) can be expressed as,

$$\widehat{\Phi} = \arg\max_{\mathbf{\Phi}_l} p_{\mathbf{y}_i}\left(\mathbf{y}|\mathbf{\Phi}_l\right) = \arg\max_{\mathbf{\Phi}_l} \|\mathbf{y}^H \mathbf{\Phi}_l\|, \tag{5.9}$$

where the second equality follows from [50, eqn. (15)]. This is a general result that applies to a large class of switching rules. For any receive-power based switching scheme, the ML decoder is given by (5.9) if unitary codes are used, which in turn is identical to the single receive antenna system decoding rule.

5.3.2. Pairwise Error Probability Analysis

Since the derivation of the exact bit error rate (BER) expression is not tractable, in what follows the PEP is analyzed. PEP is formally defined as the probability that the transmitted codeword $\mathbf{\Phi}_1$ is incorrectly decoded as $\mathbf{\Phi}_2$ from a code book containing $\{\mathbf{\Phi}_1,\ \mathbf{\Phi}_2\}$ only.

Consider the Chernoff bound on the PEP, $\Pr(\mathbf{\Phi}_1 \rightarrow \mathbf{\Phi}_2)$, which can be obtained by computing [50, 52],

$$P_{CB}(\gamma) = \frac{1}{2}\exp(\Omega(\rho, \gamma)), \tag{5.10}$$

where

$$\Omega(\rho, \gamma) = \ln\left[\mathrm{E}_{\mathbf{y}^*}\left[\exp\left(\gamma \ln p_{\mathbf{y}_i}\left(\mathbf{y}|\mathbf{\Phi}_2\right) - \gamma \ln p_{\mathbf{y}_i}\left(\mathbf{y}|\mathbf{\Phi}_1\right)\right)\right] \middle| \mathbf{\Phi}_1\right], \tag{5.11}$$

$0 < \gamma < 1$ is a free parameter that needs to be optimized to minimize the bound in (5.10), and the expectation is with respect to $p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_1)$. Substituting (5.11) in (5.10), the Chernoff bound is given by,

$$P_{CB}(\gamma) = \frac{1}{2}\int p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_2)^{1/2}\, p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_1)^{-1/2}\, p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_1)\, d\mathbf{y}, \tag{5.12}$$

where we have selected $\gamma = 1/2$ following [50, 53]. In (5.12), $p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l)$ represents the distribution of the received signal on the i^{th} antenna given that $\mathbf{\Phi}_l$ is transmitted and is given by,

$$p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l) = \frac{1}{\pi^T \det(\mathbf{R}_l)} \exp\left(-\mathbf{y}^H \mathbf{R}_l^{-1}\mathbf{y}\right) \tag{5.13}$$

where $\mathbf{R}_l = (\rho T/N)\mathbf{\Phi}_l\mathbf{\Phi}_l^H + I_T$ and $\mathbf{R}_l^{-1} = I_T - (1/(1+N/\rho T))\,\mathbf{\Phi}_l\mathbf{\Phi}_l^H$ which is obtained by using the matrix inversion lemma [54, page 50]. Using the identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$, it follows that $\det(\mathbf{R}_l) = (\rho T/N+1)^N$. Thus (5.13) can be expressed as,

$$p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l) = \frac{1}{\pi^T (\rho T/N+1)^N} \exp\left(-\mathbf{y}^H\mathbf{y} + \frac{\mathbf{y}^H\mathbf{\Phi}_l\mathbf{\Phi}_l^H\mathbf{y}}{1+N/\rho T}\right). \tag{5.14}$$

Aside from (5.14), in order to compute (5.12) an expression for $p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l)$ is needed, which is provided next.

Theorem 4. *The received signal on the selected antenna given $\mathbf{\Phi}_l$ is transmitted is distributed as,*

$$p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l) = \left[\mathrm{I}\left(\|\mathbf{y}\|^2 \geq \Theta\right) + Pr(s_{i,c} \leq \Theta)\right] p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l) \tag{5.15}$$

where $\mathrm{I}(\cdot)$ is the indicator function and $p_{\mathbf{y}_i}(\mathbf{y}|\mathbf{\Phi}_l)$ is given by (5.14).

Proof. See Appendix A. □

Different from systems implementing antenna selection at the receiver [53, eqn. (15)], the scaling factor for the SSC scheme depends directly on the choice of switching threshold Θ. In (5.15) for a given average SNR ρ, as $\Theta \to \infty$, implying excessive switching, $\mathrm{I}\left(\|\mathbf{y}\|^2 \geq \Theta\right) \to 0$ and $\Pr\left(s_{i,c} \leq \Theta\right) \to 1$ thereby leading to $p_{\mathbf{y}^*}\left(\mathbf{y}|\mathbf{\Phi}_l\right) = p_{\mathbf{y}_i}\left(\mathbf{y}|\mathbf{\Phi}_l\right)$. Similarly, as $\Theta \to 0$, implying very rare switching, $\mathrm{I}\left(\|\mathbf{y}\|^2 \geq \Theta\right) \to 1$ and $\Pr\left(s_{i,c} \leq \Theta\right) \to 0$ thereby leading to $p_{\mathbf{y}^*}\left(\mathbf{y}|\mathbf{\Phi}_l\right) = p_{\mathbf{y}_i}\left(\mathbf{y}|\mathbf{\Phi}_l\right)$. Thus, for both cases, i.e., $\Theta \to \infty$ and $\Theta \to 0$, the distribution of the received signal on the selected antenna collapses to the distribution of the received signal on the i^{th} antenna prior to selection. This illustrates that Θ has to be optimally selected to extract full receive diversity. The dependence of performance on Θ will be addressed in further detail in Section 5.3.3.

The asymptotic cumulative distributive function (CDF) of the received power on the i^{th} antenna is used here because we are interested in high SNR error rate behavior of the system. Therefore, as $\rho \to \infty$, the CDF, $\Pr(s_{i,c} \leq \Theta)$, can be expressed as [53, eqn. (29)]:

$$\Pr(s_{i,c} \leq \Theta) = \sum_{k=1}^{N} \frac{B_{1,k}}{\rho^N} \Theta^k + \sum_{k=1}^{T-N} \frac{B_{2,k}}{\rho^N} \left\{ 1 - \exp(-\Theta) \sum_{i=0}^{k-1} \frac{\Theta^i}{i!} \right\} + o\left(\rho^{-N}\right), \tag{5.16}$$

where

$$B_{1,k}:=\frac{(T-1-k)!(-1)^{N-k}}{(N-k)!(T-N-1)!(T/N)^N k!} \quad B_{2,k}:=\frac{(2N-1-k)!(-1)^N}{(N-k)!(N-1)!(T/N)^N} \ , \tag{5.17}$$

and $f(x) = o(g(x))$ means $\lim_{x\to\infty} f(x)/g(x) = 0$. By substituting for $\Pr(s_{i,c} \leq \Theta)$ from (5.16) into (5.15) we obtain the expression for $p_{\mathbf{y}^*}(\mathbf{y}|\mathbf{\Phi}_l)$. Using this result along with (5.14), in the next theorem we establish an upper bound on the Chernoff bound.

Theorem 5. *The Chernoff bound on the PEP for differential modulated MIMO SSC schemes using a ML decoder can be bounded from above as,*

$$P_{CB}(\gamma) \leq \frac{\left[\sum_{t=1}^{T} \exp\left(-\left(1-c(\rho)\delta_t\right)\Theta/T\right) + \frac{B(\Theta)}{\rho^N}\right]}{2\left(1+\rho T/N\right)^N \prod_{t=1}^{T}(1-c(\rho)\delta_t)} + o\left(\rho^{-2N}\right) \tag{5.18}$$

where $0 \leq \delta_1 \leq \delta_2 \leq \ldots \leq \delta_T \leq 1$ *are the* T *eigenvalues of the matrix*

$$\Delta = \frac{1}{2}\left(\mathbf{\Phi}_2\mathbf{\Phi}_2^H + \mathbf{\Phi}_1\mathbf{\Phi}_1^H\right), \tag{5.19}$$

$B(\Theta) = \left[\sum_{k=1}^{N} B_{1,k}\Theta^k + \sum_{k=1}^{T-N} B_{2,k}\left\{1 - e^{-\Theta}\sum_{i=0}^{k-1}\frac{\Theta^i}{i!}\right\}\right]$ *with* $B_{1,k}$ *and* $B_{2,k}$ *as defined in* (5.17) *and* $c(\rho) := (1+N/\rho T)^{-1}$.

Proof. See Appendix B. □

It will be shown in Section 5.3.3 that $\delta_T < 1$ is required for full diversity, which also is equivalent to the largest singular value of $\mathbf{\Phi}_2^H\mathbf{\Phi}_1$ being less than 1 [53]. Further, using the following simple result,

$$\mathrm{I}\left(v_1 + v_2 \ldots + v_T \geq \Theta\right) \geq \prod_{t=1}^{T} \mathrm{I}(v_i \geq \Theta/T), \tag{5.20}$$

and using similar algebra as in the proof of Theorem 5, the PEP can be approximated with the help of a lower bound on the Chernoff bound as

$$P_{CB}(\gamma) \geq \frac{\left[\exp\left(-\frac{\Theta\sum_{t=1}^{T}(1-c(\rho)\delta_t)}{T}\right) + \frac{B(\Theta)}{\rho^N}\right]}{2\left(1+\rho T/N\right)^N \prod_{t=1}^{T}\left(1-c(\rho)\delta_t\right)} + o\left(\rho^{-2N}\right). \tag{5.21}$$

Note that (5.21) is not necessarily a lower bound on the PEP even though it is a lower bound to $P_{CB}(\gamma)$. In Section 5.5 it will be illustrated that (5.21) is a tighter approximation to the actual PEP. Further, for a Θ which is not a function of the average SNR ρ, both (5.18) and (5.21) behave as $O(\rho^{-N})$. Thus, in the following section the optimal Θ that minimizes the bound is derived.

5.3.3. Optimal Switching Threshold

In what follows, (5.18) is minimized to find the optimal switching threshold. Since minimizing (5.18) with respect to Θ is not straightforward, an asymptotic approximation of (5.18) is used. The resulting optimal Θ will naturally be a function of average SNR ρ, and will also be shown to offer full diversity.

Since diversity order is a high SNR phenomenon, only the dominant terms in (5.18) are retained for optimization assuming that Θ is an increasing function of ρ,

$$\frac{1}{(\rho T/N)^N \prod_{t=1}^{T}(1-\delta_t)} \exp(-(1-\delta_T)\Theta/T) + \frac{B_{1,N}\Theta^N}{\rho^{2N}(T/N)^N \prod_{t=1}^{T}(1-\delta_t)}, \tag{5.22}$$

where for large ρ, the result $c(\rho) \sim 1$ is used. The exponent with the largest eigenvalue δ_T of (5.19) is the dominant term at high SNR and from (5.17), it is also seen that $B_{1,N}\Theta^N$ is the dominant term in $B(\Theta)$ at high SNR.

To find the optimal Θ that minimizes (5.22), the derivative of (5.22) with respect to Θ is set to zero to obtain,

$$\Theta \exp((\overline{\lambda}_{\min})\Theta/(T(N-1))) = \left(\frac{(\overline{\lambda}_{\min})\rho^N}{NTB_{1,N}}\right)^{1/(N-1)}. \tag{5.23}$$

Solving for Θ in (5.23), the optimal switching threshold minimizing (5.22) is,

$$\Theta_o(\rho) = \frac{T(N-1)}{(\overline{\lambda}_{\min})} W\left(\frac{(\overline{\lambda}_{\min})}{T(N-1)}\left(\frac{(\overline{\lambda}_{\min})\rho^N}{NTB_{1,N}}\right)^{1/(N-1)}\right), \tag{5.24}$$

where $W(\cdot)$ represents the Lambert W-function defined to be the inverse of the function $x\exp(x)$. Since we are studying the high SNR properties, the argument of the Lambert W-function is correspondingly large. From the asymptotic series expansion of the Lambert W-function in [55, eqn. (4.19)], for large x we have $W(x) \sim \ln(x)$ which means that $W(x)/\ln(x) \to 1$ as $x \to \infty$. Using this $\Theta_o(\rho)$ in (5.24) is asymptotically approximated as:

$$\Theta_o(\rho) \sim \frac{T(N-1)}{(\overline{\lambda}_{\min})} \ln\left(\frac{(\overline{\lambda}_{\min})}{T(N-1)}\left(\frac{(\overline{\lambda}_{\min})\rho^N}{NTB_{1,N}}\right)^{1/(N-1)}\right). \tag{5.25}$$

Similar to the optimal switching threshold for MIMO SSC systems with perfect channel knowledge [49], $\Theta_o(\rho)$ is again a logarithmic function of the SNR ρ. However, unlike the known channel case a closed form minimum of the PEP with respect to Θ is not tractable, requiring high SNR approximations of the preceding equations.

By using the switching threshold derived in (5.25), the behavior of the bound in (5.18) at high SNR is studied. In the following theorem we establish that the Chernoff bound in (5.18) is $P_{CB}(\gamma) = O((\ln \rho)^N/\rho^{2N})$ when (5.25) is implemented.

Theorem 6. *For a pair* $\{\mathbf{\Phi}_1, \mathbf{\Phi}_2\}$ *with maximum eigenvalue* (5.19) *of* $\delta_T < 1$*, when the optimum switching threshold* $\Theta_o(\rho)$ *in* (5.25) *is implemented with equality, the Chernoff bound on the achievable PEP in* (5.18) *can be bounded as,*

$$\lim_{\rho\to\infty} \frac{P_{CB}(\gamma)}{(\ln \rho)^N/\rho^{2N}} \leq \frac{N^{2N}}{2^N N!(\overline{\lambda}_{min})^N \prod_{t=1}^{T}(1-\delta_t)}. \tag{5.26}$$

Proof. See Appendix C. □

Recalling that diversity order [8] is defined as $d := \lim_{\rho\to\infty} -\ln P_{CB}(\gamma)/\ln \rho$, taking the natural log of both sides of (5.26) we have,

$$\ln\left(\frac{P_{CB}(\gamma)}{\ln(\rho)^N/\rho^{2N}}\right) \leq \ln\left(\frac{N^{2N}}{2^N N!(\overline{\lambda}_{\min})^N \prod_{t=1}^{T}(1-\delta_t)}\right). \tag{5.27}$$

Dividing both sides by $-\ln \rho$ and taking the limit we have

$$d = \lim_{\rho\to\infty} \frac{-\ln(P_{CB}(\gamma))}{\ln \rho} \geq 2N. \tag{5.28}$$

Since the diversity order cannot exceed $2N$ in a $N \times 2$ system, together with (5.28) we have d=$2N$. From (5.26) it is seen that for a given number of transmit antennas N, codes with larger value of $(\overline{\lambda}_{\min})^N \prod_{t=1}^{T}(1-\delta_t)$ yield

better performance. Further, (5.26) shows full diversity because when Θ is optimized the exponential terms in (5.18) behave like $O(\rho^{-N})$. Note that if Θ is a constant for all SNR, then the exponential terms in (5.18) behave like $O(1)$ which limits the diversity order of the system to N only.

To summarize, Theorem 6 establishes that full diversity is achieved when $\delta_T < 1$, (which is guaranteed by the assumptions of Theorem 5), and $\Theta_o(\rho)$ is given as (5.25) (which is obtained by optimizing the dominant terms in (5.22)). In what follows (5.26) will be used to further refine the code design criteria to distinguish between codes achieving full diversity.

Further, in Section 5.3.2 it was noted that a fixed switching threshold will lead to either excessive switching or very rare switching depending on the average SNR. The switching rate is defined as $S_r(\rho) := (1/T_c)\Pr(s_{i,c} < \theta)$, where T_c is the coherence time of the channel and $\Pr(s_{i,c} < \theta)$ is the probability of switching based on the proposed switching algorithm and the i.i.d. assumption of the two receive antennas. In the high SNR regime, for antenna selection, the switching rate $S_r(\rho) = 1/(2T_c) = O(1)$ while for a receiver implementing SSC with optimum switching threshold, the switching rate behaves as $S_r(\rho) = O((\ln(\rho)/\rho)^N)$, [49, eqn. (37)]. Therefore it follows that at large ρ, the switching rate of a SSC-based receiver is significantly smaller than that of an antenna selection based receiver. For low SNR values though, the optimal threshold θ_o in (5.25) cannot be used since it is derived by assuming large SNR. But, by calculating θ_o by simulation, the switching rate

of a SSC receiver can be shown to be less than that of an antenna selection based receiver. Since, the performance in a high SNR regime is studied in this work, low SNR analysis will not be further pursued.

5.4. Code Design

In [51], and [56] for full complexity differential systems, the rank and determinant criteria [51, eqn. (38)] for maximizing the diversity and coding gain is proposed. However, the error rate of the proposed SSC system cannot be expressed in the form $(G_c\rho)^{-d}$ where G_c is the coding gain and d is the diversity order. Instead, in Theorem 6 it is seen that the performance of our system depends on the product $(\overline{\lambda}_{\min})^N \prod_{t=1}^{T}(1-\delta_t)$ which is different from that proposed in [51, 56]. The design criteria requires not only an $\mathbf{I}-\mathbf{\Delta}$ matrix with a large determinant but also requires the largest eigenvalue of Δ, i.e., δ_T, to be small for better performance. This also suggests that the proposed metric is more sensitive to the minimum eigenvalue of $\mathbf{I}-\mathbf{\Delta}$ being close to zero than full complexity schemes. This difference in the criteria, compared to both a differentially modulated full complexity system [51] and a differentially modulated system employing antenna selection at the receiver [53], motivates code design for the SSC case. To this end a class of unitary signal constellations termed *parametric codes* proposed in [57] are considered. Prior to presenting our design criteria the parametric codes are briefly described.

Defining the integers $k_1, k_2, k_3 \in \{0, \ldots, L-1\}$, the parametric codes with constellation size L for 2 transmit antenna systems are given by,

$$\mathcal{V}(k_1, k_2, k_3) = \{A_l(k_1, k_2, k_3) | l = 0, \ldots, L-1\} \tag{5.29}$$

where $A_l(k_1, k_2, k_3)$ is a 2×2 unitary matrix obtained as,

$$A_l(k_1, k_2, k_3) = \begin{pmatrix} \exp(j\varsigma_L) & 0 \\ 0 & \exp(jk_1\varsigma_L) \end{pmatrix}^l \cdot \begin{pmatrix} \cos(k_2\varsigma_L) & \sin(k_2\varsigma_L) \\ -\sin(k_2\varsigma_L) & \cos(k_2\varsigma_L) \end{pmatrix}^l \cdot \begin{pmatrix} \exp(jk_3\varsigma_L) & 0 \\ 0 & \exp(-jk_3\varsigma_L) \end{pmatrix}^l \tag{5.30}$$

where $\varsigma_L = 2\pi/L$. As stated by the authors in [57], when $k_2 = k_3 = 0$ diagonal cyclic codes proposed in [51] are obtained.

Among all possible constellations of size L, codes based on

$$\arg \max_{k_1,k_2,k_3} \left(\min_{\mathbf{\Phi}_1, \mathbf{\Phi}_2} \left((\overline{\lambda}_{\min})^N \prod_{t=1}^{T} (1 - \delta_t) \right) \right), \tag{5.31}$$

are found where the minimization in the above expression is over all possible codeword pairs $\{\mathbf{\Phi}_1, \mathbf{\Phi}_2\}$ in the code defined by $[k_1, k_2, k_3]$. It is easy to see from the above criteria that if the code has codeword pairs that lead to $\delta_T = 1$ (i.e., $\mathbf{I}_T - \mathbf{\Delta}$, with $\mathbf{\Delta}$ as in (5.19), loses rank), then (5.31) will be zero, thereby eliminating such codes in the search process. By performing an exhaustive search over all possible constellations, a number of codes which perform better than the existing codes designed for full complexity systems are found. Table III provides a list of these codes. Interestingly, the best parametric codes

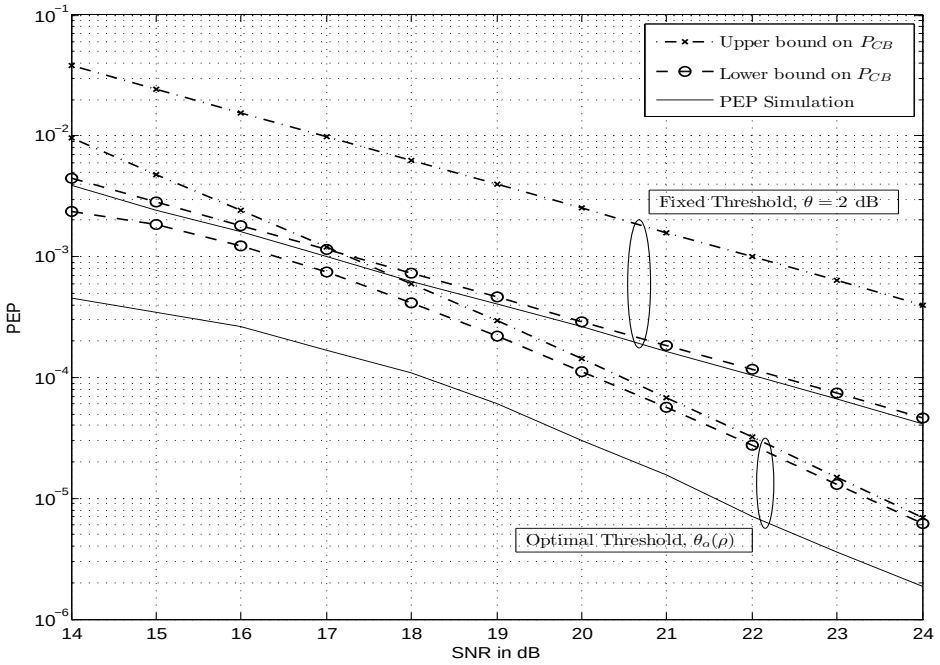

Fig. 15. PEP: Simulation vs. Analytical, $N = 2$

designed for $R = 2.5$ and $R = 3$ for differential MIMO systems implementing antenna selection at the receiver [53] are included in our set of best performing codes: $[31, 4, 2]$ for rate 2.5 and $[63, 6, 0]$ for rate 3.

5.5. Simulations

For all simulations a 2×2 system is considered. The switching algorithm proposed in (5.4) is employed at the receiver. For Figures 15, 16, 17 the transmission rate is set to $R = 1$ bit per channel use, thereby resulting in a constellation size of $L = 2^{RN} = 4$.

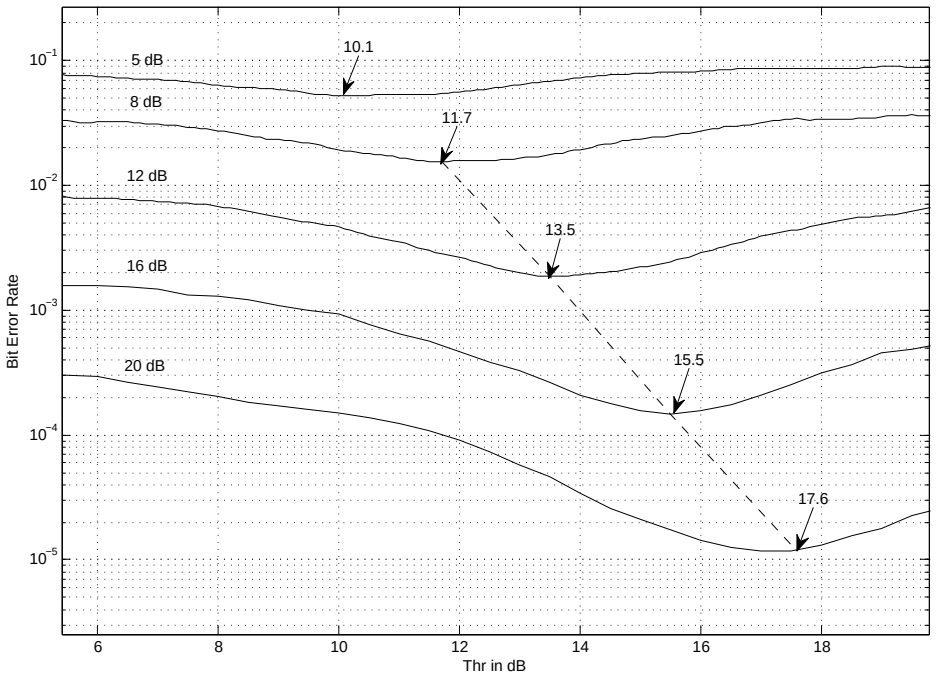

Fig. 16. Optimal Switching Threshold, $R = 1$, $N = 2$

First the PEP of receive SSC for MIMO systems using diagonal cyclic codes is illustrated. In Fig. 15 the PEP obtained for the fixed and optimal threshold cases are compared against their respective analytical plots. When a fixed threshold is used, it is seen that a diversity order of only 2 can be achieved, in agreement with the results in Section 5.3.2. But when the optimal switching threshold given by (5.25) is used, a larger diversity order is achievable. Since a diversity order of 4 is only seen at extremely large SNR values, it is not possible to show it in our plots. Further, it is also seen for the optimal threshold case the bound is about 2 dB loose at 10^{-5} BER while for the fixed threshold case it is about 5 dB loose at 10^{-3} BER implying the Chernoff bound is tighter

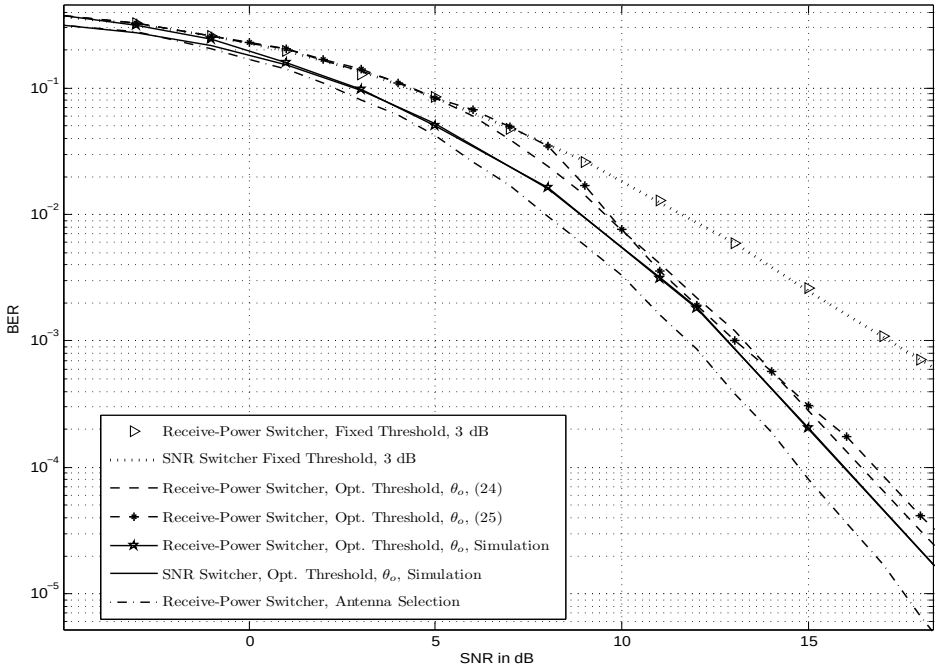

Fig. 17. Fixed Threshold vs. Optimal Threshold, $N = 2$, Diagonal Cyclic Codes, $R = 1$

when the optimum switching threshold is used. The lower bound (5.21) on the Chernoff bound is also plotted and it can be seen that for the fixed Θ case the lower bound is a very tight upper bound on the actual PEP. But when the optimal switching threshold Θ_o is used, only a gain of 0.2 dB over the upper bound on the Chernoff bound is obtained at high SNR. Therefore for the cases considered, the lower bound on the Chernoff bound in (5.21) can be used as a more accurate approximation to the actual PEP.

In (5.24) and (5.25) the optimal switching thresholds based on high SNR approximation of the Chernoff bound on the PEP is obtained. In Fig. 15 it was shown that this choice of optimal threshold still helps achieve full diversity

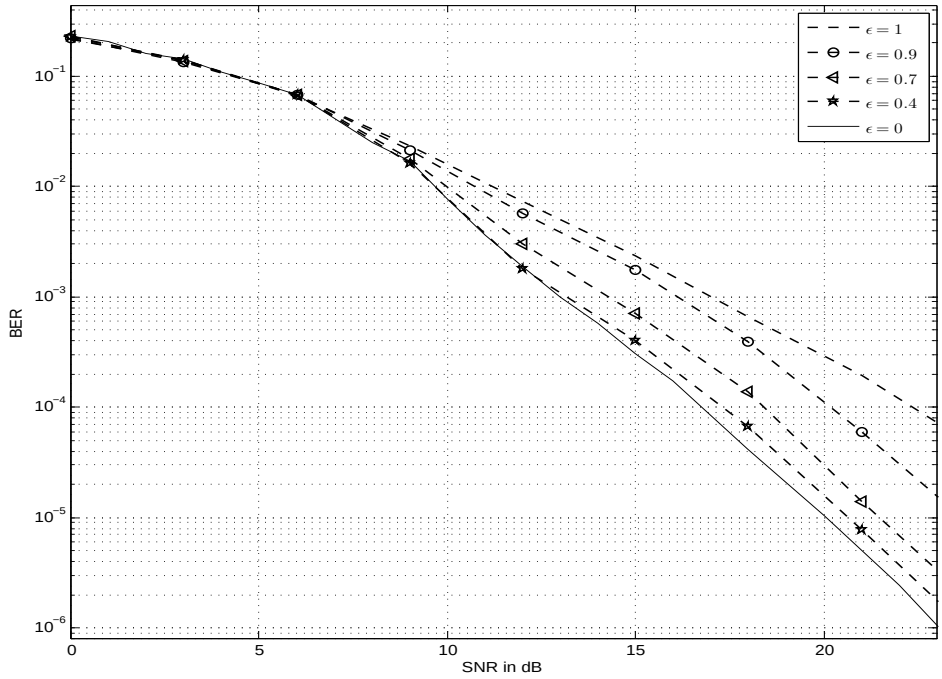

Fig. 18. Correlated Receive Branches, $N = 2$, Diagonal Cyclic Codes, $R = 1$

without being able to draw any conclusions on whether the actual PEP is minimized. To obtain better insight the threshold which minimizes the BER is calculated by simulation. In Fig. 16 the BER for various SNR values is plotted and the threshold yielding the minimum BER value for each SNR is obtained. From Fig. 16 it can be seen that optimizing the switching threshold leads to more significant improvement of BER at high SNR than at low SNR. Further, in Fig. 16, which is a $\log - \log$ plot, for sufficiently large ρ it is seen that Θ_o increases linearly with ρ. This verifies the result in (5.25) that the optimal threshold is a logarithmic function of the average SNR.

In Fig. 17 BER for fixed and optimal thresholds is plotted. The BER of

a differential 2×2 system employing receive antenna selection is also plotted for comparison purpose. Again, it is observed that when a fixed threshold is used, a maximum diversity of only 2 can be achieved. In Fig. 17, the BER for the following three cases are compared: 1) the optimal threshold in (5.24), with the Lambert W-function, is used, 2) the optimal threshold in (5.25), with the log function, is used, and 3) optimal threshold values derived from Fig. 16 are used. In all 3 cases, the slopes are identical to that obtained by using antenna selection, implying that a diversity order of 4 will be achieved. As justified in Section 5.3.2, the optimal threshold functions in (5.24) and (5.25) only help achieve maximum diversity but do not necessarily minimize the BER at high SNR. For mid SNR values, the BER obtained by (5.25) outperforms that obtained by using (5.24). For low SNR values, it is seen that for switching thresholds derived from Fig. 16, the SSC BER is very close to that of the AS scheme. For high SNR though the SSC BER is about 2 dB away. Importantly, when the optimal switching threshold is used, due to the $O((\ln \rho)^N \rho^{-2N})$ behavior of the error rate of the SSC system compared to the $O(\rho^{-2N})$ behavior of the AS system, the $(\ln \rho)^N$ penalty term mandates a larger SNR for the SSC BER curve than the AS BER curve to achieve full diversity. But both schemes achieve a diversity order of 4 at high SNR values.

Further, the BER performance of a differential MIMO receive-SSC system implementing a SNR based switching algorithm is also plotted as a benchmark. Since, in a differential system the SNR is neither known or estimated, a genie-

Table III. Best performing Parametric Codes $[k_1, k_2, k_3]$ for Differential MIMO-SSC systems with $N = 2$

Parametric Codes		
Rate 2	Rate 2.5	Rate 3
$[11, 4, 2]$	$[31, 4, 2]$	$[63, 6, 0]$
$[11, 12, 2]$	$[31, 12, 2]$	$[63, 38, 0]$
$[11, 14, 4]$	$[31, 4, 10]$	$[63, 58, 0]$
$[11, 12, 4]$	$[31, 12, 10]$	$[63, 6, 32]$
$[11, 4, 12]$	$[31, 28, 10]$	$[63, 38, 32]$
$[11, 12, 12]$	$[31, 12, 16]$	$[63, 58, 32]$

aided switching algorithm which has access to the SNR, as explained in Section 5.2.2, has been used. It can be seen that when a fixed switching threshold is used, the performance of the SNR based and the receive-power based switching algorithms are identical. But when optimal switching thresholds, calculated by simulation, are used, only at extremely low SNR the genie-aided switcher outperforms the proposed receive-power based switcher.

In practice, the receive antennas might not be i.i.d. but spatially correlated. To this end, the channel on the two receive branches is assumed to be correlated with a correlation co-efficient ϵ. With this model, in Fig. 18 the BER performance of a SSC system is illustrated. It is seen that when the optimal threshold in (5.25) is used, the BER performance deteriorates with ϵ. Further, when $\epsilon = 1$ no receive diversity will be achieved.

For the special case of a 2×2 systems, in Figs. 19, 20 and 21 the performance of the parametric codes designed for the SSC scheme with those

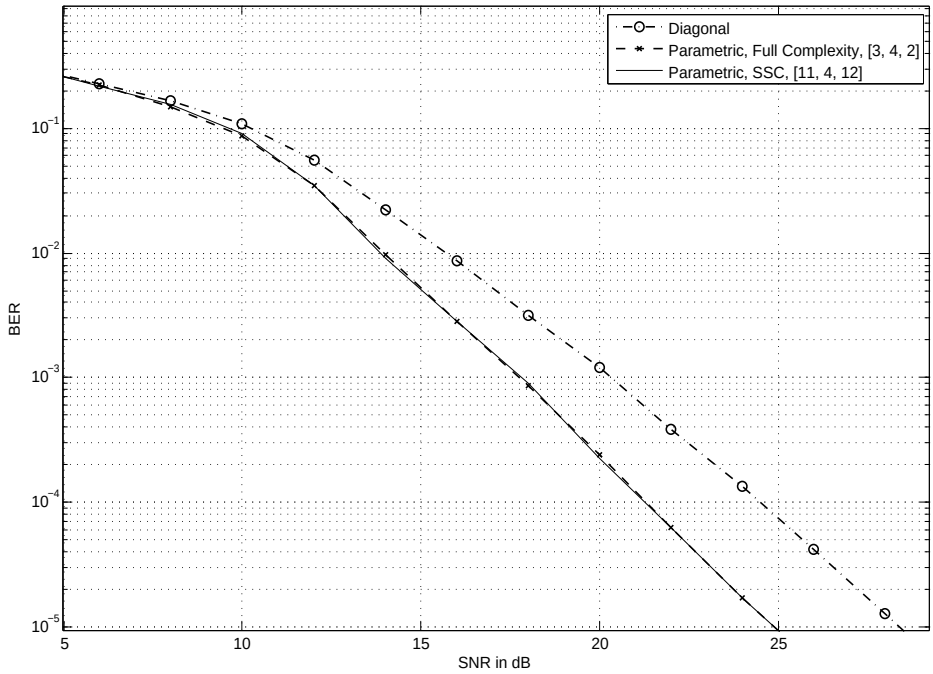

Fig. 19. Parametric Codes, $R = 2$, $N = 2$

designed for full complexity systems are compared. The performance of diagonal cyclic codes proposed in [51] is also plotted to help in this process. In Table III, since the performance of the proposed codes for a given rate are identical to one another, only one of the proposed codes for each rate is plotted for comparison purpose. In Fig. 19, codes with rate 2 bits per channel use are compared. For $N = 2$, and $R = 2$, it is seen that the parametric code leads to about 3 dB improvement in performance compared to the diagonal cyclic codes. It can be seen that compared to the best code for full complexity systems, $[3, 4, 2]$, the performance of the parametric codes designed for SSC systems $[11, 4, 12]$ are almost identical. In Fig. 20 the performance of the

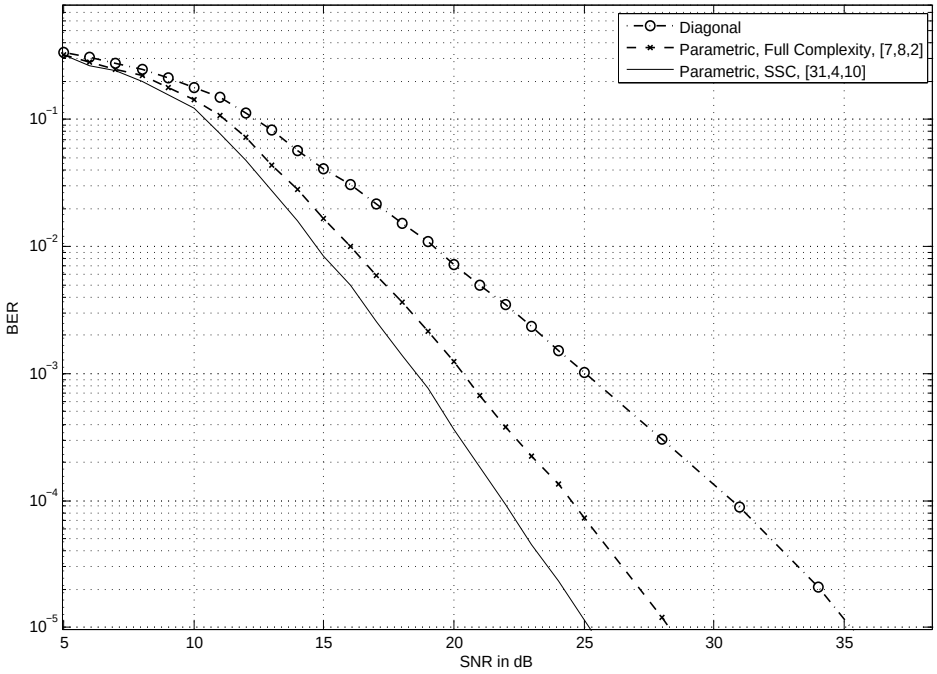

Fig. 20. Parametric Codes, $R = 2.5$, $N = 2$

system for $R = 2.5$ is compared. The parametric codes designed for full complexity systems, $[7, 8, 2]$ outperform the diagonal cyclic codes by about 7 dB at 10^{-5} BER, while our design $[31, 4, 10]$ outperforms the full complexity system codes by about 3 dB at the same BER. Finally, in Fig. 21, comparing the performance for $R = 3$, it is seen that the parametric codes designed for SSC systems $[63, 58, 0]$ outperform the full complexity parametric codes $[7, 10, 0]$ by about 3 dB.

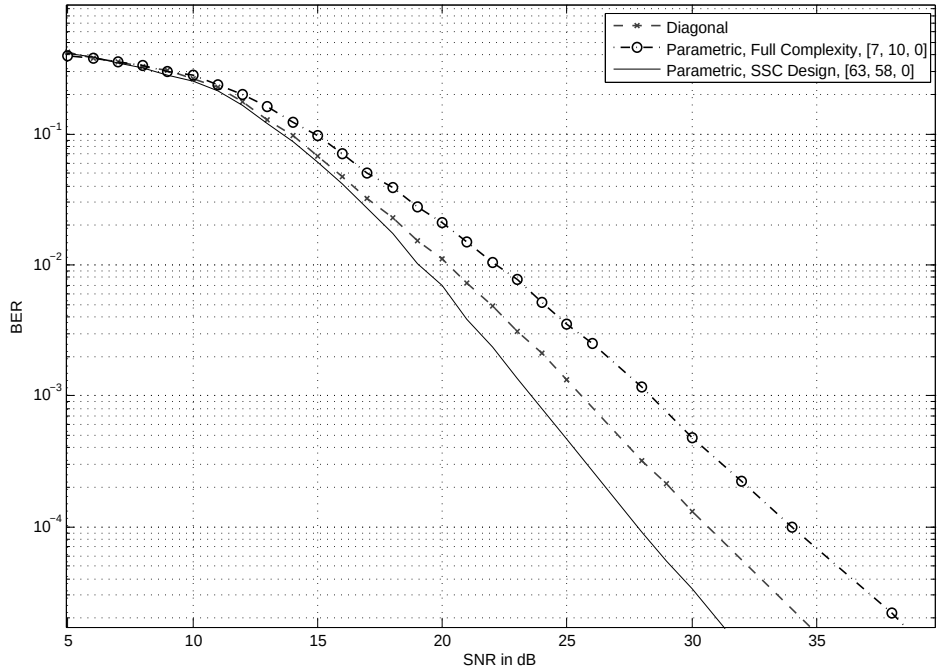

Fig. 21. Parametric Code, $R = 3$, $N = 2$

Appendix

Appendix A: Proof of Theorem 4

Using the switching algorithm defined in (5.4) we can express the probability of selecting the i^{th} antenna conditioned on the received signal as,

$$\begin{aligned}\Pr\left(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l\right) &= \Pr\left(z_c = 1|\mathbf{\Phi}_l, \mathbf{y}_1 = \mathbf{y}\right)\Pr\left(i = 1\right) + \\ &\quad \Pr\left(z_c = 2|\mathbf{\Phi}_l, \mathbf{y}_2 = \mathbf{y}\right)\Pr\left(i = 2\right), \\ &= \Pr\left(z_c = 1|\mathbf{y}_1 = \mathbf{y}, \mathbf{\Phi}_l\right), \qquad (5.32)\end{aligned}$$

which is obtained due to assumption of i.i.d. branches. Conditioning on $\mathbf{y}_1 = \mathbf{y}$ implies $s_{i,c} = \|\mathbf{y}\|^2$. So, we can rewrite (5.32), $\Pr\left(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l\right)$, as,

$$\begin{aligned}
&= \Pr\Big(\big(z_c = 1 \text{ and } z_p = 1 \text{ and } \|\mathbf{y}^*\|^2 \geq \Theta\big) \text{ or} \\
&\qquad\qquad (z_c = 1 \text{ and } z_p = 2 \text{ and } s_{2,c} < \Theta)\Big) \\
&= \Pr\big(z_c = 1 \text{ and } z_p = 1 \text{ and } \|\mathbf{y}^*\|^2 \geq \Theta\big) + \\
&\qquad\qquad \Pr\left(z_c = 1 \text{ and } z_p = 2 \text{ and } s_{2,c} < \Theta\right) \\
&= \frac{1}{2}\left[\Pr\left(\|\mathbf{y}^*\|^2 \geq \Theta\right) + \Pr\left(s_{2,c} < \Theta\right)\right] \\
&= \frac{1}{2}\left[\mathrm{I}\left(\|\mathbf{y}^*\|^2 \geq \Theta\right) + \Pr\left(s_{2,c} < \Theta\right)\right] \qquad (5.33)
\end{aligned}$$

where (5.33) is obtained due to the assumption of temporal independence of the channel. Also, we define the indicator function as $\mathrm{I}\left(\|\mathbf{y}^*\|^2 \geq \Theta\right) = 1$ if and only if the condition is satisfied. Further since both branches are i.i.d, we have

$$\Pr\left(z_c = i|\mathbf{y}_i = \mathbf{y}, \mathbf{\Phi}_l\right) = \frac{1}{2}\left[\mathrm{I}\left(\|\mathbf{y}^*\|^2 \geq \Theta\right) + \Pr\left(s_{i,c} < \Theta\right)\right] \qquad (5.34)$$

Using this in (5.8), the required result is obtained.

Appendix B: Proof of Theorem 5

Using (5.14), (5.15) and (5.16) in (5.12), the Chernoff bound on the PEP can now be expressed as,

$$P_{CB}(\gamma) = \frac{1}{2}\frac{1}{\pi^T(1+\rho T/N)^N}\int \exp\left(-\mathbf{y}^H\mathbf{\Gamma}(\rho)\mathbf{y}\right)\cdot \left[\mathrm{I}\left(\|\mathbf{y}\|^2 \geq \Theta\right) + \sum_{k=1}^{N}\frac{B_{1,k}}{\rho^N}\Theta^k + \sum_{k=1}^{T-N}\frac{B_{2,k}}{\rho^N}\left\{1- \exp(-\Theta)\sum_{i=0}^{k-1}\frac{\Theta^i}{i!}\right\} + o\left(\rho^{-N}\right)\right] d\mathbf{y}, \tag{5.35}$$

where

$$\mathbf{\Gamma}(\rho) := \frac{1}{2}\left(\mathbf{R}_2^{-1} + \mathbf{R}_1^{-1}\right) = \mathbf{I}_T - \frac{1}{1+N/\rho T}\mathbf{\Delta} \tag{5.36}$$

with $\mathbf{\Delta} = 1/2\left(\mathbf{\Phi}_2\mathbf{\Phi}_2^H + \mathbf{\Phi}_1\mathbf{\Phi}_1^H\right)$ and $B_{1,k}$ and $B_{2,k}$ as defined in (5.17). Define the eigenvalue decomposition $\mathbf{\Gamma}(\rho) := \mathbf{U}\cdot\mathrm{diag}[(1-c(\rho)\delta_1),\ldots,(1-c(\rho)\delta_T)]\cdot\mathbf{U}^H$ where it is recalled that $c(\rho) = (1+N/T\rho)^{-1}$. Further, by defining $\mathbf{r} := \mathbf{U}^H\mathbf{y}$, and performing a change of variables, the Chernoff bound can be expressed as,

$$P_{CB}(\gamma) = \frac{1}{2\pi^T(1+\rho T/N)^N}\int \exp\left(-\sum_{t=1}^{T}(1-c(\rho)\delta_t)|r_t|^2\right) \left[\mathrm{I}\left(\|\mathbf{r}\|^2 \geq \Theta\right) + \sum_{k=1}^{N}\frac{B_{1,k}}{\rho^N}\Theta^k + \sum_{k=1}^{T-N}\frac{B_{2,k}}{\rho^N}\left\{1- \exp(-\Theta)\sum_{i=0}^{k-1}\frac{\Theta^i}{i!}\right\}\right] d\mathbf{r} + o\left(\rho^{-2N}\right). \tag{5.37}$$

Defining $r_t := e^{i\alpha}\sigma_t$ and after a change of variables the Chernoff bound in (5.37) can be expressed as,

$$P_{CB}(\gamma) = \frac{2^T}{2\pi^T \left(1+\rho T/N\right)^N} \int_0^\infty \cdots \int_0^\infty \exp\left(-\sum_{t=1}^{T}(1-c(\rho)\delta_t)\sigma_t^2\right) \left[\mathrm{I}\left(\sigma_1^2+\ldots+\sigma_T^2 \geq \Theta\right) + \sum_{k=1}^{N}\frac{B_{1,k}}{\rho^N}\Theta^k + \sum_{k=1}^{T-N}\frac{B_{2,k}}{\rho^N}\left\{1-\exp(-\Theta)\sum_{i=0}^{k-1}\frac{\Theta^i}{i!}\right\}\right] \sigma_1\ldots\sigma_T d\sigma_1\ldots d\sigma_T + o\left(\rho^{-2N}\right). \quad (5.38)$$

Now, defining $v_t := \sigma_t^2$, and by a change of variables, the upper bound can now be expressed as,

$$P_{CB}(\gamma) = \frac{1}{2\pi^T \left(1+\rho T/N\right)^N} \int_0^\infty \cdots \int_0^\infty \exp\left(-\sum_{t=1}^{T}(1-c(\rho)\delta_t)v_t\right) \left[\mathrm{I}\left(v_1+\ldots+v_T \geq \Theta\right) + \sum_{k=1}^{N}\frac{B_{1,k}}{\rho^N}\Theta^k + \sum_{k=1}^{T-N}\frac{B_{2,k}}{\rho^N}\left\{1-\exp(-\Theta)\sum_{i=0}^{k-1}\frac{\Theta^i}{i!}\right\}\right] dv_1\ldots dv_T + o\left(\rho^{-2N}\right). \quad (5.39)$$

The exact evaluation of this function is quite cumbersome. Instead, to simplify the expression, the indicator function is upper bounded as $\mathrm{I}\left(v_1+v_2\ldots+v_T \geq \Theta\right) \leq \sum_{t=1}^{T}\mathrm{I}(v_t \geq \Theta/T)$, which is valid because the union of the individual indicator functions will lead to an overlap of the area of

interest. Substituting,

$$P_{CB}(\gamma) \leq \frac{1}{2\left(1+\rho T/N\right)^N} \int_0^\infty \cdots \int_0^\infty \exp\left(-\sum_{t=1}^{T}(1-c(\rho)\delta_t)v_t\right)\left[\sum_{t=1}^{T} \mathrm{I}(v_t \geq \Theta/T) + \frac{B(\Theta)}{\rho^N}\right] dv_1 \ldots dv_T + o\left(\rho^{-2N}\right) \tag{5.40}$$

where $B(\Theta) = \left[\sum_{k=1}^{N} B_{1,k}\Theta^k + \sum_{k=1}^{T-N} B_{2,k}\left\{1 - e^{-\Theta} \cdot \sum_{i=0}^{k-1} \frac{\Theta^i}{i!}\right\}\right]$. Note that if $1 - c(\rho)\delta_t = 0$, then the corresponding exponential term becomes 1 and thereby leads to $\int_{\Theta/T}^{\infty} I(v_t < \Theta/T) = \infty$. This in turn results in $P_{CB}(\gamma) = \infty$. Solving the integral in (5.40), the bound on the Chernoff bound can be expressed as,

$$P_{CB}(\gamma) \leq \frac{\left[\sum_{t=1}^{T} \exp\left(-\left(1-c(\rho)\delta_t\right)\Theta/T\right) + \frac{B(\Theta)}{\rho^N}\right]}{2\left(1+\rho T/N\right)^N \prod_{t=1}^{T}(1-c(\rho)\delta_t)} + o\left(\rho^{-2N}\right) \tag{5.41}$$

Appendix C: Proof of Theorem 6

In this Appendix, the upper bound stated in (5.26) is established. Substituting the optimal switching threshold in (5.25) for Θ into (5.18), since $\Theta(\rho)$ is a logarithmic function of ρ, it follows that,

$$\sum_{t=1}^{T} \exp\left(-\lambda_t \Theta_o(\rho)/T\right) = o(\rho^{-N}) \tag{5.42}$$

and

$$\frac{B(\Theta_o(\rho))}{\rho^N} = \frac{N^{2N}}{2^N N! (\overline{\lambda}_{\min})^N \prod_{t=1}^{T} \overline{\lambda}_t} \frac{(\ln \rho)^N}{\rho^N} + o\left(\frac{(\ln \rho)^N}{\rho^N}\right). \tag{5.43}$$

Multiplying (5.18) by $\left(\rho^{2N}/(\ln \rho)^N\right)$ and recalling that (5.18) is an upper bound to the Chernoff bound on the PEP, the theorem follows.

6. DIVERSITY IN MULTI-USER SYSTEMS

Driven by the increasing demand for wireless services, and the scarce radio spectrum, conventional diversity techniques, such as temporal, spatial and frequency diversity, offer improvements in spectral efficiency for point-to-point links. These diversity techniques, covered in the previous chapters, are more suitable in a point-to-point context because the emphasis is on improving the communication reliability (error probability performance) rather than increasing throughput (spectral efficiency). The authors in [58] have shown that there is another form of diversity available in multiuser systems called multiuser diversity, provided by independent time varying channels across the different users. Unlike conventional diversity techniques, where fading is treated as an adverse effect to be combated, in multiuser diversity fading is exploited to improve system performance. Channel fading is a source of randomization that will be taken advantage of in wireless networks.

The central idea behind multiuser diversity is that in a large system with users fading independently, it is highly likely to find a user with a very good channel at any given point of time. By assigning channel access to the best user the system throughput can be improved. One prerequisite for multiuser diversity is that the users need to have data that is delay insensitive so that they can wait for their channel to peak. Larger the number of active users in the system, the larger the multiuser diversity gain. It is shown that this is the optimal strategy for both the uplink [59], and the downlink [60] scenarios. Implementing multiuser diversity in practical systems, leads to issues such as

fairness of system resource allocation among users, delay experienced by the individual users waiting to be served, feedback delay and estimation inaccuracy of the user fading channel. Nevertheless, multiuser diversity scheduling has been adopted in 3G (third generation) wireless data systems, e.g. CDMA 2000 1x EV-DO, and HSDPA (an enhancement to WCDMA).

Next we briefly cover the information theoretic results of the multiuser diversity system. For the uplink channel, the system model can be expressed as

$$y[m] = \sum_{n=1}^{N} h_n[m]x_n[m] + w[m] \tag{6.1}$$

where m is the time index, the channel is assumed to remain constant over an entire coherence period, and it is i.i.d. across different coherence periods. Within one coherence period, the time index is dropped and thereby $h_n[m] = h_n$. A channel with L coherence periods can be modeled as a parallel uplink channel with L subchannels with independent fade. For a given channel realization h_n, $n = 1, \ldots, N$, the sum capacity can be expressed as [15, chap. 6],

$$C_{\text{sum}} = \max_{P_{n,l};n=1,\ldots,N;l=1,\ldots,L} \frac{1}{L} \sum_{l=1}^{L} \log\left(1 + \frac{\sum_{n=1} NP_{n,l}|h_{n,l}|^2}{N_o}\right) \tag{6.2}$$

subject to $P_{n,l} \geq 0$ and $(1/L)\sum_{l=0}^{L} P_{n,l} = P$ for all n. Here $P_{n,l}$ is the power allocated to the n^{th} user, $n = 1, \ldots, N$ and for the l^{th} subchannel, $l = 1, \ldots, L$. Assuming a symmetric uplink system, the optimal power allocation can be

relaxed and written as,

$$\frac{1}{L}\sum_{l=0}^{L}\sum_{n=0}^{N} P_{n,l} = NP. \tag{6.3}$$

Therefore, by assuming a total power constraint on all the users, the sum capacity can be expressed as [15],

$$C_{\text{sum}} = \mathrm{E}\left[\log\left(1 + \frac{P_{n^*}(\mathbf{h})|h_{n^*}|^2}{N_o}\right)\right] \tag{6.4}$$

where $\mathbf{h} := (h_1, \ldots, h_N)$, n^* is the index of the user with the strongest channel for a given $\mathbf{h}$, and

$$P_{n^*}(\mathbf{h}) = \begin{cases} \left(\frac{1}{\lambda} - \frac{N_o}{\max_i |h_i|^2}\right)^+ & \text{if} \quad |h_n|^2 = \max_i |h_i|^2 \\ 0 & \text{else} \end{cases} . \tag{6.5}$$

Though this result is derived assuming a total power constraint on all the users, by symmetry, the power consumption of all the users is the same under the optimal solution. Therefore the individual power constraints in (6.1) are automatically satisfied. Thereby, the original problem is solved.

C_{sum} in the symmetric case where users have identical channel statistics and power constraints has been considered. Even for the asymmetric case, the optimal strategy to achieve sum capacity is to have only the best user transmitting at a time, but the criterion of choosing the best user is different. Maximizing the sum rate may not be the appropriate objective since the user with the statistically better channel may get a much higher rate at the expense of the other users [15]. A more fair scheduling algorithm called "proportionally fair scheduling" is also presented in [15].

Next, looking at the downlink system, the received signal at the n^{th} user can be expressed as,

$$y_n[m] = h_n[m]x[m] + w_n[m] \tag{6.6}$$

where $h_n[m]$ is the channel of user n. The transmit power constraint is P and $w_n[m] \sim \mathcal{CN}(0, N_o)$ is i.i.d. in time m for each user. Here it is assumed that there is one base station (BS) transmitting to all users.

Consider the case when the users can track the channel but the BS only has statistical knowledge of the channel and not the exact channel realizations. The fading statistics are assumed to be symmetric to all users and by the assumption of ergodicity, if user n can decode its data reliably, then all the other users can also successfully decode user n's data. Thereby the sum rate can be expressed as,

$$\sum_{n=1}^{N} R_n < \mathrm{E} \log \left(1 + \frac{P|h_n|^2}{N_o} \right). \tag{6.7}$$

The bound above is achievable by transmitting to one user only or by time-sharing between any number of users. Thus in the symmetric fading channel, we obtain the same conclusion as in the symmetric AWGN downlink: the rate pairs in the capacity region can be achieved by both orthogonalization schemes and superposition coding.

For downlink AWGN channels superposition coding was successfully applied as there is an ordering of the channel strength of the users from weak to strong. In the case of asymmetric downlink fading channels, users in general

have different fading distributions and there is no longer a complete ordering of the users. In this case, we say that the downlink channel is non-degraded and little is known about good strategies for communication.

Next, consider the case when the base station also knows the channel realization and can thereby vary the powers as a function of the channel. In a downlink system where the base station tracks all the channels of the users, the metric of interest is still the sum capacity. Without fading, the sum capacity is achieved by transmitting only to the best user [15]. For fading channels, the best user in each coherence period is picked and an appropriate power is allocated to this user, subject to a constraint on the average power. Under this strategy, the downlink channel reduces to a point-to-point channel with the channel gain distributed as $\max_{n=1,\ldots,N} |h_n|^2$. The sum capacity for this case with optimal power allocation can be expressed as [15],

$$C_{\text{sum}} = \mathrm{E}\left[\log\left(1 + \frac{P_{n^*}(\mathbf{h})(\max_{n=1,\ldots,} |h_n|^2)}{N_o}\right)\right], \tag{6.8}$$

where $\mathbf{h} = (h_1, \ldots, h_N)$ is the joint fading state, and

$$P_{n^*}(\mathbf{h}) = \left(\frac{1}{\lambda} - \frac{N_o}{\max_{n=1,\ldots,N} |h_n|^2}\right)^+ \tag{6.9}$$

Next, we provide a brief literature survey of related work. The greedy scheme mentioned above assumes the ideal situation where fading statistics are the same for all users. However, in practical systems, the users' statistics are not identical, which would result in assigning the channel mostly to the statistically stronger user, and would be highly unfair. A fair strategy where

the channel is assigned to the user with the greatest instantaneous normalized signal-to-noise ratio (SNR) is also proposed [15, chap. 6]. The users are allowed to have different average channel powers in this setting. Using the fair channel assignment, the probability of being assigned to use the channel is the same for each user, even when each user has a different average SNR. A similar fair assigning algorithm which employs normalization with the time averaged throughput is proposed in [58]. Asymptotic analysis of this proportional fair algorithm is presented in [61]. In [62], by applying a first-order Markov chain model, the authors demonstrate the multiuser diversity-mobility tradeoff for the proportional fair scheduling proposed by [58].

For single user systems, optimal and practical adaptive modulation techniques have been well established. Several adaptation strategies are proposed and their spectral efficiencies are compared in [63] and [64]. Assuming perfect channel quality feedback, a constant-power, rate-adaptive scheme to maximize the average rate with an average bit error rate (BER) requirement is investigated in [65]. Good performance of adaptive modulation requires accurate channel estimation at the receiver and a reliable feedback path between the receiver and the transmitter. However, the channel feedback information will become outdated if the channel is changing rapidly. The uncertainty in channel feedback arises from the feedback delay. The effect of channel feedback delay on the average BER over Nakagami fading channels is briefly addressed in [66]. Under an instantaneous BER constraint, an approach to the design

of a robust adaptive modulation system based on only a single (most recent) outdated channel estimate is proposed in [67]. Adaptive modulation systems utilizing better channel prediction by combining a number of past channel estimates are investigated in [68], which further improve the performance of link adaptation. In [69], adaptive modulation schemes based on long-range channel prediction are presented.

When the dynamic range of the channel fluctuations are limited in environments with little scattering and/or slow fading, the authors in [58] propose a scheme called *opportunistic beamforming* to induce large and fast channel fluctuations artificially by using multiple transmit antennas. Further, in [70] the authors assume a multiuser frequency selective channel implementing OFDM where subcarriers are jointly allocated to users with favorable fading channels on these subcarriers.

In [71] multiuser diversity is studied in MIMO systems with linear receivers. The impact of adding transmitter or receiver antennas to multiuser diversity was further investigated in [72–74]. Correlated fading with multiuser diversity was analyzed in [58]. The authors compared an uncorrelated fading multiuser system with a fully correlated fading system when multiuser diversity is used for a network with a multi-antenna base station and single antenna users and showed that the fully correlated fading case has a higher network capacity. When perfect channel state information is not available in multiuser MIMO system, random beamforming was proposed to attain the multiuser

diversity gain with limited channel knowledge [75, 76].

Providing quality of service (QoS) to meet the data rate and the packet delay constraints of real-time data users is an important requirement of emerging wireless networks. To address this issue, channel aware scheduling algorithms were proposed. Being designed to satisfy the constraints of delay and data rate, these algorithms also improve the throughput performance by exploiting the channel fluctuations. Perhaps the most well-known channel aware scheduling algorithm is proportional fair scheduling algorithm used in CDMA 1xEVDO system [77], which maximizes the sum of logarithms of all user's average throughput. However, it is worth to know that the stability of the proportional fair scheduling algorithm was challenged in [78]. The authors in [79] developed another scheme, where the fraction of the time slots allocated to each user can be arbitrarily chosen based on the service objectives. A scheduling algorithm exploiting asynchronous variations of channel quality was proposed in [80]. The performance of channel aware scheduling algorithm was analyzed in [81]. Many approaches to exploit multiuser diversity require a centralized scheduler with full or limited knowledge of each user's channel state information. However, the scheme to acquire the estimates of the channel may incur excess overhead and delay, particularly if the number of active users is large or the channels change rapidly. Therefore, it is of interest to study the way to exploit the channel fluctuations by distributed channel access like random access, which may reduce overhead. The research of multiuser diversity

in random access is also driven by the proliferation of wireless LAN [24], where random access such as CSMA/CA is the primary means to provide channel access. To exploit the channel variations, many works of this kind assume decentralized channel state information where the base station broadcasts a pilot signal to all users, and each user measure its own channel using this pilot signal. The authors in [82, 83] proposed the use of channel state information to vary the transmission probability in ALOHA. They analyzed the system, and demonstrated the effect of multiuser diversity on throughput.

7. MULTI-USER DIVERSITY WITH RANDOM NUMBER OF USERS

7.1. Introduction

Point to point diversity combining schemes aim to mitigate the effects of fading in a wireless channel. Contrary to this, for multi-user systems the authors in [15] have indicated another form of diversity titled *multi-user diversity* (MUD), which thrives on the randomness of the user fading channel. The key idea here is to provide channel access to the user with the best channel at any instant of time. This has been shown to be optimal for both uplink [59] and downlink [60] scenarios.

In the literature so far MUD has been studied for the case of a fixed number of users only. While the assumption of fixed number of users holds for the analysis of instantaneous capacity and error rate, it is inaccurate for calculating time averages of the above mentioned metric. This is because in reality the number of users change rapidly with time. For example, the probability of a cell phone user requesting data communication is very low [22]. Further, certain types of data requests such as stocks, weather and email are bursty in nature. This implies that the number of users actively contending for channel access across time is a random variable.

Additionally, schemes in which a user is allowed to feedback its channel metric to request channel access, only if its metric is larger than a predefined threshold are also proposed [84–86]. Such algorithms also lead to a random number of users across time.

Based on the above reasons, in this work we study the effects of randomness of the number of users on system performance. We would like to emphasize here that in the multi-user system we consider, the base station does not know the number of users apriori but based on the received packets schedules the best user at any given time. Also, this work focuses on the system performance averaged across time and users but not the performance of an individual user.

Here we list the novel contributions of this work. For the first time in literature, analysis of MUD is considered when the number of users is random. The distribution of SNR of the best user chosen from a random set of users is derived for very general conditions. For any given user channel fading distribution, we prove that the BER of the system with a fixed number of users is a completely monotonic function of the number of users N. Further, we also prove that the capacity of the system implementing MUD with a fixed number of users has a completely monotone derivative. Using these results and the Jensen's inequality, we illustrate that randomization of the number of users always leads to deterioration of performance. Further, the user distributions can be Laplace transform ordered with respect to the BER and capacity of the system. For when the user distribution is Poisson distributed, the tightness of Jensen's inequality for BER is proved asymptotically in the number of users. We also derive formulas for outage expressions for any user fading distribution. For the special case when the number of users is Poisson distributed and

when the user channel is Rayleigh faded, a closed form expression for the BER is derived. Expressions for asymptotic capacity and outage capacity are also provided.

7.2. System Model

We consider the uplink multi-user system with one base station (BS) and multiple users. Both the BS and the users are assumed to be single antenna systems. The received signal at the BS from the n^{th} user can be expressed as,

$$y_n = \sqrt{\rho} h_n x_n + w_n, \qquad n = 1, \ldots, N. \tag{7.1}$$

The number of users N is assumed to be a random variable with a discrete non negative integer distribution. A homogeneous multi-user system is assumed where the average received power at the BS, ρ, in (7.1), is identical across all users. h_n is the channel, x_n is the transmitted symbol, and w_n is the additive white Gaussian noise corresponding to the n^{th} user, $n \in \{1, \ldots, N\}$. The channel is assumed to satisfy $\mathrm{E}[|h_n|^2] = 1$ for all n and to be independent and identically distributed (i.i.d.) across all users. The transmitted symbols satisfies $\mathrm{E}[|x_n|^2] = 1$.

7.3. SNR at the Base Station

The instantaneous SNR of all users is evaluated at the BS and the user with the best instantaneous SNR is selected for transmission. Unlike the existing

literature on multi-user diversity, the novelty of this work lies in assumption that the number of users N in the system is a random variable. Additionally, N is unknown at the BS. The BS station only selects the best user but does not feedback the quality of the channel to the user, thereby there is no rate adaptation at the user.

The instantaneous SNR of the n^{th} user at the BS, prior to selection, can be expressed as

$$\rho_n = \rho|h_n|^2, \qquad n = 1, \ldots, N \tag{7.2}$$

The best user is selected to have,

$$\rho^* = \rho|h_*|^2 \tag{7.3}$$

where $|h_*|^2| = \max_n\{|h_n|^2\}$.

Define $F_{\rho_n}(x)$ as the CDF of the instantaneous SNR of the n^{th} user ρ_n. Recalling that the total number of users N is a random variable, the CDF of the SNR of the best user selected using (7.3), conditioned on N, can be written as,

$$F_{\rho^*|N}(x) = \left[F_{\rho_n}(x)\right]^N, \tag{7.4}$$

which is obtained due to the i.i.d. assumption of the N user channels. The SNR of the best user selected from a random set of users can be obtained by

averaging (7.4) with respect to the distribution of N, i.e.,

$$G(F_{\rho_n}(x)) = F_{\rho^*}(x) = \mathrm{E}_N\left[\left[F_{\rho_n}(x)\right]^N\right] = \sum_{k=0}^{\infty} \Pr(N=k)\left[F_{\rho_n}(x)\right]^k, \tag{7.5}$$

which follows due to the assumption that N is from a discrete non negative integer distribution. Further, $G(F_{\rho_n}(x))$ is recognized to be the probability generating function of the discrete random variable N.

From (7.5) it can be seen that for any channel fading distribution and for any non-negative integer distribution on the number of users, the CDF of the SNR of the best user at the BS can be easily obtained. Further, using (7.5), the PDF of the SNR of the best user can be expressed as,

$$g(F_{\rho_n}(x)) = \sum_{k=0}^{\infty} \Pr(N=k) k \left[F_{\rho_n}(x)\right]^{k-1} f_{\rho_n}(x) \tag{7.6}$$

where $f_{\rho_n}(x) = dF_{\rho_n}(x)/dx$.

In the next section, we analyze the properties of the BER and capacity of a system with a fixed number of users implementing MUD. Using these results we characterize the behavior of the multi-user system with a random number of users implementing MUD.

7.4. Characteristics of the BER and Capacity

7.4.1. Error Rate

In this section, we first prove that the average BER of a multi-user system implementing MUD on a fixed number of users N is a completely monotonic

function of N, for very general conditions.

The BER of a multi-user system with a fixed number of users N can be expressed as,

$$\overline{\text{BER}}(\rho, N) = \int_0^\infty \text{BER}(x) d\left[F_{\rho_n}(x)\right]^N \tag{7.7}$$

where $\text{BER}(x)$ is the instantaneous BER over an additive white Gaussian noise (AWGN) channel for an instantaneous SNR x of the best user. Prior to proving that $\overline{\text{BER}}(\rho, N)$ in (7.7) is completely monotonic, we define completely monotonic functions.

Definition 7.4.1. *[87] A real function $f(x)$ is completely monotonic in argument x on the interval $(0, \infty)$, if it is infinitely differentiable on $(0, \infty)$ and for all $k \geq 0$,*

$$(-1)^k f^{(k)}(x) \geq 0, \qquad \textit{for all } \; x \in (0, \infty) \tag{7.8}$$

where $f^{(k)}(x)$ denotes the k^{th} derivative of $f(x)$. If in addition $f(0+) < \infty$, then $f(x)$ is completely monotone on $[0, \infty)$ and is represented as $f \in c.m.[0, \infty)$.

Though N is discrete, assuming that $\overline{\text{BER}}(\rho, N)$ is a function of a continuous N we prove that it is a completely monotone function of N. For discrete N a complete monotonicity definition similar to Definition. 7.4.1 exists in terms of successive differences of the function [88, pg. 118-119].

We choose to use the continuous version of the definition as it is relatively easier to prove that $\overline{\text{BER}}(\rho, N)$ is a completely monotone function using

derivatives. Further, the discrete version can be obtained by sampling its continuous counterpart, thereby validating our approach.

Integrating the $\overline{\text{BER}}(\rho, N)$ expression in (7.7) by parts, we have,

$$\overline{\overline{\text{BER}}}(\rho, N) = -\int_0^\infty \text{BER}'(x) \left[F_{\rho_n}(x)\right]^N dx, \tag{7.9}$$

where $\text{BER}'(x) = d\text{BER}(x)/dx$. Since $\text{BER}(x)$ is a decreasing function of x, $\text{BER}'(x)$ is negative. This implies that $\overline{\overline{\text{BER}}}(\rho, N)$ is still positive.

Using (7.9), the k^{th} derivative of $\overline{\text{BER}}(\rho, N)$ can be written as,

$$\frac{d^{(k)}\overline{\overline{\text{BER}}}(\rho, N)}{dN^{(k)}} = -\int_0^\infty \text{BER}'(x) \left[F_{\rho_n}(x)\right]^N \left[\log\left(F_{\rho_n}(x)\right)\right]^k dx. \tag{7.10}$$

In (7.10), since $0 \leq F_{\rho_n}(x) \leq 1$, this implies that $\log\left(F_{\rho_n}(x)\right) \leq 0$, and thereby $\left[\log\left(F_{\rho_n}(x)\right)\right]^k$ switches sign with increasing k. Therefore, we can write,

$$(-1)^k \frac{d^{(k)}\overline{\overline{\text{BER}}}(\rho, N)}{dN^{(k)}} \geq 0, \qquad for \;\; k \geq 0. \tag{7.11}$$

Further, since $\overline{\overline{\text{BER}}}(\rho, N)$ is infinitely differentiable with respect to N, and that $\overline{\overline{\text{BER}}}(\rho, 0^+) < \infty$, we have proved that $\overline{\overline{\text{BER}}}(\rho, N) \in c.m.[0, \infty)$.

For $k = 2$, the condition in (7.8) reduces to $f^{(2)}(x) \geq 0$, for all $x \in [0, \infty)$, which is the necessary condition to establish the convexity of $f(x)$. This implies that all completely monotonic functions are convex functions as well but the other way is not always true. From this result, it immediately follows that $\overline{\overline{\text{BER}}}(\rho, N)$ is a convex function in N.

For when the number of users in the system is random, by applying Jensen's inequality for convex functions, we have,

$$\mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right] \geq \overline{\mathrm{BER}}(\rho, \lambda), \tag{7.12}$$

where $\lambda = \mathrm{E[N]}$. With this result we have established that, immaterial of the distribution of the number of users N, as long as N is random, the BER averaged across the distribution of N will never be better than the BER of a system with a fixed number of users. In short, randomization of N will always hurt the BER performance of a multi-user system.

7.4.2. Capacity

Now, we focus on the properties of the capacity of a multi-user system with a fixed number of users implementing MUD. The capacity for the fixed number of users system can be expressed as,

$$\begin{aligned} C(\rho, N) &= \int_0^\infty \log_2(1+x)\, d\left[F_{\rho_n}(x)\right]^N \\ &= \frac{1}{\log(2)} \int_0^\infty \frac{1-\left[F_{\rho_n}(x)\right]^N}{1+x} dx \end{aligned} \tag{7.13}$$

where the latter expression is obtained from the former by performing integration by parts, and by assuming $\lim_{x\to 0} \log(1+x)(1-[F_{\rho_n}(x)]^N) = 0$ and $\lim_{x\to\infty} \log(1+x)(1-[F_{\rho_n}(x)]^N) = 0$.

By taking the derivative of $C(\rho, N)$ in (7.13) with respect to N, we have,

$$\frac{dC(\rho, N)}{dN} = -\frac{1}{\log(2)} \int_0^\infty \frac{\left[F_{\rho_n}(x)\right]^N \log\left(F_{\rho_n}(x)\right)}{1+x} dx. \tag{7.14}$$

Since $\log\left(F_{\rho_n}(x)\right) \leq 0$, the first derivative of $C(\rho, N)$ in (7.14) is positive thereby implying that $C(\rho, N)$ is an increasing function of N. This implies that $C(\rho, N)$ is not a completely monotonic function of N.

Defining $C_d(\rho, N) := dC(\rho, N)/dN$, it can be seen that $C_d(\rho, N)$ is a completely monotonic function since,

$$C_d^{(k)}(\rho, N) = -\frac{1}{\log(2)} \int_0^\infty \frac{\left[F_{\rho_n}(x)\right]^N \left[\log\left[F_{\rho_n}(x)\right]\right]^{k+1}}{1+x} dx. \tag{7.15}$$

In the above expression, $C_d^{(k)}(\rho, N)$ switches sign with increasing k thereby satisfying the conditions required for complete monotonicity (7.8). This further implies that $C(\rho, N)$ has a completely monotonic derivative, immaterial of the fading distribution. Once again, we would like to state here that for the above result to hold, we have assumed $C(\rho, N)$ to be a function of continuous N, but the results hold even for discrete N.

The first property we would like to point out here is that since the second derivative of $C(\rho, N)$ is negative, $C(\rho, N)$ is concave function of N. Applying Jensen's inequality for concave functions, we have

$$\mathrm{E}_N\left[C(\rho, N)\right] \leq C(\rho, \lambda). \tag{7.16}$$

This result implies that immaterial of the distribution of the number of users N, randomization of N will always hurt the capacity of a multi-user system. The completely monotone derivative property of $C(\rho, N)$ will be used in the following section discussing the Laplace transform ordering of random variables.

7.5. Laplace Transform Ordering

In this section we introduce Laplace transform (LT) ordering, a method to compare the effect that different distributions of the number of users has on the average BER and capacity.

Stochastic ordering of random variables is a branch of probability theory and statistics which deals with partial ordering of random variables with different distributions [89, 90]. There are several different orderings, such as the likelihood ratio ordering, mean residual ordering, and Laplace transform ordering. Of these, we choose LT ordering for reasons which will be evident soon.

Definition 7.5.1. *: [90, pg. 39] Let X and Y be real random variables. X is said to be less than Y in LT order (written $X \leq_{Lt} Y$), if the LTs exist and satisfy, $E\left[e^{-sX}\right] \geq E\left[e^{-sY}\right]$ for all $s > 0$.*

The existence of the Laplace transform is trivial if X and Y are non-negative, which is so for our case. Another important theorem that is relevant to our work is [89, pg. 96] mentioned next,

Theorem 7. *Let X and Y be two random variables. If $X \leq_{Lt} Y$, then, $E[\psi(X)] \geq E[\psi(Y)]$ for all completely monotone functions $\psi(\cdot)$, provided the expectation exists.*

Corollary 2. *[88, Corollary 3.2] Let X and Y be two random variables. If $X \leq_{Lt} Y$, then, $E[\psi(X)] \leq E[\psi(Y)]$ for all $\psi(\cdot)$ with completely monotone derivative, provided the expectation exists.*

Further in [90, pg. 61] it is shown that likelihood ratio (Lr) ordering implies LT ordering and in [90, pg. 63] the necessary conditions for certain discrete distributions to be likelihood ratio ordered are provided.

From this we can infer that if X and Y are both Poisson distributed with means λ and μ respectively, such that $\lambda \leq \mu$, then $X \leq_{Lr} Y$, which implies that $X \leq_{Lt} Y$. Similarly, if X and Y are geometric distributed with probability of success on each trail p and q respectively, such that $p \leq q$, then $X \leq_{Lr} Y$, which implies that $X \leq_{Lt} Y$.

From (7.10) recalling that the $\overline{\mathrm{BER}}(\rho, N)$ is a completely monotone function and that $C(\rho, N)$ has a completely monotone derivative, it follows that if two random variables N_1 and N_2 are LT ordered, i.e., $N_1 \leq_{Lt} N_2$ then,

$$\begin{aligned} \mathrm{E}_{N_1}\left[\overline{\mathrm{BER}}(\rho, N_1)\right] &\geq \mathrm{E}_{N_2}\left[\overline{\mathrm{BER}}(\rho, N_2)\right] \\ \mathrm{E}_{N_1}\left[C(\rho, N_1)\right] &\leq \mathrm{E}_{N_2}\left[C(\rho, N_2)\right] \end{aligned} \tag{7.17}$$

7.6. Poisson Distributed N

Consider a multi-user system which contains a large number of users. Each user will be with a very small probability independent of the number of users in the system or the activity of the other users. In such a system, as the

number of users present increases, the probability that N users are at a given instant will be Poisson distributed. Thereby, in this section we analyze the system when N is Poisson distributed with parameter λ.

7.6.1. Outage

For extremely slow fading or non-ergodic fading channels, the probability of outage is an important performance metric. The CDF of the SNR of the best user chosen from a random set of users given in (7.5) represents the generic outage probability.

Given N is Poisson distributed with parameter λ, the probability of outage for a given minimum SNR ρ_m can be expressed as,

$$\begin{aligned}
\Pr(\rho^* \leq \rho_m) &= G\left(F_{\rho_n}(\rho_m)\right) \\
&= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \left[F_{\rho_n}(\rho_m)\right]^k \\
&= e^{-\lambda} e^{\lambda F_{\rho_n}(\rho_m)} \\
&= e^{-\lambda(1-F_{\rho_n}(\rho_m))} \qquad (7.18)
\end{aligned}$$

From this expression it can be seen that as λ increases, $\Pr(\rho^* \leq \rho_m)$ decreases. This implies that as the average number of users increases, the outage probability decreases for any distribution of the user fading channel. This once again highlights the benefit of implementing MUD.

Next we would like to illustrate the convergence of the CDF in (7.18) to one of the three limiting distributions asymptotically with λ. First, we would like to mention a theorem from [91],

Theorem 8. *Define* $W_n = \max\{x_1, \ldots, x_n\}$ *where* $x_1, \ldots, x_n$ *are i.i.d. random variables. Let* $a_n > 0$ *and* b_n *be suitably chosen constants such that, as* $n \to \infty$,

$$F^n(a_n x + b_n) \to S(x) \tag{7.19}$$

at all continuity points of S, *which means that* F *is in the domain of maximal attraction of the limiting distribution* S. *A necessary condition for the above convergence, as* $n \to \infty$, *is given by, [91, 92],*

$$n\left[1 - F(a_n x + b_n)\right] \to w(x). \tag{7.20}$$

Depending on the properties of $F(x)$, $S(x) = e^{w(x)}$ *can be one of only three forms, where*

$$\begin{aligned} w(x) &= x^{-h} \qquad \textit{for} \quad x > 0 \quad \textit{or} \\ w(x) &= (-x)^{h} \qquad \textit{for} \quad x < 0 \quad \textit{or} \\ w(x) &= e^{-x} \qquad \textit{for all} \quad x. \end{aligned}$$

Here $h > 0$ *is a real number.*

Further details can be found in [91, 92]. Using the above theorem, the CDF of the SNR of the best user chosen from a random set of users, (7.5), (7.18), can be expressed for $\lambda \to \infty$ as,

$$\lim G(F_{\rho_n}(x)) = \lim e^{-\lambda(1-F_{\rho_n}(x))} = e^{-w((x-b_n)/a_n))}, \tag{7.21}$$

where $\lim \lambda(1 - F_{\rho_n}(x)) = w(x)$.

This implies that the extreme value distributions defined for the maximum of a large set of random variables can be used to express the distribution of the SNR of the best user chosen from a random set of users. Thus, as $\lambda \to \infty$,

$$\Pr(\rho^* \leq \rho_m) = e^{-w((\rho_m - b_n)/a_n)} \tag{7.22}$$

is defined by the properties of $F_{\rho_n}(x)$.

7.6.2. BER

In this subsection, we provide the necessary conditions for the Jensen's inequality for $\overline{\mathrm{BER}}(\rho, N)$ in (7.17) to be tight. To this end we use the results put forth by the author in [93]. We make use of the following theorem proved in their work,

Theorem 9. *Let $f \in c.m.[0, \infty)$. Define the average,*

$$I(\lambda) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} f(k)$$

so that $I(\lambda) = E_N[f(N)]$ where N is a Poisson random variable with mean λ. Then,

- *$I(\lambda)$ is completely monotone, strictly decreasing in λ (unless f is constant) and $\forall\ \lambda \geq 0$, $I(\lambda) \geq f(\lambda)$,*

- *If f is c.m. $\in [0,\infty)$ and regularly varying, i.e., $f \in \mathbb{R}_\mu$ for some $\mu \leq 0$, then $I(\lambda)$ is both c.m. and $\mathbb{R}_\mu$, and*

$$I(\lambda) = f(\lambda) + O(f(\lambda)/\lambda) \qquad as \;\; \lambda \to \infty$$

where $f(x) = O(g(x))$ means $\lim_{x\to\infty} |f(x)/g(x)| \leq b$ for some constant $b > 0$, and $\mathbb{R}_\mu$ denotes that $\lim_{x\to\infty} f(\kappa x)/f(x) = \kappa^\mu$ holds for all $\kappa > 0$. For the sake of completeness, we define regularly varying functions next.

Definition 7.6.1. *A function $f : (0,\infty) \to \infty$ is regularly varying at infinity if for all $\kappa \geq 1$, the limit,*

$$\lim_{x\to\infty} \frac{f(\kappa x)}{f(x)} \tag{7.23}$$

exists and is in $(0,\infty)$.

Regularly varying functions are those that scale homogeneously for large arguments. Once again, we would like to state here that for our result we have assumed $\overline{\text{BER}}(\rho, N)$ is a function of a continuous N. Since the discrete version can be obtained by sampling the continuous version, the asymptotic convergence in the above Definition 7.6.1 also implies the convergence for the discrete case.

To utilize the above theorems we rewrite the $\overline{\text{BER}}(\rho, N)$ expression in (7.9) as,

$$\begin{aligned}\overline{\text{BER}}(\rho, N) &= -\int_0^\infty \text{BER}'(x) e^{N \log(F_{\rho_n}(x))} dx \\ &= \int_0^\infty B(x) e^{N \log(F_{\rho_n}(x))} dx \end{aligned} \tag{7.24}$$

where $B(x) = -\text{BER}'(x) = -d\text{BER}(x)/dx$. Now, setting $u := -\log(F_{\rho_n}(x))$, and integrating by substitution we have,

$$\overline{\text{BER}}(\rho, N) = \int_0^\infty \frac{B(F_{\rho_n}^{-1}(e^{-u}))e^{-u}e^{-uN}du}{f_{\rho_n}(F_{\rho_n}^{-1}(e^{-u}))}, \tag{7.25}$$

where $F_{\rho_n}^{-1}(x)$ is the inverse CDF function and $f_{\rho_n}(x)$ is the pdf of ρ_n. Before we proceed further, we present Berstein's theorem next,

Theorem 10. *If $f \in c.m.[a, \infty)$, this implies that*

$$f(x) = \int_0^\infty e^{-sx}d\psi(s), \qquad x \in [a, \infty), \tag{7.26}$$

where $\psi(s)$ is monotone non-decreasing and bounded on $[0, \infty)$ with $\psi(0) = 0$ and $\psi(\infty) = f(a+) < \infty$.

Following this, we quote Karamata's Tauberian theorem [94] [95, XIII.4, Theorem 2] that if f is given by Berstein representation as shown above, i.e., $f \in c.m.[a, \infty)$, then $f \in \mathbb{R}_\mu$ for some $\mu \leq 0$ if and only if $\psi(1/s) \in \mathbb{R}_{-\mu}$ (i.e., $\psi(s)$ is regularly varying as $s \to 0$).

In equation (7.25) we know that $\overline{\text{BER}}(\rho, N)$ is a completely monotone function as shown in Section 7.4.1. Therefore using the above mentioned Bernstein's representation of completely monotone functions and the Tauberian theorem, $\overline{\text{BER}}(\rho, N)$ can be shown to be regularly varying at infinity by showing that

$$\frac{B(F_{\rho_n}^{-1}(e^{-u}))}{f_{\rho_n}\left(F_{\rho_n}^{-1}(e^{-u})\right)}$$

is regularly varying at the origin, i.e., as $u \to 0$.

Using the results cited above from [93], if $\overline{\mathrm{BER}}(\rho, N)$ is completely monotonic and regularly varying,

$$\mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right] = \overline{\mathrm{BER}}(\rho, \lambda) + O\left(\overline{\mathrm{BER}}(\rho, \lambda)/\lambda\right) \qquad \lambda \to \infty \qquad (7.27)$$

holds. Else, if $\overline{\mathrm{BER}}(\rho, N)$ is completely monotonic only and not regularly varying then,

$$\mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right] = \overline{\mathrm{BER}}(\rho, \lambda) + O\left(\lambda\overline{\mathrm{BER}}''(\rho, \lambda)\right) \qquad \lambda \to \infty \qquad (7.28)$$

To give specific examples for regularly varying $\overline{\mathrm{BER}}(\rho, N)$ let us consider Rayleigh faded channels with parameter 1, and let $\mathrm{BER}(x)$ have the general form of e^{-x}. For this case, we have

$$\frac{B(F_{\rho_n}^{-1}(e^{-u}))e^{-u}}{f_{\rho_n}(F_{\rho_n}^{-1}(e^{-u}))} = (1 - e^{-u})^{\rho-1}e^{-u} = t(u).$$

From the above expression it is easy to verify that $\lim_{u\to 0} t(\kappa u)/t(u)$ exists and is finite, thereby proving the regular variation of $t(u)$ near its origin and in turn proving regular variation of $\overline{\mathrm{BER}}(\rho, N)$ at infinity.

Next, for the more general case of when $\mathrm{BER}(x) = Q(\sqrt{2x})$ is considered. For this case it can shown that,

$$\overline{\mathrm{BER}}(\rho, N) = \frac{\rho}{2\sqrt{2\pi}} \int_0^\infty e^{-Nu} \frac{(1 - e^{-u})^{\rho-1}e^{-u}}{\sqrt{-\rho \log(1 - e^{-u})}} du$$

where we have used the form $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-y^2/2} dy$ and the Leibniz integral rule is used to find the derivative of $Q(x)$.

Once again, we are interested in the behavior of

$$\frac{(1-e^{-u})^{\rho-1}e^{-u}}{\sqrt{-\rho\log(1-e^{-u})}}$$

term. This can be shown to be regularly varying as $u \to 0$ using (7.23).

Next the more general case of chi-square distributed channel fading with K degrees of freedom is considered. For this, $F_{\rho_n}(x) = \gamma(K/2, X/2\rho)/\Gamma(K/2)$ and $\mathrm{BER}(x) = e^{-x}$. The average BER can be written as,

$$\overline{\mathrm{BER}}(\rho, N) = \rho \int_0^\infty e^{-x} e^{N\log(\gamma(K/2,x/2\rho)/\Gamma(K/2))} dx$$

By setting $u = -\log\left(\gamma(K/2, x/2\rho)/\Gamma(K/2)\right)$ and integrating by parts, we have,

$$\overline{\mathrm{BER}}(\rho, N) = \frac{\rho\Gamma(K/2)}{2^{k/2}} \int_0^\infty \frac{e^{\rho(K-2)\log(Q(k)u^{2/K-2})} e^{-Nu} e^{-u}}{e^{(k/2-1)\log(Q(k)u^{2/K-2})} \left(-(k-2)\log\left(Q(k)u^{2/K-2}\right)\right)^{k/2-1}} du$$

To obtain the above expression, we need to solve $\gamma(K/2, x/2) = \Gamma(K/2)e^{-u}$ for x. The solution is obtained as,

$$x = -(K-2)\mathcal{W}\left(-\frac{2[(1-e^{-u})\Gamma(K/2)]^{2/K-2}}{K-2}\right),$$

where $\mathcal{W}(\cdot)$ is the Lambert-W function [55]. Since we are interested in the behavior of the arguments of the integral as $u \to 0$, this leads to the behavior of the Lambert-W function as its argument approaches 0 from the left. This in-turn leads to the Lambert-W function approaching $-\infty$ as $u \to 0$.

Thereby, using the non-principal branch of the Lambert-W function and the asymptotic representation $\mathcal{W}(x) \sim \log(x)$ as $x \to 0^-$ on the non-principal

branch, and the substitution $(1 - e^{-u}) \sim u$ as $u \to 0$,

$$x = -(K-2)\log\left(\Omega(K)u^{2/K-2}\right),$$

where $\Omega(K) = -2(\Gamma(K/2))^{2/K-2}/(K-2)$. Putting all of this together we can express $\overline{\text{BER}}(\rho, N)$ as,

$$\overline{\text{BER}}(\rho, N) = \frac{\rho\Gamma(K/2)}{2^{K/2}} \int_0^\infty \frac{e^{\rho(K-2)\log(\Omega(K)u^{2/K-2})} e^{-Nu} e^{-u}}{e^{K/2-1}\log(\Omega(K)u^{2/K-2})\left(-(K-2)\log(\Omega(K)u^{2/K-2})\right)^{K/2-1}} du.$$

In the above expression we are interested in the behavior of,

$$u^{(2/K-2)(\rho(K-2)-(K/2-1))} \frac{(\Omega(K))^{\rho(K-2)}}{(\Omega(K))^{K/2-1}\left(-(K-2)(\log(\Omega(K)u^{2/K-2}))\right)^{K/2-1}}.$$

Once again, it is straightforward to check that the above function of u is regularly varying as $u \to 0$, thereby implying that $\overline{\text{BER}}(\rho, N)$ is regularly varying as $N \to \infty$.

In the following section, we consider a specific distribution for N and ρ_n and derive a closed form BER expression and asymptotic capacity for a system implementing MUD with a random number of users.

7.7. A Special Case: Poisson distributed N and Rayleigh Faded Channels

In this section, the case when the number of users N is Poisson distributed with parameter λ and the user channels are Rayleigh faded with parameter 1 is considered.

For this case the CDF of the instantaneous SNR of the best user chosen from a random set of users in (7.5) can be expressed as,

$$G(F_{\rho_n}(x)) = e^{-\lambda e^{(-x/\rho)}} \qquad \text{for } \; x \geq 0. \tag{7.29}$$

From the above equation it is seen that the SNR of the best user is identical to a truncated and shifted Gumbel distribution [96]. Notice that for $x = 0$ (7.29) gives $e^{-\lambda}$. For $x > 0$, (7.29) has the form of the Gumbel distribution (one of the three limiting distributions) with $a_n = \rho$ and $b_n = \rho \log(\lambda)$ corresponding to the parameters in [92, pg. 296]. The distribution is therefore of mixed type, mass of $e^{-\lambda}$ at the origin and the rest of the distribution has the form of a *truncated Gumbel distribution*.

It is important to note here that unlike the usage of the Gumbel distribution in [97] or earlier in (7.22), in (7.29) the parameter λ is finite and not an asymptotic result in λ. This implies that the result is exact and not an approximation like in the other two cases.

The CDF expression in (7.29) includes the case when $N = 0$, i.e., there are no users in the system. When there are no users, no data will be transmitted in the system and thereby the capacity and BER are both zero. Including this case into the system performance evaluation would unfairly bias the performance. Therefore, the $N = 0$ case is dropped and the *zero-truncated Poisson distribution* is chosen for N. For zero-truncated Poisson distributed N, $N \in \{1, 2, \ldots\}$ and $\underline{\lambda} = \lambda/(1 - e^{-\lambda})$. The CDF of the instantaneous SNR

of the best user can be expressed as,

$$G(F_{\rho_n}(x)) = \frac{1}{1-e^{-\underline{\lambda}}} \sum_{k=1}^{\infty} \left[F_{\rho_n}(x)\right]^k \underline{\lambda}^k \frac{e^{-\underline{\lambda}}}{k!}$$
$$= \frac{e^{\underline{\lambda}(1-e^{(-x/\rho)})} - 1}{e^{\underline{\lambda}} - 1}. \tag{7.30}$$

The pdf can be expressed as,

$$f_{\rho_n}(x) = \frac{\underline{\lambda}}{\rho(1-e^{-\underline{\lambda}})} e^{-x/\rho} e^{-\underline{\lambda}e^{(-x/\rho)}}. \tag{7.31}$$

We assume the instantaneous BER is of the form,

$$\text{BER}(x) = \alpha e^{-\eta x}. \tag{7.32}$$

For binary differential phase-shift-keying (DPSK) (7.32) is exact with $\alpha = 0.5$ and $\eta = 1$. For Gray-coded M-level quadratic amplitude modulation (M-QAM), $\alpha = 0.2$ and $\eta = 1.5/(M-1)$ yield a BER within 1 dB for $M \geq 4$ [5]. Further using the Craig's formula,

$$Q(\sqrt{x}) = \frac{1}{\pi} \int_0^{\pi/2} \text{BER}\left(\frac{x}{2\sin^2(\phi)}\right) d\phi \tag{7.33}$$

with $\alpha = 1$ and $\eta = 1$ can be used for modeling BER of other modulation schemes as well, which would lead to an additional integral to our forthcoming results. Therefore, the average BER achieved by the system across time, can be expressed as,

$$\text{E}_N\left[\overline{\text{BER}}(\rho, N)\right] = \frac{\underline{\lambda}\alpha}{\rho(1-e^{-\underline{\lambda}})} \int_0^{\infty} e^{-\eta x} e^{-x/\rho} e^{-\underline{\lambda}e^{-x/\rho}} dx. \tag{7.34}$$

Setting $y = \underline{\lambda} e^{-x/\rho}$ and integrating by substitution, (7.34) can be expressed as,

$$\begin{aligned} \mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right] &= \frac{\alpha}{1 - e^{-\underline{\lambda}}} \int_0^{\underline{\lambda}} \left(\frac{y}{\underline{\lambda}}\right)^{\eta\rho} e^{-y} dy \\ &= \frac{\alpha \underline{\lambda}^{-\eta\rho}}{1 - e^{-\underline{\lambda}}} \gamma(\eta\rho + 1, \underline{\lambda}) \\ &\leq \frac{\alpha \underline{\lambda}^{-\eta\rho}}{1 - e^{-\underline{\lambda}}} \Gamma(\eta\rho + 1), \end{aligned} \tag{7.35}$$

where $\gamma(s, x)$ is the lower incomplete gamma function [98] which has been upper bounded by the complete gamma function $\Gamma(s)$ [98]. Additionally, the average BER depends on $\underline{\lambda}$, which defines the average number of users in the system. As $\underline{\lambda}$ increases, the BER of the system improves thereby highlighting the benefit of implementing MUD once again.

To find the diversity order achievable by such a system, consider the expression in (7.35). Using the series expansion of the lower incomplete gamma function [98], i.e.,

$$\gamma(s, x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{x^{s+k}}{s+k}, \tag{7.36}$$

we can rewrite (7.35) as,

$$\begin{aligned} \mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right] &= \frac{\alpha \underline{\lambda}^{-\eta\rho}}{1 - e^{-\underline{\lambda}}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\underline{\lambda}^{\eta\rho+k+1}}{\eta\rho + k + 1}, \\ &= \frac{\alpha}{\rho(1 - e^{-\underline{\lambda}})} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\underline{\lambda}^{k+1}}{\eta + (k+1)/\rho}. \end{aligned} \tag{7.37}$$

Defining the limit,

$$\lim_{\rho \to \infty} \frac{\mathrm{E}_N\left[\overline{\mathrm{BER}}(\rho, N)\right]}{1/\rho} = \frac{\alpha \underline{\lambda}}{\eta(e^{\underline{\lambda}} - 1)}, \tag{7.38}$$

which implies that the achievable diversity order will be limited to 1. Contrary to this, it has been shown in [97] that for a system with a fixed number of users N, a diversity order of N can be achieved. Due to this difference in achievable diversity order, at high SNR, the system with a fixed number of users would always outperform a system with a random number of users significantly, for any fixed finite $\underline{\lambda}$.

In Section 7.6.2, the tightness of Jensen's inequality was shown for the specific case of N Poisson distributed for any channel fading distribution. This implies that for a fixed SNR, as the total number of users N for the fixed case, and the average number of users $\underline{\lambda}$ increases, such that $N = \underline{\lambda}$, the BER performance of the two systems become almost identical. We will further illustrate these points in the Simulations section. Next, we derive the asymptotic average capacity and the corresponding scaling laws with respect to $\underline{\lambda}$. The capacity of the multi-user system with a random number of users, averaged across the user distribution can be expressed as,

$$\mathrm{E}_N\left[C(\rho, N)\right] = \mathrm{E}_{\rho^*}\left[\log_2(1+\rho^*)\right]. \tag{7.39}$$

For the zero-truncated Poisson distributed N and Rayleigh faded channels, the capacity of the system can be written as (7.13), [99],

$$\mathrm{E}_N\left[C(\rho, N)\right] = \frac{1}{\log(2)} \int_0^\infty \frac{1 - \frac{e^{-\underline{\lambda}e^{-x/\rho}} - e^{-\underline{\lambda}}}{1-e^{-\underline{\lambda}}}}{1+x} dx. \tag{7.40}$$

We can express (7.40) for $\underline{\lambda} \to \infty$ as,

$$\mathrm{E}_N\left[C(\rho, N)\right] \sim \frac{1}{\log(2)} \int_0^\infty \frac{1 - e^{-\underline{\lambda}e^{-x/\rho}}}{1+x} dx, \tag{7.41}$$

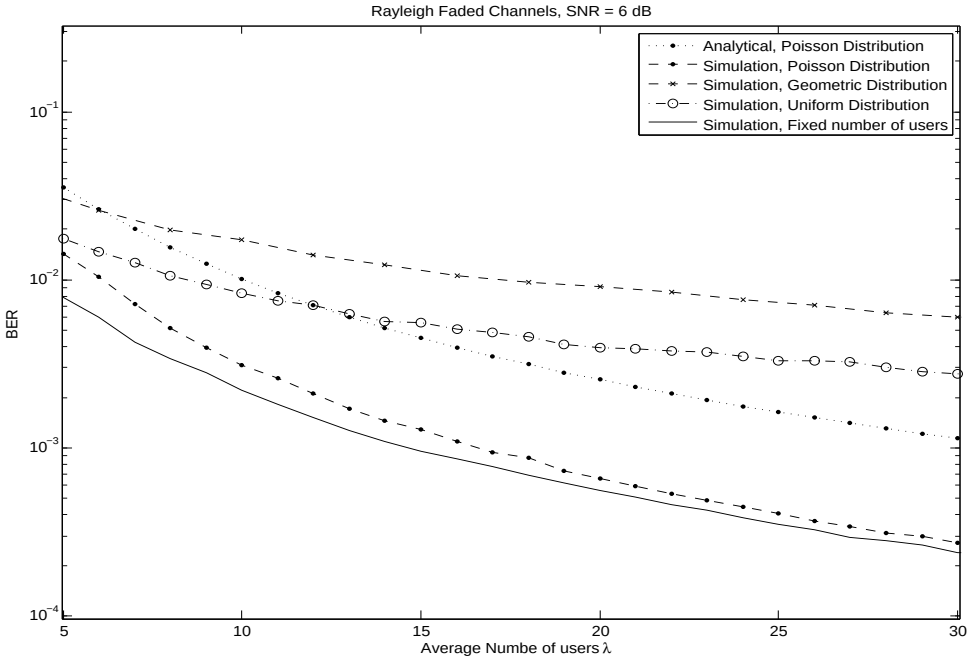

Fig. 22. BER vs. λ: Rayleigh Fading Channel, SNR = 6 dB

where $f(x) \sim g(x)$ implies that $f(x)/g(x) \to 1$ as $x \to \infty$. Defining $y := e^{-x/\rho}$ and integrating by substitution,

$$\begin{aligned} \mathrm{E}_N\left[C(\rho, N)\right] &\sim \frac{1}{\log(2)} \int_0^1 \frac{1 - e^{-\underline{\lambda} y}}{1 - \rho \log(y)} \left(\frac{\rho}{y}\right) dy \\ &\sim \frac{1}{\log(2)} \left[\int_0^{1/\underline{\lambda}} \frac{1 - e^{-\underline{\lambda} y}}{1 - \rho \log(y)} \left(\frac{\rho}{y}\right) dy + \int_{1/\underline{\lambda}}^1 \frac{\rho}{y(1 - \rho \log(y))} dy \right]. \end{aligned} \tag{7.42}$$

The limit of the first term in the integral in (7.42) becomes negligibly small because $\lim_{\underline{\lambda} \to \infty} 1/\underline{\lambda} = 0$, while the second term can be computed as,

$$\frac{1}{\log(2)} \int_{1/\underline{\lambda}}^1 \frac{\rho}{y(1 - \rho \log(y))} dy = \frac{1}{\log(2)} \log\left(1 + \rho \log(\underline{\lambda})\right).$$

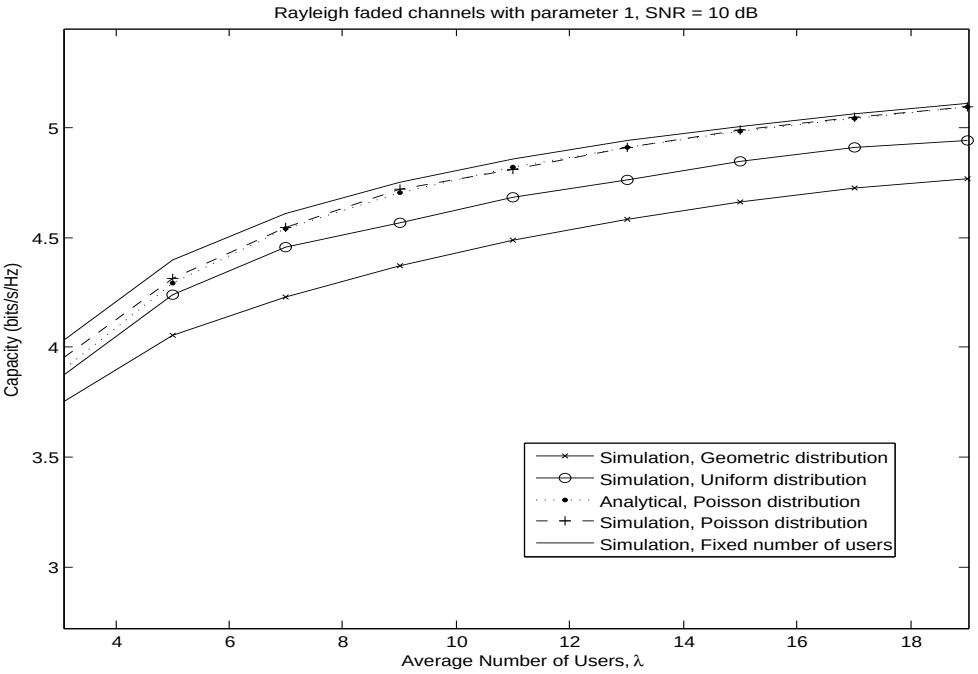

Fig. 23. Capacity vs. λ: Rayleigh Fading Channel, SNR = 10dB

This implies that the capacity of the system grows like $\log\log(\underline{\lambda})$, i.e.,

$$\lim_{\underline{\lambda}\to\infty} \frac{\mathrm{E}_N\left[C(\rho, N)\right]}{\log(\log(\underline{\lambda}))} = 1.$$

7.7.1. Outage Capacity

In this section, we consider the case when fading is non-ergodic. ϵ-Outage capacity is defined as the largest rate of transmission R such that the outage probability is less than ϵ; $\epsilon > 0$. To help compare the outage performance of a system implementing MUD with a random number of users, we first derive the outage capacity of a system implementing MUD with a fixed number of users N. By solving $\Pr(\log(1+\rho^*) \leq R) = \epsilon$ in (7.18), the outage capacity for

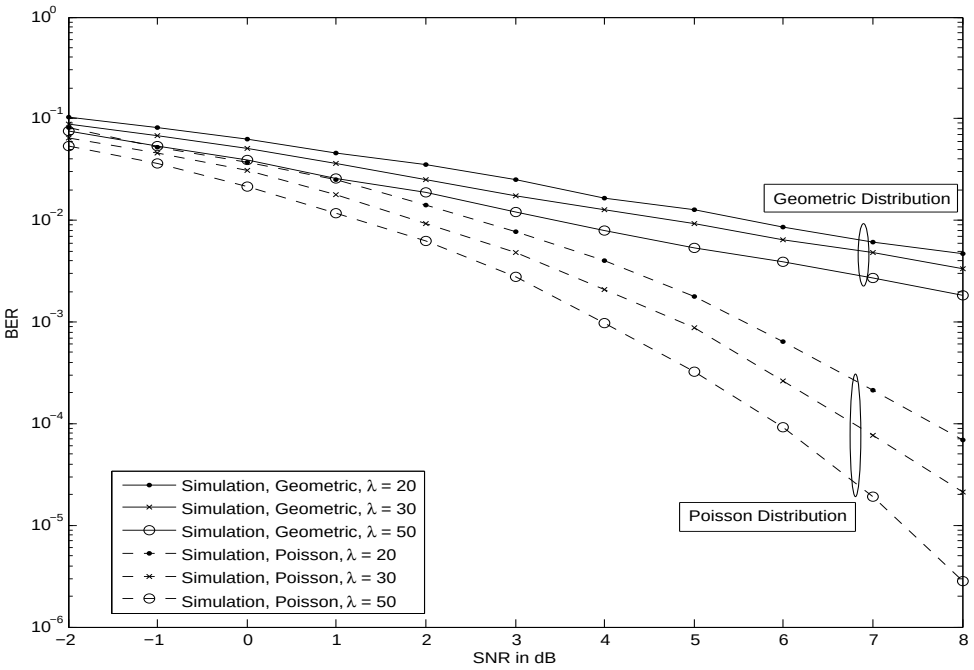

Fig. 24. BER vs. SNR: Rayleigh Fading Channel

the system with a fixed number of users can be expressed as,

$$C_{o,f}(\epsilon) = \log_2\left(1 + \rho F^{-1}\left(\epsilon^{1/N}\right)\right)$$

where $F(\cdot)$ is the CDF of the user fading channel $|h_n|^2$. For the case of Poisson-exponential distribution, $F^{-1}(x) = \log\left(1/(1-x)\right)$, we have

$$C_{o,f}(\epsilon) = \log_2\left(1 + \rho \log\left(\frac{1}{1-\epsilon^{1/N}}\right)\right). \tag{7.43}$$

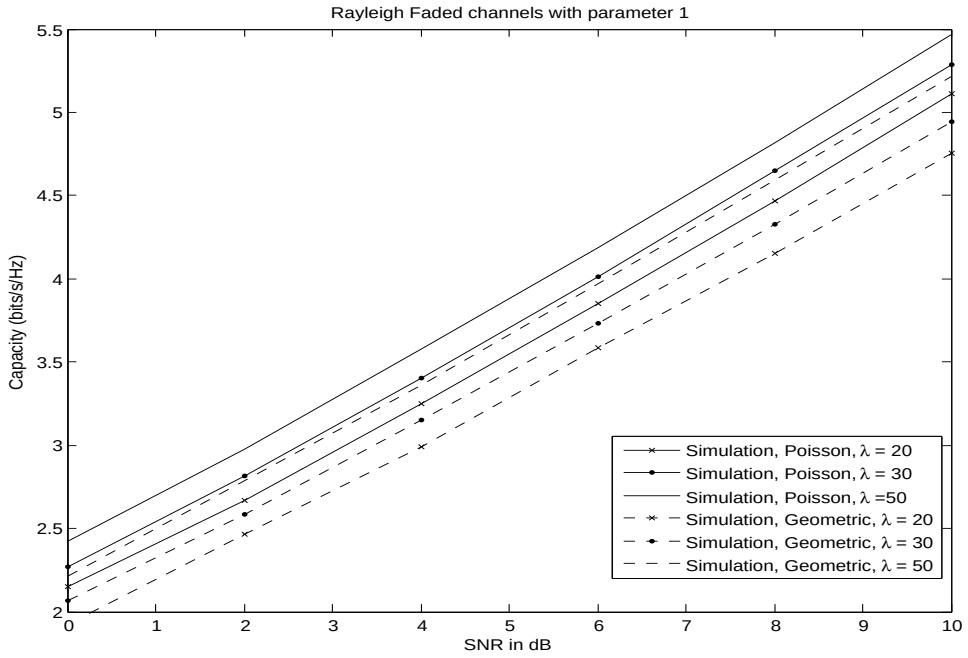

Fig. 25. Capacity vs. SNR: Rayleigh Fading Channel

For when the number of users is random, the outage capacity can be similarly expressed as

$$\begin{aligned} C_{o,r}(\epsilon) &= \log_2\left(1+\rho G^{-1}\left(\epsilon\right)\right) \\ &= \log_2\left(1+\rho\left(-\log\left(-\frac{\log\left(\epsilon\right)}{\underline{\lambda}}\right)\right)\right), \end{aligned} \tag{7.44}$$

where $G^{-1}(x) = -\log\left(-\log(x)/\underline{\lambda}\right)$.

From [92, 97, 100] it is well known that the distribution of the maximum SNR chosen from a set of N i.i.d. Rayleigh faded users converges to the Gumbel limiting distribution as $N \to \infty$. For when the number of users in the system is not fixed but random and Poisson distributed, in Section 7.6.1, we

show that the distribution of the maximum SNR also approaches the Gumbel limiting distribution as the average number of users λ grows asymptotically. Therefore, for the case where $\lambda = N$, it follows that the distribution of the SNR of the maximum in a system with random number of users converges to the distribution of the SNR of the maximum in a system with fixed number of users. It is straightforward to prove that this holds even when the number of users is zero truncated Poisson distributed. Since the CDFs converge, it implies that their inverses also converge for values close to 1. Therefore it follows that the ratio of outage capacity of the fixed number of users system in (7.43) and the outage capacity of the random number of users system in (7.44) approaches 1 as $\underline{\lambda} = N \to \infty$.

7.8. Simulations

An uplink multi-user system where both the BS and users having a single antenna is considered. In this section, using Monte-Carlo and semi Monte-Carlo simulations the BER, capacity and outage capacity are simulated to corroborate our analytical results. For all simulations considered herein, the wireless channel between any user and BS is assumed to be Rayleigh faded with parameter 1.

In Section 7.4.1 we proved that the BER is a completely monotone function of number of users N. In Fig. 22, assuming $\pi/4$ QPSK symbols are transmitted, the BER of the system with fixed N is compared with the BER

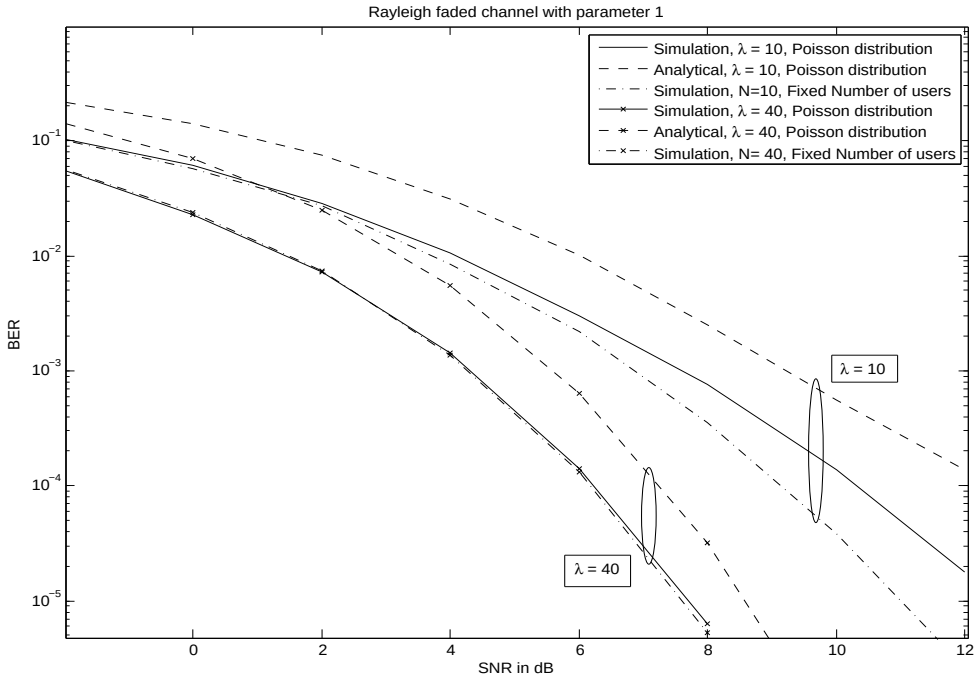

Fig. 26. BER vs. SNR: Poisson Users and Rayleigh Fading Channel

averaged across user distributions for various distributions. In Fig. 22, it is seen that the fixed number of users system performs better than all the random number of users systems. Only the error rate performance of the Poisson distributed users case comes close to the fixed case as λ increases. This follows our result in (7.27). Also the BER on the $\log - \log$ plot is seen to be a convex function which agrees with out result in Section 7.4.1.

Similarly, the capacity was shown to have a completely monotone derivative with respect to N. Thereby in Fig. 23, the capacity is plotted against λ for both the fixed and the random user case. Following the result in Section 7.4.2, it is seen that the capacity of the fixed number of users system is the

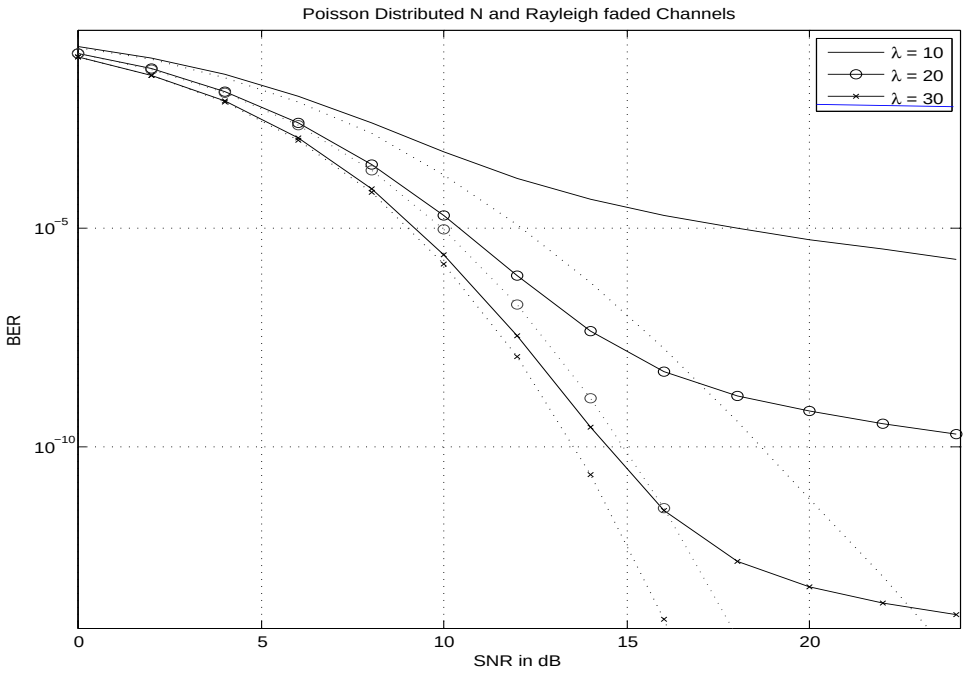

Fig. 27. Diversity Analysis: Poisson Users and Rayleigh Fading Channel

highest while for all distributions of N, the capacity is worse.

In Section 7.5, we showed that Poisson distributed random variables and geometric distributed random variables can be LT ordered, which would also order their respective BER and capacity performance when averaged across the respective user distributions. In Fig. 24 and 25 it can be seen that both BER and capacity follow their corresponding ordering, and that larger the average number of users the better the performance.

Next, in Fig. 26 we plot the BER of the multi-user system implementing MUD against average SNR for different λ's. From Fig. 26, it can be seen that the analytical approximation derived is within 1 dB of the Monte-Carlo

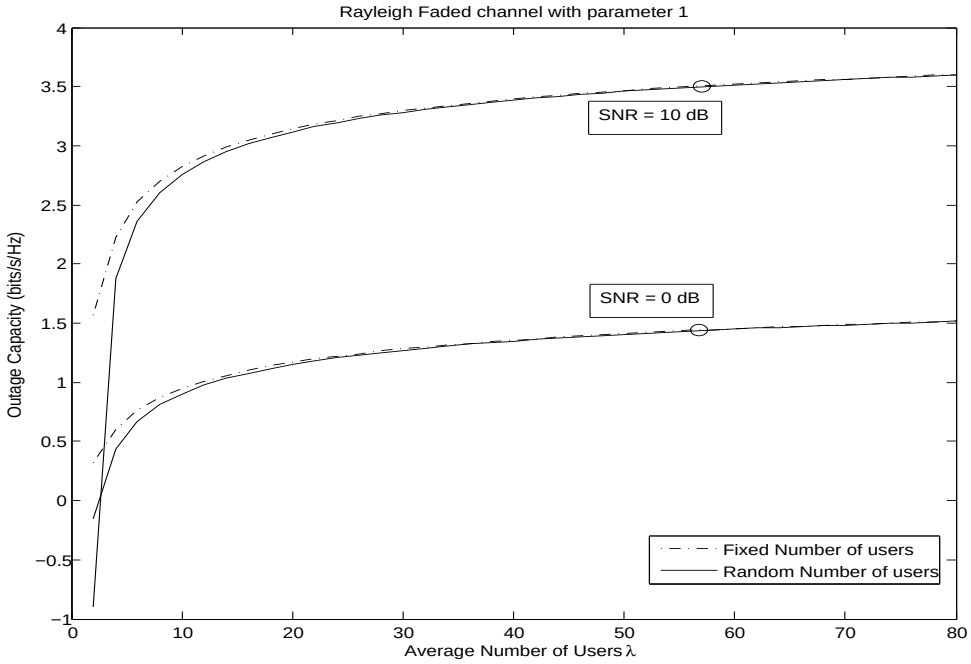

Fig. 28. Outage Capacity vs. λ: Rayleigh Fading Channel

simulation result. Following the result in (7.35) it can also be seen that larger the value of λ, the lower the error rate.

In previous simulations we saw that increasing λ leads to an improvement in performance, and for a fixed SNR the performance of the system with random number of users approaches the performance of the system with fixed number of users. Contrary to this, for a fixed λ, as SNR increases, in (7.38), it is shown that the diversity order achievable by the random number of user system is only 1. In Fig. 27 the simulation plots corroborate this result. As λ increases, the BER of random case coincides with the fixed user case for larger SNR values but eventually the slope of the curve is equal to 1. This leads us

to conclude that for low SNR's but sufficiently large λ the performance of the random number of users is identical to that of the fixed case. But for high SNR's, the performance of the random number of users case is significantly worse due to the loss in diversity order.

Finally, in Fig. 28 we plot the ϵ-capacity for both the fixed and random number of users case. It is seen that immaterial of the operating SNR, for λ significantly larger, the outage capacity is identical for both the systems. This justifies our analytical findings in Section 7.7.1.

8. CONCLUSIONS

To further reduce the implementation complexity of diversity combining scheme, switch and stay diversity combining at the receiver has been proposed and studied for space time coded MIMO systems for the first time to the best of our knowledge. A channel energy based switching algorithm is proposed. For the ML decoder, the PEP is upper bounded and the optimal switching threshold is derived, which is a logarithmic function of the average SNR ρ. The optimal switching threshold not only minimizes the PEP bound but also is shown to lead to the system achieving full diversity.

Additionally, we show that for the SSC scheme implemented, the asymptotic PEP expression behaves as $O\left(\ln(\rho)^N/(\rho)^{2N}\right)$ instead of $O\left(1/\rho^{2N}\right)$, which mandates a higher SNR value to achieve full diversity. The high SNR analysis in our work helps compare our results with the MIMO systems employing antenna selection at the receiver. This makes a complexity versus performance tradeoff analysis between antenna selection schemes and SSC in an analytical framework. From the analysis and the simulation results, it can be seen that implementing the optimal switching threshold and optimal power allocation is advisable to achieve better error rate performance.

In practice since perfect channel information at the receiver is not available it has to be estimated. Channel estimation issues have not been studied for the SSC scheme, for either SIMO or MIMO systems in the literature. To fill this gap, a training scheme for the MMSE based channel estimator at the receiver has been described and the switching algorithm based on the received

power has been proposed. In the proposed received power based switching algorithm, the switching rule is implemented prior to channel estimation leading to significant savings in implementation complexity. We also show that, by optimizing the power allocation between training and data, significant performance improvement can be achieved for all values of K. As K gets large, implying a large coherence time, the PEP performance of the channel estimation system approaches the perfect channel case. The high SNR analysis in our work helps compare our results with the MIMO systems employing antenna selection at the receiver. This makes a complexity versus performance tradeoff analysis between antenna selection schemes and SSC in an analytical framework.

From the analysis and the simulation results, it can be seen that implementing the optimal switching threshold and optimal power allocation is advisable to achieve better error rate performance. Further, when optimal switching threshold is implemented, the receive power based switching algorithm achieves the same switching rate and error rate performance as the estimated channel case but at a significantly reduced system complexity, thus making it an attractive choice for implementation.

Finally we consider the case when no channel state information is available either at the transmitter or the receiver. A receive SSC scheme for MIMO systems using differential space-time coding has been proposed and analyzed. The proposed system model has minimum complexity among all MIMO sys-

tems aiming to achieve full spatial diversity. Being limited to only 2 receive antennas due to the assumption of i.i.d branches, a novel switching algorithm is proposed and the Chernoff bound on the achievable PEP is derived. As an immediate consequence of the Chernoff bound, it is shown that for any fixed switching threshold a diversity order of only N can be achieved. Further the optimal switching threshold is derived to be a function of the SNR and it is proved that by using this optimal threshold full spatial diversity of $2N$ can be achieved while not necessarily minimizing the BER. When the derived optimal threshold is used it is seen that the error rate behaves as $O((\ln \rho)^N/\rho^{2N})$. The $(\ln \rho)^N$ penalty term mandates a higher SNR value to achieve full diversity as compared to the AS scheme, which scales as $O(\rho^{-2N})$.

Based on the PEP bound using the optimal switching threshold, the code design criteria for differential MIMO systems employing SSC at the receiver are proposed. For the special case of 2×2 system, parametric codes which outperform the existing full complexity diagonal cyclic codes and the full complexity parametric codes are designed. Though there is degradation in performance compared to the systems where CSI is available at the receiver, the proposed system realizes the benefits of a MIMO system at the least possible implementation complexity making it an attractive choice for implementation.

In the second part, for multiuser systems, unlike the model commonly assumed in literature, multiuser diversity is analyzed for when the number of users in the system is random. The BER of multiuser systems implementing

multiuser diversity is proved to be a completely monotone function of the number of users in the system, which also implies convexity. Further the throughput is shown to have a completely monotone derivative with respect to the number of users. Using the above mentioned properties along with Jensen's inequality, it is shown that the error rate performance and throughput averaged across the user distribution will always perform inferior to the corresponding performance of a system with fixed number of users. Further,for very general conditions, using Laplace transform ordering, we provide a method to compare the performance of the system for different user distributions.

When the number of users are Poisson distributed, for any user channel fading distribution, we show that the ratio of the error rate performance of the random number of users system and the fixed number of users system approaches 1 asymptotically in the average number of users. As a special case, when the user fading channels are Rayleigh distributed, closed form BER expression is provided. Also, the throughput and outage capacity is analyzed.

REFERENCES

[1] N. Chiurtu, B. Rimoldi, and E. Telatar, "On the capacity of multi-antenna gaussian channels," *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, p. 53, 2001.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," in *Wireless Personal Communication*, vol. 6, no. 2, March 1998, pp. 311–335.

[3] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Laboratories Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[4] A. Goldsmith, S. A. Jafar, Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," in *Selected Areas in Communication, IEEE Journal on*, vol. 21, no. 5, June 2003, pp. 46–56, submitted.

[5] A. Goldsmith, *Fundamentals of Wireless Communication*. NY: Cambridge : Cambridge University Press, 2005.

[6] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over rich-scattering wireless channel," *Bell Laboratories Technical Journal*, no. 2, 1998.

[7] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[8] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in mulitple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, May 2003.

[9] S. M. Alamouti, "A simple transmit diversity technique for wireless communication," *Selected Areas in Communication, IEEE Journal on*, vol. 16, pp. 1451–1458, Oct. 1998.

[10] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal design," *Information Theory, IEEE Transactions on*, vol. 45, pp. 1456–1458, July 1999.

[11] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *Information Theory, IEEE Transactions on*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[12] T. S. Rappaport, *Wireless Communications*, 2nd ed. Prentice Hall, 2005.

[13] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, August 2000.

[14] G. L. Stuber, *Principles of Mobile Communication*. Springer, 2001.

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

[16] L. Tong, G. Xu, and T. Kailath, "Blind channel identification and equalization based on second–order statistics: A time–domain approach," *Information Theory, IEEE Transactions on*, vol. IT-40, pp. 340–349, March 1994.

[17] E. Moulin, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *Signal Processing, IEEE Transactions on*, vol. SP-43, pp. 516–525, Feb. 1995.

[18] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification for multichannel fir system," in *Proc. IEEE ICASSP*, vol. 4, May 1994, pp. 581–584.

[19] "*Group Special Mobile (GSM) Recommendations*," in *GSM Series*, vol. 01–12 Std., 1990.

[20] "TIA/EIA, TIA/EIA/IS-136: TDMA *Cellular/PCS-Radio Interface - Mobile Station - Base Station Compatibility- Digital Control Channel*," Englewood Cliffs, NJ: Prentice Hall Std., 1997.

[21] "*Physical Channels and Mapping of Transport Channels onto Physical Channels FDD*," in *WCDMA Std.*, 2001.

[22] "*Physical Layer Standard for CDMA2000 Spread Spectrum Systems*," in *CDMA2000 Std.*, 2001.

[23] "*Broadband Radio Access Networks: HIPERLAN Type 2 Physical Layer*," in *HIPERLAN II Std.*, 2001.

[24] "*ANSI/IEEE Std. 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer Specifications*," in *IEEE 802.11 Std.*, 1999.

[25] "*IEEE Std 802.11a-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer Specifications: High Speed Physical Layer Extension in the 5 GHz Band*," in *IEEE 802.11a Std.*, 1999.

[26] "*IEEE Std 802.11b-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer Specifications: High Speed Physical Layer Extension in the 2.4 GHz Band*," in *IEEE 802.11a Std.*, 1999.

[27] "*Digital Video Broadcasting (DVB) Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television*," in *DVB–T Std.*, 2001.

[28] "*ATSC Digital Television Standard (Revision B), Doc. A/53B*," in *ATSC Std.*, 2001.

[29] A. Narasimhamurthy and C. Tepedelenlioglu, "Antenna selection for MIMO-OFDM systems with channel estimation error," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 5, pp. 2269–2278, Jun 2009.

[30] Z. Liu, Y. Xin, and G. B. Giannakis, "Space time frequency coded OFDM over frequency selective fading channels," *Signal Processing, IEEE Transactions on*, vol. 50, no. 10, Oct. 2002.

[31] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple antenna wireless links?" *Information Theory, IEEE Transactions on*, vol. 49, no. 4, April 2003.

[32] I. Bahceci, T. M. Duman, and Y. Altunbasak, "Antenna selection for multiple-antenna transmission systems: Performance analysis and

code construction," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2669–2681, Oct. 2003.

[33] D. Gore and A. Paulraj, "MIMO antenna subset selection with space time coding," in *Signal Processing, IEEE Transactions on*, vol. 50, March 2002, pp. 2580–2588.

[34] A. Rustako, Y. Yeh, and R. Murray, "Performance of feedback and switch space diversity 900 MHz FM mobile radio systems with Rayleigh fading," *Communications, IEEE Transactions on*, vol. 21, no. 11, pp. 1257–1268, Nov 1973.

[35] M. Blanco and K. Zdunek, "Performance and optimization of switched diversity systems for the detection of signals with Rayleigh fading," *Communications, IEEE Transactions on*, vol. 27, no. 12, pp. 1887–1895, Dec 1979.

[36] A. Abu-Dayya and N. Beaulieu, "Analysis of switched diversity systems on generalized-fading channels," *Communications, IEEE Transactions on*, vol. 42, no. 11, pp. 2959–2966, Nov 1994.

[37] ——, "Switched diversity on microcellular Ricean channels," *Vehicular Technology, IEEE Transactions on*, vol. 43, no. 4, pp. 970–976, Nov 1994.

[38] Y.-C. Ko, M.-S. Alouini, and M. Simon, "Analysis and optimization of switched diversity systems," *Vehicular Technology, IEEE Transactions on*, vol. 49, no. 5, pp. 1813–1831, Sep 2000.

[39] C. Tellambura, A. Annamalai, and V. Bhargava, "Unified analysis of switched diversity systems in independent and correlated fading channels," *Communications, IEEE Transactions on*, vol. 49, no. 11, pp. 1955–1965, Nov 2001.

[40] H. C. Yang and M.-S. Alouini, "Performance analysis of multibranch switched diversity systems," *Communications, IEEE Transactions on*, vol. 51, no. 5, pp. 782–794, May 2003.

[41] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[42] H. C. Yang and M.-S. Alouini, "Markov chains and performance comparison of switched diversity systems," *Communications, IEEE Transactions on*, vol. 52, no. 7, pp. 1113–1125, July 2004.

[43] Q. Ma and C. Tepedelenlioglu, "Antenna selection for space-time coded systems with imperfect channel estimation," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, 2005.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[45] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *Selected Areas in Communications, IEEE Journal on*, vol. 5, no. 5, pp. 871–878, Jun 1987.

[46] A. Ghrayeb and T. M. Duman, "Performance analysis of MIMO systems with antenna selection over quasi-static fading channels," *Vehicular Technology, IEEE Transactions on*, vol. 52, pp. 281–288, March 2003.

[47] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," in *IEEE Microwave Mag.*, vol. 5, no. 1, March 2004, pp. 46–56.

[48] M. Blanco, "Diversity receiver performance in Nakagami fading," in *Proc. 1983 IEEE Southeastern Conf., Orlando*, 1983, p. 529532.

[49] A. B. Narasimhamurthy and C. Tepedelenliogly, "Space-time coding for receive switch and stay combining," *Wireless Communications, IEEE Transactions on*, 2009, accepted for Publication.

[50] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 543–564, Mar 2000.

[51] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *Communications, IEEE Transactions on*, vol. 48, no. 12, pp. 2041–2052, Dec 2000.

[52] A. Dogandzic, "Chernoff bounds on pairwise error probabilities of space-time codes," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1327–1336, May 2003.

[53] Q. Ma and C. Tepedelenlioglu, "Antenna selection for unitary space-time modulation," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, pp. 3620–3631, Oct. 2005.

[54] G. H. Golub and C. F. V. Loan, *Matrix computations*, 3rd ed. Baltimore, MD: John Hopkins University Press, 1996.

[55] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the LambertW function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[56] B. Hughes, "Differential space-time modulation," *Information Theory, IEEE Transactions on*, vol. 46, no. 7, pp. 2567–2578, Nov 2000.

[57] X.-B. Liang and X.-G. Xia, "Unitary signal constellations for differential space-time modulation with two transmit antennas: parametric codes, optimal designs, and bounds," *Information Theory, IEEE Transactions on*, vol. 48, no. 8, pp. 2291–2322, Aug 2002.

[58] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1277 –1294, jun 2002.

[59] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, vol. 1, 18-22 1995, pp. 331 –335 vol.1.

[60] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *IEEE International Symposium on Information Theory*. Citeseer, 1997, pp. 27–27.

[61] J. Holtzman, "Asymptotic analysis of proportional fair algorithm," vol. 2, sep. 2001, pp. F–33 –F–37 vol.2.

[62] D. Piazza and L. Milstein, "Multiuser diversity-mobility tradeoff: modeling and performance analysis of a proportional fair scheduling," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 1, 17-21 2002, pp. 906 – 910 vol.1.

[63] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power mqam for fading channels," *Communications, IEEE Transactions on*, vol. 45, no. 10, pp. 1218 –1230, oct. 1997.

[64] S. T. Chung and A. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *Communications, IEEE Transactions on*, vol. 49, no. 9, pp. 1561 –1571, sep. 2001.

[65] B. Choi and L. Hanzo, "Optimum mode-switching-assisted constant-power single- and multicarrier adaptive modulation," *Vehicular Technology, IEEE Transactions on*, vol. 52, no. 3, pp. 536 – 560, may. 2003.

[66] M. S. Alouini and A. J. Goldsmith, "Adaptive M-QAM modulation over Nakagami fading channels," in *Proc. of IEEE Global Commun. Conf*, Nov. 1997, p. 218223.

[67] D. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *Communications, IEEE Transactions on*, vol. 47, no. 6, pp. 844 –855, jun. 1999.

[68] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, "Adaptive modulation systems for predicted wireless channels," *Communications, IEEE Transactions on*, vol. 52, no. 2, pp. 307 – 316, feb. 2004.

[69] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *Signal Processing Magazine, IEEE*, vol. 17, no. 3, pp. 62 –75, may. 2000.

[70] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser ofdm with adaptive subcarrier, bit, and power allocation," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 10, pp. 1747 –1758, oct. 1999.

[71] R. Heath Jr., M. Airy, and A. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 2, 2001, pp. 1194–1199.

[72] R. Gozali, R. Buehrer, and B. Woerner, "The impact of multiuser diversity on space-time block coding," *Communications Letters, IEEE*, vol. 7, no. 5, pp. 213 – 215, may. 2003.

[73] J. Jiang, R. Buehrer, and W. Tranter, "Antenna diversity in multiuser data networks," *Communications, IEEE Transactions on*, vol. 52, no. 3, pp. 490 – 497, mar. 2004.

[74] E. Larsson, "On the combination of spatial diversity and multiuser diversity," *Communications Letters, IEEE*, vol. 8, no. 8, pp. 517 – 519, aug. 2004.

[75] J. Chung, C.-S. Hwang, K. Kim, and Y. K. Kim, "A random beamforming technique in mimo systems exploiting multiuser diversity," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 5, pp. 848 – 855, jun. 2003.

[76] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channels with partial side information," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 506 – 522, feb. 2005.

[77] E. F. Chaponniere, P. J. Black, J. M. Holtzman, and D. Tse, "Transmitter directed code division multiple access system using path diversity to equitably maximize throughput," US Patent No. 6449490, Sept. 2002.

[78] M. Andrews, "Instability of the proportional fair scheduling algorithm for hdr," *Wireless Communications, IEEE Transactions on*, vol. 3, no. 5, pp. 1422 – 1426, sep. 2004.

[79] X. Liu, E. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 10, pp. 2053 –2064, oct. 2001.

[80] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *Communications Magazine, IEEE*, vol. 39, no. 2, pp. 150 –154, feb. 2001.

[81] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *Networking, IEEE/ACM Transactions on*, vol. 13, no. 3, pp. 636 – 647, jun. 2005.

[82] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," vol. 2, mar. 2003, pp. 1084 – 1094 vol.2.

[83] ——, "Opportunistic splitting algorithms for wireless networks," vol. 3, mar. 2004, pp. 1662 –1672 vol.3.

[84] T. Tang, R. W. Heath, S. Cho, and S. Yun, "Opportunistic feedback for multiuser mimo systems with linear receivers," *Communications, IEEE Transactions on*, vol. 55, no. 5, pp. 1020 –1032, may. 2007.

[85] S. Y. Park, D. Park, and D. Love, "On scheduling for multiple-antenna wireless networks using contention-based feedback," *Communications, IEEE Transactions on*, vol. 55, no. 6, pp. 1174 –1190, jun. 2007.

[86] T. Tang and R. Heath, "Opportunistic feedback for downlink multiuser diversity," *Communications Letters, IEEE*, vol. 9, no. 10, pp. 948 – 950, oct. 2005.

[87] D. V. Widder, *An introduction to transform theory.* Academic Press, 1971.

[88] A. Alzaid, J. S. Kim, and F. Proschan, "Laplace ordering and its applications," *Journal of Applied Probability*, vol. 28, no. 1, pp. 116–130, Mar. 1991, ArticleType: research-article / Full publication date: Mar., 1991 / Copyright 1991 Applied Probability Trust. [Online]. Available: http://www.jstor.org/stable/3214745

[89] M. Shaked and J. G. Shanthikumar, *Stochastic orders and their applications*. Academic Press, 1994.

[90] A. Mller and D. Stoyan, *Comparison methods for stochastic models and risks*. John Wiley and Sons, 2002.

[91] J. Galambos, "The distribution of the maximum of a random number of random variables with applications," *Journal of Applied Probability*, vol. 10, no. 1, pp. 122–129, Mar. 1973, ArticleType: primary_article / Full publication date: Mar., 1973 / Copyright 1973 Applied Probability Trust. [Online]. Available: http://www.jstor.org/stable/3212500

[92] H. A. David, *Order Statistics*, 2nd ed. New York: Wiley, 1981.

[93] P. Downey, "An abelian theorem for completely monotone functions," 1993.

[94] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*. Cambridge University Press, July 1989.

[95] W. Feller, *An Introduction To Probability Theory And Its Applications, 2Nd Ed.* Wiley India Pvt. Ltd., 2009.

[96] E. Castillo, A. S. Hadi, N. Balakrishnan, and J. M. Sarabia, *Extreme value and related models with applications in engineering and science*. Wiley, 2005.

[97] C. Tepedelenlioglu, Y. Zhang, and O. Rahman, "Asymptotic ber analysis of a simo multiuser diversity system," vol. 58, no. 9, nov. 2009, pp. 5330 –5335.

[98] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Courier Dover Publications, 1964.

[99] M.-O. Pun, V. Koivunen, and H. Poor, "SINR analysis of opportunistic MIMO-SDMA downlink systems with linear combining," in *Communications, 2008. ICC '08. IEEE International Conference on*, 19-23 2008, pp. 3720 –3724.

[100] Q. Ma and C. Tepedelenlioglu, "Practical multiuser diversity with outdated channel feedback," *Vehicular Technology, IEEE Transactions on*, vol. 54, no. 4, pp. 1334 – 1345, jul. 2005.