Multiple-Input Multiple-Output Wireless Systems: Coding, Distributed Detection and Antenna Selection

A Thesis
Presented to
The Academic Faculty

by

Israfil Bahceci

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
December 2005

Multiple-Input Multiple-Output Wireless Systems: Coding, Distributed Detection and Antenna Selection

Approved by:

Dr. Yucel Altunbasak, Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Russell Mersereau
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Faramarz Fekri,
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Glenn Smith
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Xiaoming Huo
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: August 25, 2005

This dissertation is dedicated to my family.

Thank you for your love, encouragement, and support.

ACKNOWLEDGEMENTS

I would like to offer my thanks and gratitudes for Dr. Yücel Altunbaşak and for Dr. Tolga M. Duman for their thorough support throughout the course of this work. I am also thankful to Dr. Russell M. Mersereau, Dr. Faramarz Fekri, Dr. Glenn Smith and Dr. Xiaoming Huo for their helpful advice as my committee members, to Dr. Ghassan Al-Regib for his valuable suggestions, to CSIP staff Christy Ellis, Kay Gilstrap and Charlotte Doughty for their help in my administrative problems, to the National Science Foundation for the financial support, and to all my friends at MCCL, CSIP, Georgia Tech and Atlanta for making the last four years in Atlanta precious.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Multiple-input multiple output (MIMO) wireless systems have become a widespread technology in many different forms and in various application areas because of the rich benefits they serve. For example, deploying multiple antennas at both sides of a wireless communication system improves the system performance and increases the information rates. The users in a wireless network can cooperate with each other and establish a multiple-input multiple-out wireless system that result in enhanced rates and improved reliability. The performance of a wireless sensor network becomes significantly better by the cooperation of the individual nodes to perform distributed and/or collaborative processing to complete some given task, which, again, results in a multiple-input-multiple output wireless system. Due to their potential applications, the MIMO systems will be an important part of future communication systems.

In this dissertation, we study a number of important issues that arise in multiple-input multiple-out wireless systems. We first deal with a MIMO antenna system and study two problems: (i) the performance of the MIMO system with antenna subset selection and (ii) joint source-channel coding for MIMO systems. Antenna subset selection reduces the implementation cost of a MIMO antenna system by allowing a reduction in the number of required RF chains that would otherwise be needed for a full-complexity system. In this dissertation, we specifically investigate the performance of receive antenna selection and derive Chernoff upper bounds on the pairwise error probability for the energy-based selection. We consider three different situations: (i) selection over an independently and identically distributed MIMO fading channel, (ii) selection over correlated fading channel where the subchannels among the antenna pairs are correlated, and (iii) selection for a space-time coded orthogonal frequency multiplexing system. In all cases, explicit upper bounds are derived and it is shown that using the energy-based antenna selection, one can achieve the same diversity order as that attained by a full-complexity MIMO system. The resulting

upper bounds are used to design optimal space-time codes for the the MIMO system using antenna selection. We then look into the joint source-channel coding problem for a MIMO antenna system and develop a turbo-coded multiple description code for multiple antenna transmission. Multiple description codes generate a number of correlated streams of some signal to be transmitted over preferably independent on-off type channels so that upon the reception of all or a subset of the streams, one can attain acceptable level of reconstruction qualities. Motivated by the observation that independently fading channels intrinsically inhabit in MIMO antenna systems, we propose the use of MDC over MIMO wireless channels. We show via simulations that by the proposed iterative joint source-channel decoding that exchanges the extrinsic information between the source code and the channel code, we can achieve better reconstruction quality than that can be achieved by the single-description codes at same rate.

In the rest of the dissertation, we deal with a network of wireless nodes having single antennas. We study two problems: (i) channel coding for user cooperation diversity in wireless networks, and (ii) distributed detection in wireless sensor networks. Cooperative diversity refers to the spatial diversity obtained by a virtual MIMO antenna system consisting of a number of geographically separated mobile single-antenna nodes that can cooperate between each other. Channel coding even for simple cooperative networks is a widely open problem. In this dissertation, we develop a turbo-code based distributed channel code for a three-terminal wireless relay channel where a source node is assisted by a relay node for communication with the destination node. We consider a full-duplex relay that employs a simple decode-and-forward method. Both the source and the relay nodes use turbo codes, and due to the full-duplex transmission, the destination node observes the superposition of the transmitted signals. We propose an iterative turbo decoding algorithm that exploits the information arriving from both the source and relay nodes. Simulation results show that with the proposed scheme, one can perform very close to the capacity of a wireless relay channel. In addition, the proposed scheme can readily be extended to multiple relay networks. We next consider the binary distributed detection problem in wireless sensor

networks where spatially distributed sensor nodes make individual observations, locally process their observations and transmit some related information through a noisy channel to a fusion center where a final decision is made. We consider two network configurations: (i) parallel network where each sensor node transmits the data to a fusion center, and (ii) serial network where the sensor nodes constitute a serially connected multi-hop network and the individual sensor nodes base their decisions on their own observations as well as on the signal received from its ascending node. We consider detection strategies based on single-bit and multiple-bit decisions. We derive expressions for the detection and false alarm rates that is used for designing the optimal detection rules (thresholds) at all sensor nodes. Observing that determining the optimal designs might be formidable even for small-scale networks, we propose an analog approach to the distributed detection in wireless sensor networks where each sensor nodes simply amplifies-and-forwards its observation (or sufficient statistics) to the fusion center. This method requires very simple processing at the local sensor and the optimal design reduces to a suitable power allocation across sensor nodes. Numerical examples indicate that the analog approach is superior to the digital approach in most cases.

CHAPTER I

INTRODUCTION

In this chapter, we briefly summarize the necessary background for several topics that will be considered in the dissertation. Since we deal with wireless systems, we first present the typical characteristics of wireless channels in Section 1.1. In Section 1.2, we discuss the use of antenna diversity over fading channels and also present the space-time codes and turbo coded modulation for multiple antenna transmission. Section 1.3 describes a diversity method that can be obtained by user cooperation in the presence of multiple users. In Section 1.4, we summarize the multiple description coding which can be imagined as a method that provides source diversity. Section 1.5 explains the distributed detection problem, and finally Section 1.6 summarizes the contribution of the dissertation.

1.1 Fading Channels

There are various physical mediums through which the information may be exchanged. Regardless of which transmission medium is used, "non-ideal" communication channel corrupts the transmitted signals via many possible mechanisms such as noise, attenuation, fading and interference from other users. In the simplest case, the received signal is affected only by additive ambient noise modelled as a white Gaussian process, resulting in additive white Gaussian noise (AWGN) channel.

In wireless communications, the information is transmitted via propagation of electromagnetic waves. Several mechanisms like reflection, refraction and scattering are effective in the propagation process. Due to these three main mechanisms, the transmitted signal is received via multiple paths. In general, this multipath propagation medium is time varying, therefore, it is reasonable to characterize the time varying nature of the multi-path propagation channels using statistical techniques. The time varying channel impulse response of a typical wireless channel to a pulse at time $t-\tau$ can be expressed as (in low-pass equivalent

form)

$$c(\tau, t) = \sum_n \alpha_n(t) e^{-j2\pi f_c \tau_n(t)} \delta(\tau - \tau_n(t))$$

where f_c is the carrier frequency, and $\alpha_n(t)$ and $\tau_n(t)$ are the attenuation factor and the propagation delay for the n^{th} path, respectively. If there are a large number of multipaths in the medium, which is a typical case for a mobile subscriber generally surrounded by many scatterers, by applying the law of large numbers, we conclude that the channel impulse response at time t is Gaussian distributed for each delay τ. Thus, the channel impulse response is a Gaussian process. If it is a zero-mean Gaussian process, then the envelope is Rayleigh distributed and the phase is uniformly distributed in the interval $(0, 2\pi)$. That is, the envelope

$$R = |c(\tau, t)|$$

has the probability density function (p.d.f.) of

$$p_R(r) = \frac{2r}{\Omega} e^{-r^2/\Omega}, \ r \geq 0$$

where $\Omega = E(R^2)$ is the signal power.

Since the wireless environment is characterized by a statistical process, the correlation functions and power spectral densities are the useful tools to express these characteristics. We assume that $c(\tau, t)$ is a *wide sense stationary* process in the t-variable, and that the random processes for different path delays, τ, are uncorrelated. These assumptions yield to the classical *wide sense stationary uncorrelated-scattering* channel model. The average power output of the channel as a function of the time delay, τ, is then defined by the *delay power spectrum*, $\phi_c(\tau)$. The range of values of τ over which $\phi_c(\tau)$ is essentially non-zero is called the *multipath spread of the channel* and denoted by T_m. A similar characterization is obtained in the frequency domain starting with the time varying transfer function of the channel. The same assumptions lead to the *spaced-frequency spaced-time correlation function of the channel*, $\Phi_c(\Delta f; \Delta t)$, which provides a measure of the frequency coherence of the channel. The *coherence bandwidth of the channel*, $(\Delta f)_c$, is defined as the maximum

spacing Δf between frequencies at which two sinusoids are affected by approximately the same complex gain, *i.e.*, $\Phi_c(\Delta f; 0) \approx \Phi_c(0; 0)$, and is approximately given by

$$(\Delta f)_c \approx \frac{1}{T_m}.$$

Similarly, the *coherence time of the channel*, T_c, is defined as the maximum spacing Δt between instants at which two impulses are affected by approximately the same complex gain, *i.e.*, $\Phi_c(0; \Delta t) \approx \Phi_c(0; 0)$. In addition, we define the *Doppler spread*, B_d, as the maximum frequency shift due to the variations in the channel or the relative motion of the transmitter and receiver. An approximate value for B_d is given by $B_d \approx \frac{1}{T_c}$.

Suppose that we transmit digital information over the channel by modulating the basic pulse at a rate $1/T$, where T is the symbol duration. Then, the bandwidth of the pulse is approximately $W \approx 1/T$. If the bandwidth of the signal is much smaller than the coherence bandwidth of the channel, *i.e.*, $W \ll (\Delta f)_c$, then all the frequency components in the pulse are affected by the same attenuation and phase shift during the transmission of one symbol. Hence, the channel is said to be *frequency non-selective*. Conversely, if the bandwidth of the signal W is larger than the coherence bandwidth, $(\Delta f)_c$, the pulse is subject to different complex gains across the frequency band and of the symbol and the channel is said to be *frequency selective*.

In this dissertation, we mainly consider the transmission of digital signals over frequency non-selective slowly fading channels. For this channel, if the low-pass transmitted signal is $\hat{s}(t)$, the received equivalent low-pass signal in the signaling interval is

$$y(t) = \alpha e^{-j\phi} \hat{s}(t) + z(t), \quad 0 \leq t \leq T$$

where $z(t)$ represents the complex valued zero-mean white Gaussian noise process (with variance $N_0/2$ per dimension), α is the Rayleigh distributed attenuation and ϕ is the uniformly distributed phase shift due to the channel. Let us assume that fading is sufficiently slow and the phase shift ϕ can be estimated from the received signal without any error, then the channel is described by

$$y(t) = \alpha \hat{s}(t) + z^{'}(t)$$

where $z^{'}(t)$ is also white Gaussian.

To see how fading deteriorates the system performance, let us consider the use of binary phase shift keying (BPSK) modulation over a Rayleigh fading channel [125]. The average bit error probability over a frequency non-selective fading channel is given by

$$\begin{aligned} P_{b,fading} &= \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\gamma}_b}{1+\bar{\gamma}_b}}\right) \\ &\approx \frac{E_b}{4N_0} \end{aligned} \tag{1}$$

where $\bar{\gamma}_b = \frac{E_b}{N_0}E[\alpha^2]$ is the average SNR, E_b is the energy in the transmitted pulse $\hat{s}(t)$ and $E[.]$ denotes the expectation operator. On the other hand, the bit error probability for BPSK modulation over AWGN channel is given by

$$\begin{aligned} P_{b,AWGN} &= Q(\sqrt{2E_b/N_0}) \\ &\leqslant \frac{1}{2}\exp\left(\frac{-E_b}{N_0}\right) \end{aligned} \tag{2}$$

From (1) and (2) , it is clear that while the error probability decreases exponentially with SNR for the AWGN channel, it decreases only inversely for the Rayleigh fading channel case. Therefore, fading degrades the performance of a wireless communication system significantly.

In order to combat fading, the receiver is typically provided with multiple replicas of the transmitted signal so that it can use these replicas to extract the transmitted information [124, 153]. The transmitted information will be recovered with high probability since all the replicas will not typically fade simultaneously. This method is called *diversity* and it is one of the most effective techniques for combating multi-path fading. There are many diversity techniques including temporal, frequency and space diversity techniques. In the next section, we elaborate on spatial (antenna) diversity that can be obtained by employing multiple antenna elements at the transmitter and/or receiver.

1.2 Antenna Diversity

Assume that the receiver is equipped with L antennas that are sufficiently separated so that the signals received from each antenna element are (almost) independent from each other.

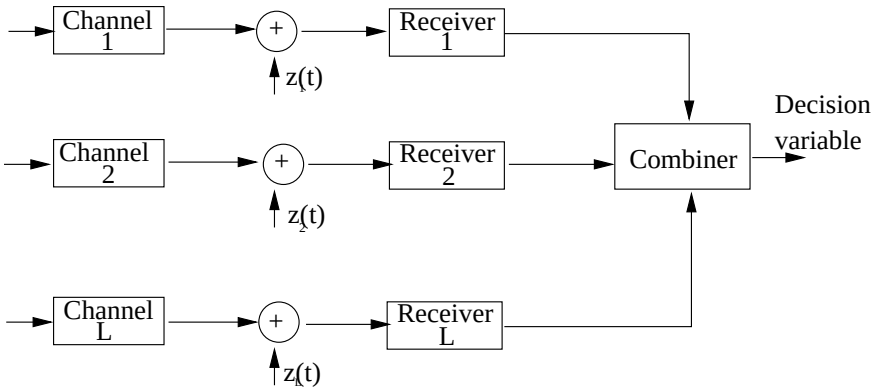


Figure 1: Model of a digital communication system with diversity.

Note that this system can be viewed as a single-input multiple-output (SIMO) wireless system. A block diagram of such a system is shown in Figure 1. The signals from different diversity branches are combined with a diversity combiner. The combiner simply uses the information received from the different branches and forms the decision variable. One way of combining is to weight the signals received from each diversity branch using the complex conjugates of the corresponding channel gains and add them up. This combiner is called *maximal ratio combiner* and the signal to noise ratio at the output is the sum of signal to noise ratios of individual branches. This combining scheme is optimal, however, it is only applicable when the channel estimates are available. There are also various other possible combining schemes. For example, in *equal gain combining*, the signals from each branch are added after co-phasing. In *switched diversity*, the combiner switches between the branches such that if the signal to noise ratio drops below a threshold value in the current branch, it selects another branch with SNR larger than the threshold.

In this dissertation our main interest will be the *selection diversity combining* in which the branches are all monitored and the branch with the strongest signal level is selected at any one time. As described later, while this method is a suboptimal one, it reduces the implementation complexity significantly while retaining the diversity order of system.

1.2.1 Multiple-Input Multiple-Output Antennas and Space-Time Codes

Antenna diversity at the receiver side is well-known and has been widely implemented in many communication systems such as in cellular networks. However, it is also possible to deploy multiple antennas at transmitter side, leading to a multiple-input multiple-output (MIMO) antenna system. Consider, for example, a mobile communication system where we have M transmit and N receive antennas. Figure 2 displays such a wireless link. For each channel use, the signal $s_i(t),\ i = 1, \cdots, M$, is transmitted from i^{th} antenna element. The signal at each receive antenna is a noisy superposition of the M transmitted signals, *i.e.*,

$$x_i = \sum_{k=1}^{M} h_{i,k} s_k + w_i$$

where $h_{i,k}$ is the path gain from k^{th} transmit antenna to i^{th} receive antenna and w_i is the additive Gaussian noise at the i^{th} receive antenna. The channel capacity of such a multiple antenna system in the presence of Gaussian noise is determined by [164] and [48, 49], where it is shown that for a fading channel where the sub-channels between each pair of transmit and receive antennas are independent Rayleigh distributed and each use of the channel corresponds to an independent realization of channel transfer function, the average channel capacity scales linearly with the the number of the transmit antennas (as long as the number of antenna elements at the receiver is grater than or equal to the number of antenna elements at the transmitter) provided that exact channel state information is available at the receiver.

These information theoretic results led to the development of so-called "space-time codes" [163]. These coding schemes propose the joint design of coding and modulation along with transmit and receive diversity. To accomplish this, space-time trellis codes, space-time block codes and turbo coded modulation systems have been developed [149–151, 160, 163]. Several space-time coding schemes for multiple antenna transmission have also been proposed in [2, 73, 141, 144, 154, 161, 162, 182]

In [163], the multi-antenna performance criteria for designing space-time codes are derived under the assumption that the fading is quasi-static and frequency non-selective. Two performance measures based on the matrices constructed from the pairs of code sequences have been developed: the *rank criterion* and the *determinant criterion*. Let us denote the

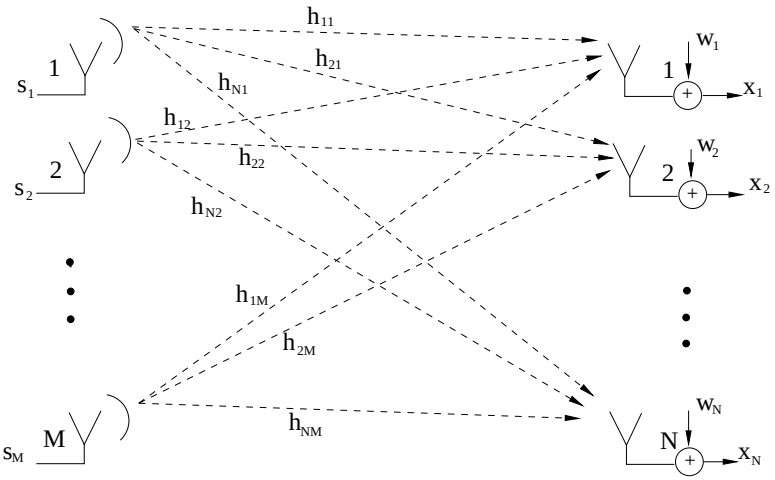


Figure 2: A wireless link comprising M transmit and N receive antennas.

codeword difference matrix

$$B(\boldsymbol{c},\boldsymbol{e}) = \begin{pmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \dots\dots & e_l^1 - c_l^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \dots\dots & e_l^2 - c_l^2 \\ e_1^3 - c_1^3 & e_2^3 - c_2^3 & \ddots\ \vdots & e_l^3 - c_l^3 \\ \vdots & \vdots & \ddots\ \ddots & \vdots \\ e_1^M - c_1^M & e_2^M - c_2^M & \dots\dots & e_l^M - c_l^M \end{pmatrix}$$

where $\boldsymbol{c}$ is the transmitted codeword and $\boldsymbol{e}$ is the decoded codeword such that

$$\boldsymbol{c} = c_1^1\, c_1^2 \dots c_1^M\, c_2^1\, c_2^2 \dots c_2^M \dots c_l^1\, c_l^2 \dots c_l^M$$

and

$$\boldsymbol{e} = e_1^1\, e_1^2 \dots e_1^M\, e_2^1\, e_2^2 \dots e_2^M \dots e_l^1\, e_l^2 \dots e_l^M.$$

Define $A(\boldsymbol{c},\boldsymbol{e}) = B(\boldsymbol{c},\boldsymbol{e})B(\boldsymbol{c},\boldsymbol{e})^H$. Here, c_k^m denotes the signal transmitted from m^{th} antenna element at time k. The minimum rank of $B(\boldsymbol{c},\boldsymbol{e})$ for any codeword pair determines the *diversity gain* while the minimum determinant of the $A(\boldsymbol{c},\boldsymbol{e})$ determines the *coding gain*. Using these criteria, the authors design trellis codes with performance within $2-3$ dB of the outage capacity at a frame error rate of 0.1.

The space-time coding schemes mentioned so far assume the availability of channel state information at the receiver. However, estimating the fading coefficients between each pair

of transmit and receive antenna elements becomes difficult, if not impossible, if the fading is fast or a large number of antenna elements are used. For such cases, it is necessary to develop modulation techniques that do not need the channel state information. Assuming that no channel state information is available at the transmitter and the receiver, Hochwald et. al. present the information theoretic limits of multi antenna systems for Rayleigh block fading channels and demonstrate that the capacity achieving signals are orthogonal to each other with respect to time across transmit antennas [84, 109]. The resulting constellations are called unitary space-time constellations. Such constellations have been designed and shown to perform well for fast fading scenarios [8, 85]. Other recently proposed methods that do not require the estimation of the channel state information include differential space-time modulation schemes [86, 89, 159]. These schemes can be considered as extensions of the standard differential phase-shift keying, where the transmitted signals are space-time symbols in the form of complex matrices.

A natural drawback of the multiple antenna systems is the increased complexity due to the need for multiple RF chains. Therefore, there is a considerable effort in exploring multiple input multiple output (MIMO) systems that significantly reduce this complexity, but still provide similar capacity and performance improvements. A promising technique to achieve this goal is to select a subset of antennas at the transmitter and/or receiver [5–7] [62, 113, 135]. For example, for the case of single antenna selection at the receiver, assuming that the fading is slow, the received signal power can be monitored periodically, and only the signal of the receive antenna observing the largest instantaneous SNR can be fed to the RF chain for processing. Thus significant reduction in hardware costs can be attained while reaping the benefits of MIMO signaling.

Antenna subset selection for MIMO systems has been investigated by many researchers [62–64, 66, 82, 98, 113, 117, 135, 181]. For example, the capacity of MIMO systems with antenna selection (only at the receiver) is considered in [113]. The selection is based on the capacity, *i.e.*, those antennas that achieve the largest capacity are selected. The authors evaluate upper bounds on the capacity of the system and conclude that one can achieve a capacity very close to that of the full-complexity system as long as the number of antennas

selected is greater than or equal to the number of transmit antennas. In [62] and [135], transmit antenna selection is studied for systems where limited feedback on the channel state information is available to the transmitter. In these systems, channel capacity is used as the optimality criterion and the selection is performed by an exhaustive search. Gorokhov proposes a suboptimal selection algorithm in [66] that decreases the computational complexity significantly. An antenna selection method seeking the minimization of the error rate using linear receivers is considered in [82]. These studies indicate that antenna subset selection attains an acceptable performance while reducing the implementation costs.

1.2.2 Turbo Codes and Turbo-Coded Modulation for Multiple Antenna Transmission

Turbo coding is a way of constructing powerful codes from two or more component codes, which are generally convolutional codes. It is distinguished with its excellent performance at very low signal to noise ratios, *i.e.*, within 1 dB of the channel capacity [16, 17]. The main idea in turbo codes is to concatenate two recursive systematic convolutional encoders in parallel using an interleaver as shown in Figure 3. The information sequence is partitioned into blocks of size N_i. The input to the first encoder is the information sequence itself while the input to the second one is an interleaved version. The bits from the systematic block, first and second parity blocks are then multiplexed to produce the encoded sequence (see Figure 3). The k^{th} element of the encoded sequence is thus defined as $\boldsymbol{s}_k = (s_k^s, s_k^{1p}, s_k^{2p})$. In addition to parallel concatenation, one can also use a serial of convolutional encoders [14], and instead of using convolutional codes, one can use linear block codes [15]. And as shown in [36], we can concatenate more than two component codes in parallel.

In Figure 4, we depict the block diagram of the iterative turbo decoder. Since the optimum decoding algorithm has a very high complexity, a suboptimum iterative decoding algorithm is employed. In the iterative decoding of the turbo codes, the maximum a posteriori (MAP) algorithm proposed in [9] is employed. In the figure, the two central units, MAP, denote the Maximum A-posteriori Probability decoders associated with the two component encoders. In the first iteration, the first MAP decoder computes the log-likelihood ratio

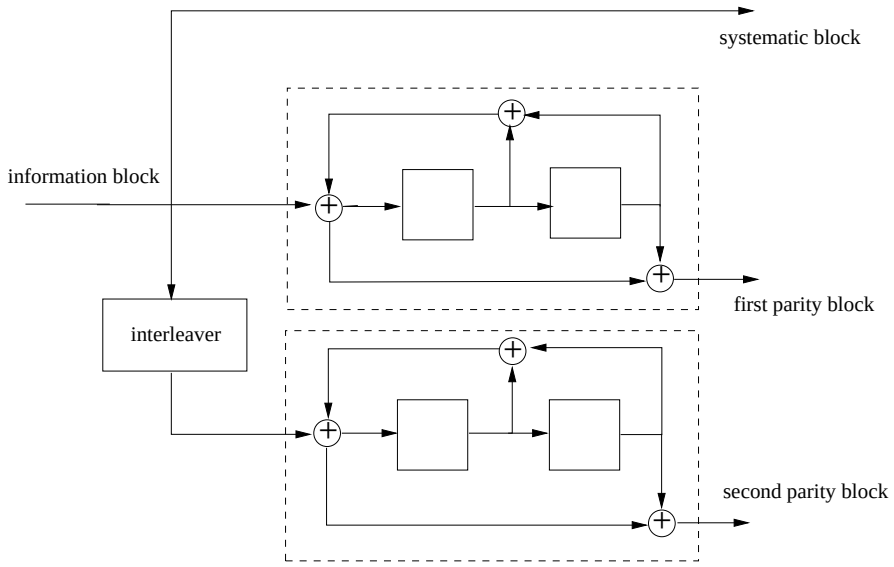


Figure 3: Rate 1/3 turbo code with component encoders $(5, 7)_{octal}$.

(LLR) of transmitted bits

$$\Gamma(d_k) = log\frac{Pr(d_k = 1|x^s, x^{1p})}{Pr(d_k = 0|x^s, x^{1p})}$$

using the received sequence x^s and x^{1p} where, for an additive white Gaussian noise channel and binary phase shift keying modulation scheme, the k^{th} element of the received sequence R is given as

$$R_k = (x_k^s, x_k^{1p}, x_k^{2p}) = (2s_k^s - 1 + n_k^s, 2s_k^{1p} - 1 + n_k^{1p}, 2s_k^{2p} - 1 + n_k^{2p}).$$

This overall LLR can be written as the sum of three components: LLR of the uncoded bit d_k at the decoder input, an a-priori term, and another information that does not depend on decoder input s^s, which is called as *extrinsic information*. This extrinsic information, L_{1e}, is passed to the second MAP decoder as an a-priori probability and together with x^s and x^{2p}, MAP decoding unit computes the log-likelihood ratio again. The extrinsic information from the second decoder is then fed back to the first decoder and the iterations proceed as the extrinsic information is exchanged between the decoders until a desired performance is achieved at which point a final decision is made based on the final log-likelihood ratio of each information bit. Note that in all these operations, all the likelihood ratio sequences and received sequences are suitably reordered and delayed. The iterative decoding algorithm

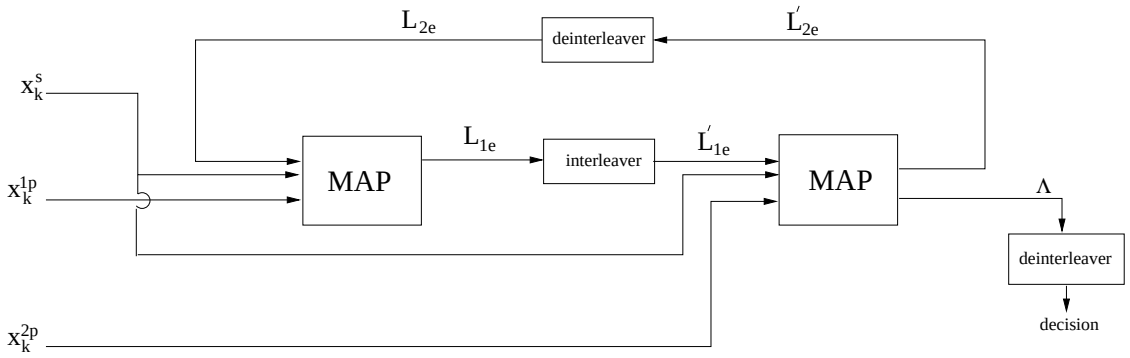


Figure 4: Block diagram of iterative turbo decoding.

is a sub-optimum algorithm, but empirical evidence suggests that this algorithm performs remarkably well and converges to the optimal decoding solution with high probability [17]. For the details of the iterative turbo decoding, see [77, 127].

In the Turbo coding scheme of Figure 3, the over all code rate is 1/3, which is too low for many applications. Using different rate component convolutional encoders and certain puncturing schemes, one can obtain higher rate turbo codes. Nevertheless, in general, turbo codes are low rate codes and require a considerable bandwidth expansion. In this case, turbo coded modulation scheme is a suitable alternative. In turbo coded modulation, a turbo code is concatenated with a higher order modulation scheme. There are various approaches for turbo coded modulation [13, 60, 128]. For example, the block diagram of the scheme in [60] is shown in Figure 5. In this scheme, the turbo coded bits are partitioned and directly mapped to a signal point in the constellation. The decoding of turbo coded modulation can also be accomplished using the suboptimum iterative decoding algorithm. In this case, we first compute the log-likelihood of the transmitted bits and then use these likelihood values as if they are the likelihoods of the observations from a BPSK modulation scheme over an AWGN channel in the iterative turbo decoding process [60].

The turbo coded modulation above can be employed for multiple antenna transmission [151]. In this case, after the data is partitioned into blocks of N_i bits and encoded by a binary turbo encoder, the resulting turbo-code stream are first set partitioned by a serial-to-parallel converter which results in parallel streams for each transmit antenna. Then, the

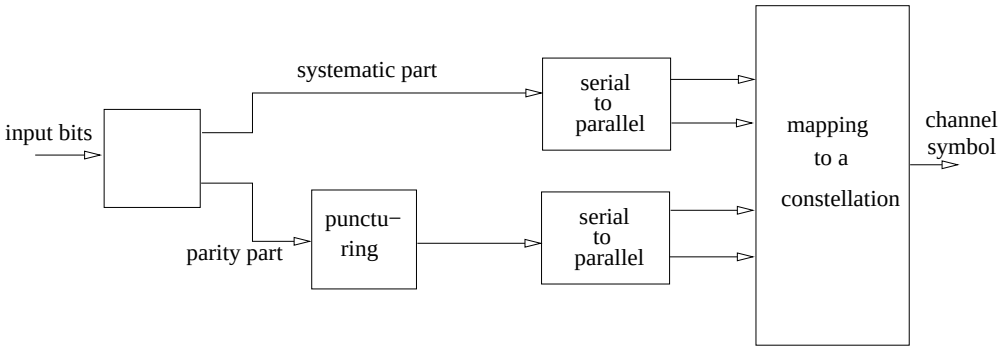


Figure 5: The block diagram of a turbo coded modulation scheme.

bits of each partition are mapped to the signal constellation and transmitted using different antennas. Since the transmission take place simultaneously at each antenna, one can attain high spectral efficiencies.

1.3 Cooperative Diversity

In the previous section, we assumed that at both sides of the communication system there are multiple antenna elements. On the other hand, because of the limited sizes of the portable devices, it might not be possible to deploy more than one antenna at the mobile nodes. Assume that there are a number of mobile users that are equipped with single antennas. If one user, say U_A, wants to communicate with another user U_B, the diversity order that can be achieved in the absence of any cooperation from other users is just 1, *i.e.*, there is no diversity gain. However, assume that there is genie node U_C who knows the signal transmitted by U_A,[1] and wishes to help U_A. In this case, U_A and U_C can cooperate to attain the performance of a double-transmit single-receive antenna system, providing a diversity order of 2. This diversity technique is referred to as the *user cooperative diversity* [138–140, 152].

In a wireless medium, noting that the transmissions for wireless relaying occur at different spatial locations, it is shown by Laneman et al. that one can obtain a *distributed spatial diversity* using relays [99, 101]. Valenti et. al. study a similar system where *macro*

[1]Since the channel between U_A and U_C is noisy, this may not be feasible, but for the time being, assume that the information transmitted from U_A can be recovered at U_C without errors

diversity can be obtained using relay terminals [171]. A simple coding scheme that achieves user cooperative diversity is presented in [90] where Hunter and Nosratinia consider a user cooperative scheme for two nodes each of which employs a rate compatible punctured convolutional code. If the node receives the other's data without error, it forwards the partner's bits to the base station, and otherwise, it transmits its own bits. Since the base station observes the data stream from independently fading channels, one can achieve spatial diversity.

The simplest system that can attain the cooperative diversity is perhaps the three-terminal relay channel introduced by [172]. In compliance with the above example, U_A becomes the source node, U_B becomes the destination node and U_C becomes the relay node. The capacity of the relay channel is investigated by El Gamal and Cover in [32,33], and it is shown that the information rate in a relay channel is higher than that is attained by a direct transmission from U_A to U_B. While there has been some efforts to design coding schemes for the relay channel to attain this capacity, the code design problem for the relay channels needs further investigation. In this dissertation we will propose a distributed channel coding method that can perform very close to the capacity limits of the relay channel.

1.4 Multiple Description Coding

A particular source coding method, known as multiple description coding (MDC), can be viewed as a joint source – channel coding technique. MDC generates multiple bitstreams, also called descriptions, of a source so that various quality levels of reconstruction can be obtained from any subset of the descriptions. The descriptions are transmitted over independent channels with the hope that upon the reception of all or some of the descriptions, a superior or an acceptable quality reconstruction is possible. This can be accomplished by introducing a certain amount of correlation between the individual descriptions. In an ideal transform coding, the aim is to represent the signal as efficiently as possible by removing all the redundancy. However, in that case, it is difficult to estimate the parts that are lost from those that remain. Introducing redundancy among the transform domain coefficients can provide robustness against such losses.

There are various methods of implementing MDC. A practical approach, known as multiple description scalar quantization (MDSQ), uses a pair of scalar quantizers that generates two indices (descriptions) of a source sample [170]. The quantization levels and index assignment are designed such that if only one index is received, the quantizer functions as a coarse quantizer, but if both indices are received, it functions as a fine quantizer. Another MDC scheme uses correlating transforms: Multiple Description Transform Coding (MDTC) [68,69,180]. In [180], a rotation matrix was applied to create pairwise correlation between the uncorrelated variables obtained using the Karhunuen-Loeve transform. This method is studied in a generalized framework in [68] and [120]. Other techniques of multiple description coding uses the lapped orthogonal transform framework [31] and projections onto convex sets framework [29].

Although there are various ways of multiple description encoding schemes [29,31,68,170], in this disseration, in order to produce two correlated descriptions of the source samples, we will consider the use of multiple description scalar quantization (MDSQ) introduced in [170] and Multiple Description Coding using pairwise correlating transforms (MDTC) first proposed by [180]. In this and the following subsections, we will briefly describe the MDSQ and MDTC, respectively.

1.4.1 Multiple Description Scalar Quantization

The block diagram of the MDSQ is shown in Figure 6. As depicted, each source sample is input to a cascade connection of a quantizer block $q(\cdot)$ and index assignment block $a(\cdot)$.

Let X denote a stationary and ergodic random process with zero mean and variance σ_X^2 that generates our source symbols. The encoder of an MDSQ operates as follows: First, the source sample is mapped to an index l using an N-level quantizer $q(\cdot)$ which is defined using a threshold vector $t = (t_0, t_1, \cdots, t_N)$ such that $q(x) = i$ if $x \in [t_{i-1}, t_i), i = 1, 2, \cdots, N$. Quantization is followed by an index assignment $a(\cdot)$ by which each index l is mapped to a pair of indices $(i, j) \in \mathcal{J}$ where $\mathcal{J} \subseteq \mathcal{I}_1 \times \mathcal{I}_2$ and the individual indices $i \in \mathcal{I}_1 = \{1, 2, \cdots, M_1\}$ and $j \in \mathcal{I}_2 = \{1, 2, \cdots, M_2\}$. We are considering balanced MDSQ's, therefore we assume $M_1 = M_2 = M$ where M is the number of indices for each description.

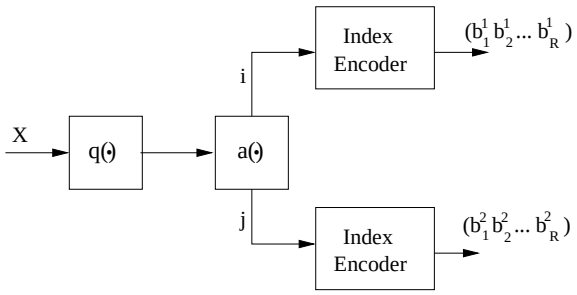


Figure 6: Block diagram for the MDSQ.

In this case, the rate is $R = log_2(M)$ bpss (bit per source sample) for each description. The MDSQ is mainly devised for an "on-off" channel model, however, it was shown in [183] that one can apply the multiple description quantizer for Rayleigh fading channels as well.

The generation of the indices is an important task in the design of MDSQ schemes. In [170], for on-off channels, design algorithms to generate good index assignments resulting in the optimum (exponential side and central distortion) decay rates have been presented. An example of an index assignment obtained using these techniques is illustrated in Table 1. The numbers in the table entries correspond to the cells of a quantizer, which are numbered $1, 2, \cdots, N$, in increasing order from left to right. The coordinates of the entry in which l is located is assigned to the quantizer cell l. For example, for the index assignment depicted on Table 1.b, if the source symbol lies in the interval corresponding to the 13^{th} quantizer cell, then the index pair $(5, 6)$ is transmitted. We note that an index assignment scheme obtained with the techniques of [170] consists of $(2k+1)$ diagonals, $k = 1/2, 1, 2, \cdots, 2^R - 2$ and the redundancy can be controlled by varying the number of the non central-diagonals covered by the index assignment. In general, the amount of correlation between the indices (i, j) decreases as the number of diagonals in the index assignment is increased. For a given index assignment scheme, it is easy to compute the conditional probabilities $P(i|j)$ or $P(j|i)$ using the threshold values of the quantization and input statistics.

1.4.2 Multiple Description Transform Coding

A different technique for multiple description coding was proposed in [180] by Wang *et. al.* and [70]. This method makes use of linear transforms to introduce correlations between the

Table 1: Examples of MDSQ index assignment for $R = 3\, b/s$, $M_1 = M_2 = 8$.

i/j	1	2	3	4	5	6	7	8
1	1	2						
2		3	4					
3			5	6				
4				7	8			
5					9	10		
6						11	12	
7							13	14
8								15

(a)

i/j	1	2	3	4	5	6	7	8
1	1	2						
2	3	4	6					
3		5	7	8				
4			9	10	11			
5					12	14		
6					13	15	16	
7						17	18	20
8						7	19	21

(b)

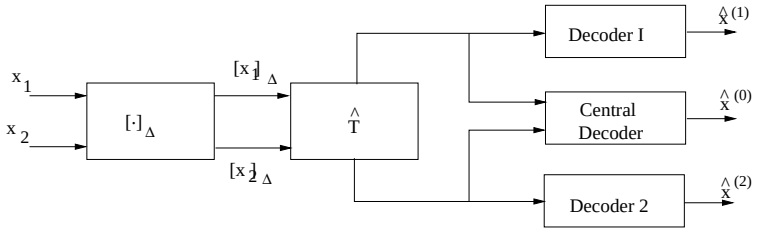


Figure 7: Block diagram for the MDTC.

pairs of random variables. Unlike MDSQ where a scalar source sample is described by the indices produced by two separate quantizers, MDTC generates multiple streams of a pair of uncorrelated source samples (or transform coefficients), first by creating some redundancy within the pairs using a correlating transform and then quantizing the resulting pair. Such correlation reinforces the robustness against coefficient losses, since one can estimate the lost coefficients from those that are received using the correlation. Consider the transmission of a pair of samples $X = (X_1, X_2)^T \in \mathcal{C}$ that denote the source descriptions that are independent real-valued random variables. The block diagram of the MDTC is depicted in Figure 7. In the figure, $[\cdot]_\Delta$ represents the quantization operator that quantizes the samples to the nearest multiple of Δ. We note that the quantization is performed before the transform, T, since quantization after the continuous transform, *i.e.*, $Y_\Delta = [Tx]_\Delta$, results in non-square partition cells that are suboptimal [70]. Therefore, the source pairs are first quantized and then using a discrete version of the continuous transform, $\hat{T}$, the output vectors $(Y_1, Y_2)^T$ are calculated. The discrete transform $\hat{T}$ is achieved in two steps: first, the continuous transform T is represented by a product of upper- and lower-triangular matrices, e.g., using LDU decomposition, and second, the transformation is performed with

intermediate roundings as

$$\hat{T}(x) = [T_1[T_2[T_3 x]_\Delta]_\Delta]_\Delta$$

We can adjust the entries of the transform matrix to achieve any desired *redundancy*, or ρ, which is defined to be the additional rate in excess of the minimum average rate R^* that achieves the minimum average distortion in redundancy rate distortion (RRD) curve. Simple analysis gives [70]

$$\rho = \frac{1}{4} log \frac{(a_{11}^2\sigma_1^2 + a_{12}^2\sigma_2^2)(a_{21}^2\sigma_1^2 + a_{22}^2\sigma_2^2)}{\sigma_1^2\sigma_2^2}$$

For brevity, we will not go into the details of the MDTC, instead, we refer the interested reader to [70]. Here, we wish to focus on the computation of conditional probabilities, *i.e.*, the probability of second description given the first description, $P(Y_{k_2}|Y_{k_1})$, $(k_1, k_2) \in \mathcal{C}_{k_1,k_2}$, where $\mathcal{C}_{k_1,k_2}$ is the set that is mapped to $(k_1, k_2)\Delta$. The probability of $x \in \mathcal{C}_{k_1,k_2}$, p_{k_1,k_2}, is given by

$$P((Y_1, Y_2) = (k_1, k_2)\Delta) = p_{k_1,k_2} = \int_{\mathcal{C}_{k_1,k_2}} f_{x_1,x_2}(x_1, x_2) dx_1 dx_2 \tag{3}$$

The conditional probability of one of the descriptions given the other follows as

$$P(Y_2 = K_2\Delta | Y_1 = K_1\Delta) = \frac{p_{K_1,K_2}}{\sum_l p_{K_1,l}}$$

If the source pairs $x = (x_1, x_2)^T$ have zero mean Gaussian components with variance σ_1^2 and σ_2^2, respectively, then the expression in (3) have a closed form expression in terms of standard error function

$$p_{k_1,k_2} = \frac{1}{4}(erf((j_1 + \frac{1}{2})\Delta/\sqrt{2\sigma_1^2}) - erf((j_1 - \frac{1}{2})\Delta/\sqrt{2\sigma_1^2}) \cdot (erf((j_2 + \frac{1}{2})\Delta/\sqrt{2\sigma_2^2}) - erf((j_2 - \frac{1}{2})\Delta/\sqrt{2\sigma_2^2}))$$

where $(j_1, j_2) : \hat{T}((j_1, j_2)\Delta) = (k_1, k_2)\Delta$. We will later show in Chapter 5 how the conditional probabilities can be utilized in the decoding to improve the performance by making use of the residual information due to correlation between the multiple descriptions.

1.5 Distributed Detection

The detection of an event based on noisy observations is a standard problem in statistics, communications and radar signal processing. For example, in radar detection theory, the

detection problem is to make a decision whether there is a signal of interest present in the noisy observations or not. This is nothing but a binary hypothesis-testing problem where there are two possible states of nature (null hypothesis H_0, alternative hypothesis H_1) associated with two different probability distributions on the observation space. At the end of the statistical test, a *false alarm* occurs if the underlying hypothesis is H_0 while the test chooses H_1, a *detection* occurs if the underlying hypothesis is H_1 and the test also chooses in H_1, and a *missed detection* occurs if the test result is H_0 while the underlying hypothesis is H_1. The probabilities of these occurrences is central in designing the optimal detection rules in the detection theory.

In this dissertation, we deal with the detection problem in a network of spatially separated sensor nodes, which is referred to as *distributed detection*. In distributed detection, each sensor sends a summary of its observation to a fusion center where a global decision is made. In classical multisensor detection, such as in the case of radar and sonar applications, it is assumed that all local sensors can communicate all their data to the central processor in which case optimal detection can be performed. However, in typical wireless sensor networks, because of the bandwidth and energy limitations as well as the presence of noisy channels, such *centralized detection* is not feasible. Therefore, instead of transmitting the raw data, it is more appropriate to perform a local detection process at each sensor node and then send the local decisions (possibly consisting of a few bits) to the fusion center. The fusion center, after collecting all decisions from all sensors, performs a final decision on the hypothesis under investigation.

Several distributed detection algorithms have been investigated in the literature [39, 83, 91, 119, 121, 129, 130, 155–158, 167]. In this section, we summarize a few of these algorithms. Optimal distributed detection algorithms have been focused on optimality under the Neyman-Pearson and Bayesian detection criteria, which lead in standard situations to likelihood ratio tests at the individual sensors and at the fusion center. The optimum distributed signal detection methods under these criteria satisfy a set of coupled constraints. In nonparametric detection, the classical requirement of parametric statistical models for the signal and noise observations is relaxed and the competing composite hypotheses are stated

in terms of statistical models for the observations that are not tied down by probability density function models from specific parametric families. Distributed detection methods based primarily on signs and ranks of observations have been investigated [91, 110, 133]. These algorithms, in many cases, provide constant false-alarm rate (CFAR) that is independent of the exact noise distribution assuming a symmetric distribution. Viswanathan and Ansari [176] studied distributed nonparametric detection that uses Wilcoxon signed-rank test statistics while Nasipuri and Tantaratana [3] used the Wilcoxon statistics to generate multi-bit sensor decisions. Han *et al.* [78] studied the sign detector and the dead-zone detector in a distributed context. Some generalizations of the sign and dead-zone detector have been studied by Al-Ibrahim and Varshney [112] and by Hussaini *et al.* [44]. Blum [18] showed that nonparametric detection based on signs and ranks can be constructed to provide locally most powerful (LMP) performance under a given nominal model for distributed signal detection problems.

CFAR detection is generally based on parametric models with an unknown clutter-plus-noise power parameter, which is often encountered in radar and sonar applications. CFAR techniques determine a new threshold setting for each cell probed to achieve a constant false-alarm probability. The threshold is determined using a set of reference observations, typically obtained from a nearby cell in space. Several distributed CFAR methods have been proposed in [6, 10, 11, 168].

Robust detection deals with the robustness against uncertainties in the statistical models used for detection design [91, 97, 110, 123]. A basic principle of robust detection is to design optimum detectors based on certain least favorable models. Veeravalli *et al.* [179] considered minimax robustness for distributed detection networks both with and without fusion centers. An asymptotic version of this problem is considered in [43].

In all the above methods, the detectors are fixed sample-size. In sequential distributed detection, the number of observations used is a random quantity that depends on the observations themselves [91]. As soon as enough data is collected to meet a given reliability requirement, a decision is finalized. This is important in time-critical applications and when data acquisition is costly. Sequential detection can be conducted either at the fusion

center or at the individual sensors. In the former, the nodes pass their local decisions to the fusion center, which performs a sequential test [88, 174, 177]. In the latter, the sensors can perform local sequential test without the need for a fusion center [40, 178]. Another version of running sequential detection is referred to as quickest detection where the detection of an abrupt change at an unknown time is desired [40,132,175]. More recently, a capacity-constraint distributed detection algorithms have been investigated [42, 111, 148]. Chamberland and Veeravalli [94] showed that under certain conditions, for an N-sensor network with a capacity constraint of N bits per time unit, having each sensor transmitting one bit is optimum. Thomopoulos and Zhang investigates the distributed detection in the case of non-ideal channels [165]. In [42], Duman and Salehi consider the distributed detection over multi access channels where the fusion center gathers the decisions from local sensors via a multi-access channel.

All the aforementioned algorithms assume that *the sensor decision statistics, either quantized or at full precision, can be transmitted error-free to the fusion center.* Even though this assumption is valid in traditional sensor networks such as radars and sonar [122], it is impractical in wireless sensor networks where wireless links are subject to fading and interference. Furthermore, due to bandwidth and energy constraints, the use of powerful error correction codes is not viable. Recently, Chen et al. introduced channel-based decision fusion for a parallel network of sensors linked with fading channels [24, 25, 119]. Assuming parallel configuration, the authors incorporate the effect of fading in the detection process, and derive optimal fusion rules and some alternative fusion rules based on diversity combining techniques. In [105], a similar decision fusion for a multihop transmission is considered. While the performance of the decision fusion based on some suboptimal methods are evaluated in these work, the optimality of the decision rules at local sensors and at the fusion center, and optimal designs are not considered. Recently, Chen and Willet have sown that optimal local decisions that minimize the error probability at the fusion center becomes a likelihood-ratio test (LRT) under some particular constraints on the fusion rule [26].

1.6 Contributions of the Dissertation

In Chapter 2, we study antenna subset selection for systems with MIMO antenna antennas. Assuming that (i) the channel is characterized by quasi-static Rayleigh flat fading, and the sub-channels fade independently, (ii) the channel state information (CSI) is exactly known at the receiver, (iii) the selection is available only at the receiver, and it is based on the instantaneous signal-to-noise ratio at each receive antenna, and (iv) space-time codes are used at the transmitter, we analyze the system performance by deriving explicit upper bounds on the pairwise error probability (PEP). The performance analysis shows that (i) by selecting the set of antennas that observe the largest instantaneous signal-to-noise ratio, one can achieve the same diversity gain as the one obtained by using all the receive antennas, provided that the underlying space-time code has full spatial diversity, and (ii) in the case of rank-deficient space-time codes, the diversity gain may be dramatically reduced when antenna selection is used. In addition, based on the upper bounds derived, we describe code design principles suitable for antenna selection. Specifically, for systems with two transmit antennas, we design space-time codes that perform better than the known ones when antenna selection is employed. We present numerical examples and simulation results that validate our analysis and code design principles.

We note that in Chapter 2, antenna selection for multiple antenna transmission systems is studied under the assumption that the subchannels between antenna pairs fade independently. In Chapter 3, we consider the performance of such systems when the subchannels experience correlated fading. We again assume that the channel state information is available only at the receiver, and the antenna selection is performed only at the receiver, and the selection is based on the instantaneous received signal power. We quantify the effects of channel correlation on the diversity and coding gain when the receiver system uses all or a subset of the antennas. Theoretical results indicate that the correlations in the channel does not degrade the diversity order provided that the channel is full-rank. However, it does result in some performance loss in the coding gain. Furthermore, for non-full-rank channels, the diversity order of the system degrades significantly and is determined by the rank of the channel correlation matrix.

The wireless channels in Chapters 2 and 3 are modeled as a frequency flat fading channel. In an effort to increase the transmission rate, one needs to reduce the symbol duration which may result in frequency selective fading. In such cases, orthogonal frequency division multiplexing can be employed to increase the symbol duration to convert the frequency selective fading channel to a flat fading one. Chapter 4 deals with the antenna selection for space-time coded orthogonal frequency division multiplexing (OFDM) systems that employ multiple transmit and receive antennas. Assuming the CSI is known at the receiver, the selection is based on the instantaneous signal-to-noise ratio at each receive antenna averaged over all carrier frequencies. We again analyze the performance of such systems and derive closed-form upper bounds on the PEP. It turns out that it is difficult to make remarks about the diversity order since the expressions are not simple. However, for the special case of double transmit diversity over a channel order of two, we explicitly show that (i) with single antenna selection out of $N = 2$ receive antennas, and (ii) with the selection of $N-1$ antennas out of N receive antennas, one can attain the same diversity order as that attained by the full-complexity system. For other cases, numerical results indicate that diversity order is preserved with antenna selection. Therefore, the proposed antenna selection technique can be used to reduce the implementation cost of the multiple-input multiple-output OFDM systems while resulting in no degradation in the asymptotic performance of the system.

In Chapter 5, we propose a joint source-channel coding scheme for wireless communication systems with multiple transmit and receive antennas. The source coder is realized by a multiple description encoder that generates multiple bit streams of the same source. Each description is then separately turbo coded and transmitted using multiple antennas. For the receiver, we describe a suitable iterative joint source-channel decoding technique that exploits the correlations between the descriptions. Extensive numerical results illustrate that the performance of the proposed system is superior to its single description counter parts. The proposed scheme can be imagined as a combination of two different diversity schemes: (i) the source diversity provided by the multiple correlated descriptions and (ii) the channel diversity provied by the MIMO antenna system.

Chapter 6 deals with coding for wireless relay channels. We propose a turbo coded

modulation technique for fading relay channels. According to this scheme, the source sends turbo coded bits to the relay and the destination, while the relay, which has full-duplex capability, forwards simultaneously the estimate for the previous turbo coded block to the destination after decoding and re-encoding while it is receiving the signals transmitted from the relay node. The destination observes the superposition of the transmitted codewords and uses an iterative soft-input soft-output decoding algorithm to estimate the transmitted information. Various decoding schemes are presented and compared to each other and also to the previous results. The performance of the system is within $1-2$ dB of capacity at a bit error rate of 10^{-5}. While the proposed scheme is presented for the case of a single relay node, it can readily be generalized to multiple relay case.

In Chapter 7, we address the binary distributed detection problem for wireless sensor networks. Contrary to the assumption that the links between the sensor nodes are noiseless, we assume that the links are subject to multi-path fading and therefore, the transmitted signals from the local sensors are received corrupted at the destination sensor nodes. Several detection strategies for different network configurations are analyzed and their performance are compared to each other. We classify two main approaches: (i) digital approach where the sensor nodes first make a local decision (consisting of a single bit or multiple bits) and then transmit those decisions to the destination nodes, (ii) analog approach where the locally observed data or the sufficient statistics is directly transmitted to destination node. Two network configurations are considered: serial and parallel. The resulting analysis is used for optimizing the distributed detection schemes for sensor networks interconnected with fading links. It turns out that the optimal designs for the digital approach becomes impractical because of the computational burden of the exhaustive search required to determine thresholds at all sensor nodes. In the analog approach, we analyze the detection performance according to Neyman-Pearson lemma and show that the optimal design reduces to determination of optimal power allocation scheme. This approach alleviates the design problem since one only needs to determine a global decision rule for the fusion center and a suitable power allocation for the local sensors which can easily be found analytically or via numerical optimization techniques. Furthermore, the simulation results illustrate that

for a given energy budget, the analog approach has a better detection performance than a system where only one-bit decisions are allowed. In this chapter, we also propose the use of sequential detection for wireless sensor networks under consideration.

CHAPTER II

ANTENNA SUBSET SELECTION FOR MULTIPLE ANTENNA TRANSMISSION

In this chapter, we will study the performance of receive antenna selection for a system equipped with MIMO antennas. This chapter deals with the case of independently and identically distributed fading among the subchannels between transmit-receive antenna pairs. The organization of the chapter is as follows: In Section 2.1 we summarize the related work. Section 2.2 describes the multiple antenna channel model, and summarizes important results on pairwise error probability. In Section 2.3, we compute the upper bounds on pairwise error probability for space-time codes when antenna selection is employed for both full-rank and rank-deficient space-time codes. In Section 2.4, we derive a tighter upper bound for systems with double transmit antenna diversity. We consider the space-time code design with antenna selection in Section 2.5. In Section 2.6, we present several numerical examples and simulation results that validate our analysis and the new code design principles. We provide the conclusions in Section 2.7.

2.1 Introduction

Deploying multiple antennas to the wireless systems increases the implementation cost and complexity due to the need for multiple RF chains. Recently, there has been a considerable effort in exploring multiple input multiple output (MIMO) systems that significantly reduce this complexity, but at the same time provide similar capacity and performance improvements. A promising technique to achieve this goal is to select a subset of antennas at the transmitter and/or receiver [62, 113, 135]. To see how antenna selection can be used, consider the case of single antenna selection at the receiver when there are say 10 antenna elements. Assuming that the fading is slow, the received signal power can be monitored periodically, and only the signal of the receive antenna observing the largest instantaneous

SNR can be fed to the RF chain for processing. Thus we need only one RF chains and significant reduction in hardware costs can be attained while reaping the benefits of MIMO signaling.

The capacity of MIMO systems with antenna selection (only at the receiver) is considered in [113]. The selection is based on the capacity, *i.e.*, those antennas that achieve the largest instantaneous capacity are selected. The authors evaluate upper bounds on the capacity of the system and conclude that one can achieve a capacity very close to that of the full-complexity system as long as the number of antennas selected is greater than or equal to the number of transmit antennas. In [62] and [135], transmit antenna selection is studied for systems where limited feedback on the channel state information is available to the transmitter. In these systems, channel capacity is used as the optimality criterion and the selection is performed by an exhaustive search. Gorokhov proposes a suboptimal selection algorithm in [66] that decreases the computational complexity significantly. An antenna selection method seeking the minimization of the error rate using linear receivers is considered in [82].

In [65], the authors considered the use of antenna selection in conjunction with orthogonal space-time block codes. They present antenna selection algorithms for cases when exact channel knowledge or statistical channel knowledge is available. For the case of exact channel knowledge, the expressions for the average SNR and the outage capacity improvement are derived assuming that the selection criterion used is the maximization of the channel Frobenius-norm. This selection criterion is equivalent to minimizing the error probability for the case of space-time block codes. Using the outage probability analysis, the authors hint that the diversity gain is preserved for this system. However, they do not explicitly provide an analysis that includes the evaluation of the pairwise error probability for the system with antenna selection. Furthermore, these results are valid only for orthogonal space-time codes and cannot be directly applied to the case of more general space-time codes. In [57], Ghrayeb and Duman present an approximate analysis for the pairwise error probability of the space-time coding system using antenna selection. They show that the

diversity order available is maintained. However, this analysis is based on several approximations. Thus, it is not an explicit proof. Other work on antenna selection is also reported in [63, 64, 98, 117, 181].

In this chapter, we present a comprehensive theoretical performance analysis for MIMO systems over quasi-static Rayleigh fading channels that use antenna selection at the receiver. We base our selection criterion on the maximization of the received signal power. Under certain cases this selection criterion may be optimal in the sense that it may achieve the maximum channel capacity, *e.g*, for the case of single antenna selection at the receiver. The pairwise error probability (PEP) will be central in our approach. We calculate the diversity and coding gains by computing upper bounds on the PEP. For the case of single antenna selection, we present a performance analysis based on the PEP, and demonstrate that the diversity gain with antenna selection is preserved for space-time codes with full spatial diversity. Since it is essential to employ full-rank space-time codes to make sure the diversity order is not reduced with antenna selection, it may be beneficial to use full-rank full-rate space-time code designs recently proposed in [51, 52, 108]. We also study the performance bounds when the space-time codes do not achieve full spatial diversity, and show that the diversity gain degrades substantially when antenna selection is employed compared to the full-complexity system. Furthermore, we present the pairwise error probability analysis when more than one antenna is selected, and generalize our results. We also compute tighter upper bounds on the pairwise error probability for the case of double transmit diversity systems.

An immediate consequence of the performance analysis is the development of code design principles for space-time codes suitable for the systems employing antenna selection. Based on the bounds on the pairwise error probability we propose two simple design criteria. In particular for double transmit and double receive diversity systems, we design space-time codes that perform better than the known ones when antenna selection is used. For example, we design a 2 bits/sec/Hz 8-state space-time code for two transmit antennas employing $4-$ary Phase Shift Keying (PSK) modulation. We show that, while achieving the same performance with 8-state code in [163] for the full-complexity system, the new

code provides about 0.7 dB performance improvement when antenna selection is employed at the receiver.

2.2 Channel and Signal Model

We consider a single user communication system where the transmitter has M antennas and the receiver has N antennas. Each receive antenna observes a noisy superposition of the M transmitted signals corrupted by Rayleigh flat fading. The sub-channels between the transmit/receive antenna pairs are assumed to be independent identically distributed (i.i.d.). The signal at the n^{th} receive antenna element at time t, x_{tn}, is given by

$$x_{tn} = \sqrt{\rho/M} \sum_{m=1}^{M} h_{nm} s_{tm} + w_{tn},\ t = 1, 2, \cdots, l \tag{4}$$

where h_{nm} is the complex-valued channel gain from the m^{th} transmit antenna to the n^{th} receive antenna, and w_{tn} is the additive noise at the n^{th} receive antenna. Both h_{nm} and w_{tn} $\sim \mathcal{CN}(0, 1)$. The transmitted signals, s_{tm}, can be chosen from any signal constellation. We assume that the average energy of the transmitted signal at time t is normalized to unity over M antennas so that ρ is the expected signal-to-noise ratio at each receive antenna. We can rewrite (4) in vector form as

$$\mathbf{X} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{S} + \mathbf{W} \tag{5}$$

where $\mathbf{X}$ is the $N \times l$ received signal vector, $\mathbf{S}$ is the $M \times l$ transmitted signal vector, $\mathbf{H}$ is the $N \times M$ channel transfer matrix, and $\mathbf{W}$ is the $N \times l$ additive white Gaussian noise vector. We assume that the CSI, *i.e.,* $\mathbf{H}$, is known at the receiver, but not at the transmitter.

The PEP conditioned on $\mathbf{H}$ is given by [163]

$$P(\mathbf{S} \to \hat{\mathbf{S}}|\mathbf{H}) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\rho}{4M}} \parallel \mathbf{H}\mathbf{S} - \mathbf{H}\hat{\mathbf{S}} \parallel\right) \tag{6}$$

$$\leq \exp\left(-\frac{\rho}{4M} \parallel \mathbf{H}\mathbf{S} - \mathbf{H}\hat{\mathbf{S}} \parallel^2\right). \tag{7}$$

Defining the codeword difference matrix $\mathbf{B} = \mathbf{S} - \hat{\mathbf{S}}$ and $\mathbf{A} = \mathbf{B}\mathbf{B}^H$, and denoting $\text{tr}\{\cdot\}$ as

the trace operator, we can write

$$\begin{aligned} \| \mathbf{HB} \|^2 &= \mathrm{tr}\{\mathbf{HBB}^H\mathbf{H}^H\} \\ &= \mathrm{tr}\{\mathbf{HAH}^H\} \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} \lambda_m |\beta_{nm}|^2 \end{aligned} \tag{8}$$

where the last equality follows by using the eigenvalue decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ with $\boldsymbol{\Lambda}$ being a diagonal matrix whose entries (λ_m) are the eigenvalues of $\mathbf{A}$, and $\boldsymbol{\beta} = \mathbf{HU}$ where β_{nm} are independent Gaussian random variables. The average PEP for a Rayleigh fading channel is obtained by averaging the conditional PEP over the statistics of $\mathbf{H}$ resulting in

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) &= E_{\mathbf{H}}\{P(\mathbf{S} \to \hat{\mathbf{S}}|\mathbf{H})\} \\ &\leq \left(\frac{\rho}{4M}\right)^{-Nr} \left(\prod_{m=1}^{r} \lambda_m\right)^{-N} \end{aligned} \tag{9}$$

where $r = \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$. From this expression, we see that the diversity gain of the coded system is Nr and that the coding gain is $(\prod_{m=1}^{r} \lambda_m)^{1/r}$.

2.3 Upper Bounds on PEP with Antenna Selection

In this section, we derive upper bounds on the pairwise error probability for the Rayleigh fading channel. We start with the case when only one antenna is selected, and then generalize the results to the selection of more than one antenna at the receiver.

2.3.1 Pairwise Error Probability

We first consider the case when only one antenna is selected at the receiver. The upper bound on the conditional PEP in (7) is then given by

$$P(\mathbf{S} \to \hat{\mathbf{S}}|\hat{\mathbf{r}}) \leq \exp\left(-\frac{\rho}{4M}\hat{\mathbf{r}}\mathbf{BB}^H\hat{\mathbf{r}}^H\right)$$

where $\hat{\mathbf{r}}$ is the row of $\mathbf{H}$ having the maximum Frobenius-2 norm. In order to obtain the average PEP, we simply evaluate the expected value of this upper bound with respect to the distribution of $\hat{\mathbf{r}}$. That is,

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) &\leq E_{\hat{\mathbf{R}}}\{\exp(-\frac{\rho}{4M}\mathbf{r}\mathbf{B}B^H\mathbf{r}^H)\} \\ &= \int_{\mathcal{C}^M} \exp\left(-\frac{\rho}{4M}\mathbf{rBB}^H\mathbf{r}^H\right) f_{\hat{\mathbf{R}}}(\mathbf{r})d\mathbf{r} \end{aligned}$$

where $\mathcal{C}^M$ is the M-dimensional complex space and $f_{\hat{\mathbf{R}}}(\mathbf{r})$ denotes the probability density function (pdf) of $\hat{\mathbf{r}}$. In order to compute $f_{\hat{\mathbf{R}}}(\mathbf{r})$, we introduce the auxiliary event $\mathcal{F} = \{n^{th} \text{ row has the largest norm}\}$. Then, we can write

$$\begin{aligned} f_{\hat{\mathbf{R}}}(\mathbf{r}) &= f_{\mathbf{R}_n}(\mathbf{r}|\mathcal{F}) \\ &= \frac{P(\mathcal{F}|\mathbf{R}_n = \mathbf{r}) f_{\mathbf{R}_n}(\mathbf{r})}{P(\mathcal{F})} \\ &= \frac{P(\|\mathbf{r}_1\|^2 < \|\mathbf{r}\|^2, \cdots, \|\mathbf{r}_{n-1}\|^2 < \|\mathbf{r}\|^2, \|\mathbf{r}_{n+1}\|^2 < \|\mathbf{r}\|^2, \cdots, \|\mathbf{r}_N\|^2 < \|\mathbf{r}\|^2) f_{\mathbf{R}_n}(\mathbf{r})}{P(\mathcal{F})} \\ &= \frac{P(\|\mathbf{r}_1\|^2 < \|\mathbf{r}\|^2)^{N-1} f_{\mathbf{R}_n}(\mathbf{r})}{P(\mathcal{F})} \end{aligned} \tag{10}$$

where we use Bayes' rule, and the fact that the $\mathbf{r}_i$'s are independent identically distributed. Since all rows have the same statistics, we have

$$P(\mathcal{F}) = \frac{1}{N}. \tag{11}$$

Furthermore, using the statistics of $\mathbf{H}$, we can say that the squared norm of each row, (*i.e.*, $\|\mathbf{r}_i\|^2 = \sum_{m=1}^{M} |h_{im}|^2,\ i \in \{1, \cdots, N\}$), is a Chi-square random variable of order $2M$, $\chi^2(2M)$. Hence, we can write

$$P(\|\mathbf{r}_i\|^2 < \|\mathbf{r}\|^2) = 1 - e^{-\|\mathbf{r}\|^2} \sum_{m=0}^{M-1} \frac{\|\mathbf{r}\|^{2m}}{m!}. \tag{12}$$

Finally, the term, $f_{\mathbf{R}_n}(\mathbf{r})$ in (10), is the unconditional pdf of the n^{th} row, which is simply

$$f_{\mathbf{R}_n}(\mathbf{r}) = \frac{1}{\pi^M} \exp(-\|\mathbf{r}\|^2). \tag{13}$$

Substituting (11), (12) and (13) into (10), we obtain the desired pdf as

$$f_{\hat{\mathbf{R}}}(\mathbf{r}) = N \left(1 - e^{-\|\mathbf{r}\|^2} \sum_{m=0}^{M-1} \frac{\|\mathbf{r}\|^{2m}}{m!}\right)^{(N-1)} \frac{1}{\pi^M} e^{-\|\mathbf{r}\|^2}.$$

Hence, the upper bound on the average PEP follows as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N \int_{\mathcal{C}^M} e^{-\frac{\rho}{4M} \mathbf{r}\mathbf{B}B^H\mathbf{r}^H} \left(1 - e^{-\|\mathbf{r}\|^2} \sum_{m=0}^{M-1} \frac{\|\mathbf{r}\|^{2m}}{m!}\right)^{(N-1)} \frac{1}{\pi^M} e^{-\|\mathbf{r}\|^2} d\mathbf{r}.$$

We can further simplify this expression by using the singular value decomposition of $\mathbf{B}\mathbf{B}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and by applying the change of variable $\mathbf{z} = \mathbf{r}\mathbf{U}$ as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N \int_{\mathcal{C}^M} e^{-\frac{\rho}{4M} \sum_{i=1}^{M} \lambda_i |z_i|^2} \left(1 - e^{-\|\mathbf{z}\|^2} \sum_{m=0}^{M-1} \frac{\|\mathbf{z}\|^{2m}}{m!}\right)^{(N-1)} \frac{1}{\pi^M} e^{-\|\mathbf{z}\|^2} d\mathbf{z}. \tag{14}$$

It is not easy to evaluate the integral in (14) to obtain a closed form expression. However, it is possible to further upper bound this expression to simplify the analysis as we will show in the next section.

2.3.2 Simplified Upper Bounds on the PEP for Single Antenna Selection

In this section, we will derive simple expressions for the upper bounds on the PEP. In Section 2.3.2.1, we will consider the case when the space-time codes achieve full-spatial diversity, *i.e.,* the rank of the codeword difference matrices, $\mathbf{B}$, is M for all codeword pairs. For such codes, we will show that the diversity order achieved with antenna selection is the same as that of the full-complexity system. Then, in Section 2.3.2.2, we perform approximate analysis of the PEP for rank-deficient space-time codes, and show that the diversity order is dramatically reduced with antenna selection.

2.3.2.1 Full-rank Space-time Codes with Antenna Selection

Let us simplify (14) by the change of variables $z_i = \sigma_i e^{j\theta_i}$, which yields

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 2^M N \int_0^\infty \cdots \int_0^\infty e^{-\frac{\rho}{4M}\sum_{i=1}^{M}\lambda_i\sigma_i^2} \left(1 - e^{-(\sigma_1^2+\cdots+\sigma_M^2)} \times \sum_{m=0}^{M-1} \frac{\left(\sigma_1^2+\cdots+\sigma_M^2\right)^m}{m!}\right)^{(N-1)} e^{-(\sigma_1^2+\cdots+\sigma_M^2)}\sigma_1\cdots\sigma_M d\sigma_1\cdots d\sigma_M. \tag{15}$$

We would like to find a simpler expression or bound that directly provides information about the diversity order and coding gain with antenna selection. To this end, we need following auxiliary lemma:

Lemma 1: Define

$$g(v) = 1 - e^{-v}\sum_{m=0}^{M-1}\frac{v^m}{m!}.$$

Then,

$$g(v) \leq \frac{v^M}{M!} \text{ for } v > 0.$$

Proof: Observing that $g(v)$ is the incomplete Gamma function, the proof follows easily,

i.e.,

$$g(v) = \frac{1}{(M-1)!}\int_0^v u^{M-1}e^{-u}du \leq \frac{1}{(M-1)!}\int_0^v u^{M-1}du = \frac{v^M}{M!} \qquad \blacksquare$$

Since the the value of the integrand in (15) is always greater than zero, we can further upper bound the right hand side of (14) by substituting $\frac{v^M}{M!}$ in place of $g(v)$ with $v = \sigma_1^2 + \cdots + \sigma_M^2$ to obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2^M N}{(M!)^{N-1}} \int_0^\infty \cdots \int_0^\infty e^{-\frac{\rho}{4M}\sum_{i=1}^M \lambda_i \sigma_i^2} \left(\sigma_1^2 + \cdots + \sigma_M^2\right)^{M(N-1)} \times e^{-(\sigma_1^2+\cdots+\sigma_M^2)} \sigma_1\sigma_2\cdots\sigma_M d\sigma_1 d\sigma_2 \cdots d\sigma_M. \tag{16}$$

Let $v_i = \sigma_i^2$, we then obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \int_0^\infty \cdots \int_0^\infty e^{-\sum_{i=1}^M (\frac{\rho}{4M}\lambda_i + 1)v_i} \left(v_1 + \cdots + v_M\right)^{M(N-1)} dv_1 dv_2 \cdots dv_M. \tag{17}$$

We note that

$$\begin{aligned} (v_1 + \cdots + v_M)^{M(N-1)} &= \left(\sum_{i=1}^M v_i\right)^{MN-M} \\ &= \sum_{i_1=1}^M \cdots \sum_{i_{MN-M}=1}^M v_{i_1} \cdots v_{i_{MN-M}} \end{aligned} \tag{18}$$

where the indices i_k in v_{i_k}, $k \in \{1, \cdots, MN-M\}$, take values from the set $\mathcal{J} = \{1, \cdots, M\}$. Assume the subscript index j appears l_j times among the subscripts of the term $v_{i_1} \cdots v_{i_{MN-M}}$ in (18). Then,

$$\begin{aligned} v_{i_1} \cdots v_{i_{MN-M}} &= \prod_{k=1}^{MN-M} v_{i_k} \\ &= \prod_{j=1}^M (v_j)^{l_j} \end{aligned} \tag{19}$$

such that $\sum_{j=1}^M l_j = MN - M$. Using (18) and (19) in (17), and changing the order of integration and summation, we obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \sum_{i_1=1}^M \cdots \sum_{i_{MN-M}=1}^M \left(\int_0^\infty \cdots \int_0^\infty e^{-\sum_{i=1}^M (\frac{\rho}{4M}\lambda_i + 1)v_i} \prod_{i=1}^M (v_i)^{l_i} dv_1 dv_2 \cdots dv_M \right). \tag{20}$$

Using

$$\int_0^\infty x^m e^{-ax} dx = \frac{m!}{a^{m+1}}$$

(20) simplifies to

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{(1+\frac{\rho\lambda_1}{4M})^{l_1+1} \cdots (1+\frac{\rho\lambda_M}{4M})^{l_M+1}}. \tag{21}$$

For high SNR, it follows that

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1+1} \cdots \lambda_M^{l_M+1} (\rho/4M)^{\sum_{i=1}^{M}(l_i+1)}}. \tag{22}$$

Since $\sum_{l_i=1}^{M}(l_i+1) = MN$, we finally arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \left(\frac{N}{(M!)^{N-1}} \frac{1}{\prod_{i=1}^{M} \lambda_i} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1} \cdots \lambda_M^{l_M}} \right) \left(\frac{\rho}{4M} \right)^{-MN}. \tag{23}$$

Inequality (23) clearly shows that a diversity advantage of MN can be achieved when only one antenna is selected based on the instantaneous SNR at the receiver. This diversity gain is equal to the diversity order of the system that uses all the antenna elements in the decoding. However, we note again that this is the case only if the space-time code has full spatial diversity. Although the diversity order is preserved, there will be a loss in the amount of coding gain with antenna selection. We will consider this loss later in more detail when we specifically study the case with two transmit antennas.

2.3.2.2 Rank-deficient Space-time Codes with Antenna Selection

In the analysis of the previous section, we assumed that the eigenvalues λ_i, $i = 1, 2, \cdots, M$, of the matrix $\mathbf{B}\mathbf{B}^H$ were all non-zero. In this case, $\text{rank}(\mathbf{B}) = \text{rank}(\mathbf{B}B^H) = M$ and the maximum diversity advantage, MN, is achieved. When the codeword difference matrix is rank-deficient, *i.e.*, $\text{rank}(\mathbf{B}) = r < M$, the diversity gain obtained for the system using all the antenna elements is Nr. On the other hand, with antenna selection based on the largest SNR observed, the diversity gain degrades dramatically as we will demonstrate shortly.

Assume that there are r non-zero eigenvalues, $\lambda_1, \cdots, \lambda_r$. The analysis for rank-deficient space-time codes follows the same lines as (15) – (21); thus, we will not repeat it. However, it differs following (21) since some of the eigenvalues vanish in this expression. When the

SNR is high, with the assumption that $\lambda_i = 0$ for $i \in \{r+1, \cdots, M\}$, we can write (from (21))

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \frac{1}{\prod_{i=1}^{r} \lambda_i} \left[\sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1} \cdots \lambda_r^{l_r}} \left(\frac{\rho}{4M}\right)^{-\sum_{i=1}^{r} l_i} \right] \left(\frac{\rho}{4M}\right)^{-r}. \tag{24}$$

Recall that $\sum_{i=1}^{M} l_i = MN - M$, and therefore, $0 \leq \sum_{i=1}^{r} l_i \leq MN - M$. In (24), the term in the square brackets is a function of $\rho/4M$. Note that there certainly exist $i_1, \cdots, i_{MN-M}$ such that $\sum_{i=1}^{r} l_i = 0$. We can re-group the terms in (24) to arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \frac{1}{\prod_{i=1}^{r} \lambda_i} \left[\sum_{j=0}^{MN-M} \xi_j \left(\frac{\rho}{4M}\right)^{-j} \right] \left(\frac{\rho}{4M}\right)^{-r} \tag{25}$$

where $j = \sum_{i=1}^{r} l_i$, and ξ_j, is the sum of the terms multiplying $\left(\frac{\rho}{4M}\right)^{-\sum_{i=1}^{r} l_i}$ with the same exponents. For sufficiently high SNR, the terms with $\left(\frac{\rho}{4M}\right)^{-j}$ goes to 0 whenever $\sum_{i=1}^{r} l_i > 0$. Thus, we get

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{(M!)^{N-1}} \frac{1}{\prod_{i=1}^{r} \lambda_i} \xi_0 \left(\frac{\rho}{4M}\right)^{-r}. \tag{26}$$

This expression suggests a diversity order of r as opposed to MN. However since this is only an upper bound on the PEP, we need further analysis. Recall that we have used the Chernoff bound to obtain these upper bounds. We now approximate the PEP with the help of a lower bound on the bound in (14).

First, we note the following simple result.

Lemma 2: If $g(v)$ is as defined in *Lemma 1*,

$$g(v) \geq \frac{e^{-v} v^M}{M!}$$

Proof:

$$\begin{aligned} g(v) &= 1 - e^{-v} \sum_{m=0}^{M-1} \frac{v^m}{m!} \\ &= e^{-v} \left(e^v - \sum_{m=0}^{M-1} \frac{v^m}{m!} \right) = e^{-v} \left(\sum_{m=M}^{\infty} \frac{v^m}{m!} \right) \\ &\geq \frac{e^{-v} v^M}{M!}. \end{aligned} \tag{27}$$

■

Using this bound on $g(v)$ in (15) with $v = \sigma_1^2 + \cdots + \sigma_M^2$, we obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \approx \frac{2^M N}{(M!)^{N-1}} \int_0^\infty \cdots \int_0^\infty e^{-\frac{\rho}{4M}\sum_{i=1}^r \lambda_i \sigma_i^2} e^{-N(\sigma_1^2+\cdots+\sigma_M^2)} \times \left(\sigma_1^2 + \cdots + \sigma_M^2\right)^{M(N-1)} \sigma_1\sigma_2\cdots\sigma_M d\sigma_1 d\sigma_2 \cdots d\sigma_M. \tag{28}$$

Evaluation of the above integration results in

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \approx \frac{N}{(M!)^{N-1}} \sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{(N + \frac{\rho\lambda_1}{4M})^{l_1+1} \cdots (N + \frac{\rho\lambda_M}{4M})^{l_M+1}}. \tag{29}$$

Assuming that $\lambda_i = 0$ for $i \in \{r+1, \cdots, M\}$, for high SNR, we can write,

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \approx \frac{N}{(M!)^{N-1} N^{MN}} \frac{1}{\prod_{i=1}^r \lambda_i} \left[\sum_{i_1=1}^{M} \cdots \sum_{i_{MN-M}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1} \cdots \lambda_r^{l_r}} \left(\frac{\rho}{4MN}\right)^{-\sum_{i=1}^r l_i} \right] \left(\frac{\rho}{4MN}\right)^{-r}. \tag{30}$$

With similar arguments used to obtain (26), we arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \approx \frac{N}{(M!)^{N-1} N^{MN-r}} \frac{1}{\prod_{i=1}^r \lambda_i} \xi_0' \left(\frac{\rho}{4M}\right)^{-r}. \tag{31}$$

From (26) and (31), we observe that the Chernoff bound on the PEP is squeezed between two curves that have the same order in the exponent of ρ, which is $r = \text{rank}(\mathbf{B})$. Since the Chernoff bound is tight in the exponential sense, *i.e.*, $PEP = K * \text{Chernoff Bound} + O(1)$, where K is a constant that does not depend on the SNR, the slopes of the exact PEP plot and the Chernoff bound plot will have the same slopes on a log-log scale. Hence, we conclude that the diversity gain of the system with antenna selection is only r, contrary to the case of the full-complexity system where the diversity gain is Nr. Therefore, to exploit the diversity gain promised by MIMO systems when antenna selection is employed, space-time codes with full spatial diversity should be employed.

2.3.3 Upper Bound for any M and N when $L > 1$ Antennas are Selected

In this section, we will extend the performance analysis presented in the previous sections to the more general case of $L > 1$. For the full spatial diversity system, since selecting a single receive antenna results in full diversity, we expect that the diversity obtained by

selecting L out of N antennas will be the same. However, the coding gain may be different. Also, for a rank-deficient system, it will be interesting to observe the effect of the number of antennas selected at the receiver on the overall diversity order achieved.

Let us denote the rows of $\mathbf{H}$ with the largest L norms by $\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \cdots, \hat{\mathbf{r}}_L$. Similar to the case of single antenna selection, let us introduce an event

$$\mathcal{F}' = \{\mathbf{r}_{i_1}, \cdots, \mathbf{r}_{i_L} \text{ have the largest norms among all the rows}\}.$$

We also define another auxiliary event

$$\mathcal{A}_l = \{i_l^{th} \text{ row has the minimum norm among } \mathbf{r}_{i_1}, \cdots, \mathbf{r}_{i_L}\}.$$

Then, we can obtain the joint pdf for the rows having the largest norms as follows:

$$\begin{aligned}
f_{\hat{\mathbf{R}}_1,\cdots,\hat{\mathbf{R}}_L}(\mathbf{r}_1,\cdots,\mathbf{r}_L) &= f_{\mathbf{R}_{i_1},\cdots,\mathbf{R}_{i_L}}(\mathbf{r}_1,\cdots,\mathbf{r}_L|\mathcal{F}') \\
&\overset{(1)}{=} \sum_{l=1}^{L} f_{\mathbf{R}_{i_1},\cdots,\mathbf{R}_{i_L}}(\mathbf{r}_1,\cdots,\mathbf{r}_L|\mathcal{F}',\mathcal{A}_l)P(\mathcal{A}_l) \\
&\overset{(2)}{=} \sum_{l=1}^{L} \frac{P(\mathcal{F}',\mathcal{A}_l|\mathbf{R}_{i_1}=\mathbf{r}_1,\cdots,\mathbf{R}_{i_L}=\mathbf{r}_L) f_{\mathbf{R}_{i_1},\cdots,\mathbf{R}_{i_L}}(\mathbf{r}_1,\cdots,\mathbf{r}_L)P(\mathcal{A}_l)}{P(\mathcal{F}',\mathcal{A}_l)} \\
&\overset{(3)}{=} \frac{C(N,L)}{L}\sum_{l=1}^{L} P(\|\mathbf{r}_{i_{L+1}}\|^2 < \|\mathbf{r}_l\|^2,\cdots,\|\mathbf{r}_{i_N}\|^2 < \|\mathbf{r}_l\|^2)\prod_{j=1}^{L} f_{\mathbf{R}_{i_j}}(\mathbf{r}_j) I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L) \\
&= \frac{N!}{(N-L)!L!L}\left(\sum_{l=1}^{L}\left[1-e^{-\|\mathbf{r}_l\|^2}\sum_{m=0}^{M-1}\frac{\|\mathbf{r}_l\|^{2m}}{m!}\right]^{N-L} I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L)\right)\frac{e^{-(\|\mathbf{r}_1\|^2+\cdots+\|\mathbf{r}_L\|^2)}}{\pi^{ML}}
\end{aligned} \tag{32}$$

where (1) follows because of the total probability theorem, (2) follows because of Bayes' rule, and (3) follows because of the facts that $P(\mathcal{F},\mathcal{A}_l) = 1/C(N,L)$ and $P(\mathcal{A}_l) = 1/L$. $I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L)$ is the indicator function

$$I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L) = \begin{cases} 1 \text{ if } (\mathbf{r}_1,\cdots,\mathbf{r}_L) \in \mathcal{R}_l \\ 0 \qquad\qquad \text{else} \end{cases}$$

where the region $\mathcal{R}_l$ is defined as

$$\mathcal{R}_l = \{\mathbf{r}_1,\cdots,\mathbf{r}_L : \|\mathbf{r}_l\| < \|\mathbf{r}_k\|, \;\; k = 1,\cdots,l-1,l+1,\cdots,L\}$$

The pairwise error probability can thus be obtained by averaging the conditional pairwise

error probability over this pdf

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \sum_{l=1}^{L}\int_{\mathcal{R}_l} e^{-\frac{\rho}{4M}\|\tilde{\mathbf{H}}\mathbf{B}\|^2}\frac{N!}{(N-L)!L!L}\left[1-e^{-\|\mathbf{r}_l\|^2}\sum_{m=0}^{M-1}\frac{\|\mathbf{r}_l\|^{2m}}{m!}\right]^{N-L}\times \frac{1}{\pi^{ML}}e^{-(\|\mathbf{r}_1\|^2+\cdots+\|\mathbf{r}_L\|^2)}d\mathbf{r}_1\cdots d\mathbf{r}_L \tag{33}$$

However, the exact evaluation of (33) over this region is quite difficult. Instead, for its analytic tractability, we will evaluate the integral throughout the whole space which results in a looser upper bound. We also note that because of the symmetry of the pdf, the integral over $\mathcal{R}_l$ for each l will have the same value. We now consider the evaluation of the integral in the l^{th} term. Before we proceed further, we note that

$$\begin{aligned}\|\tilde{\mathbf{H}}\mathbf{B}\|^2 &= \mathrm{tr}\{\tilde{\mathbf{H}}\mathbf{B}\mathbf{B}^H\tilde{\mathbf{H}}^H\}\\ &= \mathrm{tr}\{\tilde{\mathbf{H}}\mathbf{U}\mathbf{\Lambda}(\tilde{\mathbf{H}}\mathbf{U})^H\}\\ &= \sum_{i=1}^{M}\lambda_i\|\mathbf{c}_i\|^2\end{aligned} \tag{34}$$

where we used the eigenvalue decomposition of $\mathbf{B}\mathbf{B}^H$. Here, $\mathbf{c}_i$ is the i^{th} column of $\tilde{\mathbf{H}}\mathbf{U}$. Using standard integration techniques, the l^{th} term of (33) can be re-written as

$$\mathcal{I}_l = \frac{N!}{(N-L)!L!L}\int_0^\infty\cdots\int_0^\infty e^{-\frac{\rho}{4M}\left(\lambda_1(u_{11}+\cdots+u_{L1})+\cdots+\lambda_M(u_{1M}+\cdots+u_{LM})\right)} \times\left(1-e^{-(u_{l1}+\cdots+u_{lM})}\sum_{k=0}^{M-1}\frac{(u_{l1}+\cdots+u_{lM})^k}{k!}\right)^{N-L} e^{-(u_{11}+\cdots+u_{lM})}du_{11}\cdots du_{LM}. \tag{35}$$

After some manipulations, we obtain

$$\begin{aligned}\mathcal{I}_l = \frac{N!}{(N-L)!L!L}&\int_0^\infty\cdots\int_0^\infty e^{-\frac{\rho}{4M}\sum_{i=1}^M\lambda_i\left(\sum_{d=1,d\neq l}^L u_{di}\right)}e^{-\left(\sum_{i=1}^M\sum_{d=1,d\neq l}^L u_{di}\right)}\prod_{i=1}^{M}\prod_{d=1,d\neq l}^{L}du_{di}\\ &\times\int_0^\infty\cdots\int_0^\infty e^{-\frac{\rho}{4M}\sum_{i=1}^M\lambda_i u_{li}}\left(1-e^{-(u_{l1}+\cdots+u_{lM})}\times\right.\\ &\left.\sum_{k=0}^{M-1}\frac{(u_{l1}+\cdots+u_{lM})^k}{k!}\right)^{N-L} e^{-(u_{l1}+\cdots+u_{lM})}du_{l1}\cdots du_{lM}\end{aligned} \tag{36}$$

Denote the first integral in this expression by $\mathcal{I}_l^{(1)}$ and the second integral by $\mathcal{I}_l^{(2)}$. The

result of the first integral, $\mathcal{I}_l^{(1)}$, is

$$\mathcal{I}_l^{(1)} = \left(\frac{1}{\prod_{i=1}^{M} 1 + \frac{\rho \lambda_i}{4M}} \right)^{L-1} . \tag{37}$$

The second integral is very similar to the one that we obtained for the case of single antenna selection and can be evaluated as described in Section 2.3.2. Using *Lemma 1* in (36), we can upper bound $\mathcal{I}_l^{(2)}$, as

$$\mathcal{I}_l^{(2)} \leq \frac{1}{(M!)^{N-L}} \left(\sum_{i_1=1}^{M} \cdots \sum_{i_{MN-ML}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1+1} \cdots \lambda_M^{l_M+1}} \right) \left(\frac{\rho}{4M} \right)^{-M(N-L+1)} \tag{38}$$

Using these results, we obtain

$$\mathcal{I}_l \leq \frac{N!}{(N-L)!L!L(M!)^{N-L}} \left(\frac{1}{\prod_{i=1}^{M} 1 + \frac{\rho \lambda_i}{4M}} \right)^{L-1} \left(\sum_{i_1=1}^{M} \cdots \sum_{i_{MN'-M}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1+1} \cdots \lambda_M^{l_M+1}} \right) \left(\frac{\rho}{4M} \right)^{-M(N-L+1)} . \tag{39}$$

Note that this resulting bound is independent of l. Hence, substituting (39) into (33), and performing simple algebraic manipulations for the high SNR region, we finally arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N!}{(N-L)!L!(M!)^{N-L}} \frac{1}{\left(\prod_{i=1}^{M} \lambda_i \right)^L} \left(\sum_{i_1=1}^{M} \cdots \sum_{i_{MN-ML}=1}^{M} \frac{l_1! \cdots l_M!}{\lambda_1^{l_1} \cdots \lambda_M^{l_M}} \right) \left(\frac{\rho}{4M} \right)^{-MN} . \tag{40}$$

The inequality in (40) clearly indicates that the diversity order is MN when a selection of L antennas out of N antennas is made available at the receiver and the space-time code achieve full spatial diversity. Note also that this expression reduces to the expression in (23) when $L = 1$, and takes the same value as the full-complexity system for $L = N$. Clearly, (23) is also an upper bound on the PEP for the case under consideration. However, the bound in (40) is tighter than (23) and provides a better assessment of the coding gain of the system.

For rank-deficient space-time codes, the analysis is very similar to the case when $L = 1$, which is described in Section 2.3.2.2. Hence, we will not repeat these steps. If we have $r < M$ non-zero eigenvalues, one can obtain the upper bound on the PEP as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N!}{(N-L)!L!(M!)^{N-L}} \frac{1}{(\prod_{i=1}^{r} \lambda_i)^L} \xi_0'' \left(\frac{\rho}{4M} \right)^{-Lr} . \tag{41}$$

Similarly, the lower bound on the Chernoff bound (an approximation to the PEP) is given by (using (31)),

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \approx \frac{N!}{(N-L)!L!(M!)^{N-L}\left(\prod_{i=1}^{r}\lambda_i\right)^{L}(N-L+1)^{MN-r}}\xi_0'''\left(\frac{\rho}{4M}\right)^{-Lr}. \qquad (42)$$

In (41) and (42), ξ_0'' and ξ_0''' are the coefficients that can be computed in a similar fashion to that used to compute ξ_0 and ξ_0' in (26) and (31), respectively. The upper bound in (41) and the approximation in (42) obtained for the rank-deficient space-time codes indicate that the diversity order is Lr. To verify this approximation, we have conducted extensive simulations, and we have observed that the diversity gain is exactly Lr for this case as the expression suggests. These results indicate that, with antenna selection, the diversity order that can be obtained when the underlying space-time code is rank-deficient is only Lr while the diversity order of the full-complexity system is Nr.

Such a degradation in the achievable diversity order may be somewhat counterintuitive since one might have expected the results to be Nr as opposed to Lr as an extension of the full-rank space-time case. To explain this further, let us consider a space-time code with rank $r < M$. Such a code will achieve full spatial diversity for a system with r transmit antennas. Assume this code is used for a system having M transmit antennas, which can be trivially done by transmitting the same (dummy) symbols from the additional $M - r$ antennas. Hence, these antennas will not give the receiver any useful information that will improve the diversity gain of the system. However, when we perform antenna selection, the channel gains from these antennas may enforce the selection of a "bad" subset of receive antennas, which will degrade the performance of the system severely. Even if this is the case for a fraction of time, asymptotic behavior of the PEP will depend on these falsified selections resulting in the reduced diversity order.

Clearly the performance with antenna selection will improve if we do not use the channel coefficients from the last $M-r$ antennas for selection, in which case a diversity order of Nr will be achieved. However this means that the spatial structure of the underlying space-time code has to be used in the selection process. For this example, this is simple. However, in general it is not an easy task even for simple non full-rank space-time codes.

2.4 Tighter Upper Bounds for Systems with Two Transmit Antennas

Although the analysis in the previous section accurately predicts the diversity order, the upper bounds derived are not very tight. In this section, we will evaluate the exact value of the bound given in (14) to obtain a tighter bound for the case of double transmit antenna diversity. For $M = 2$, (14) reduces to

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N \int_{\mathcal{C}^2} e^{-\frac{\rho}{8}(\lambda_1|\hat{z}_1|^2+\lambda_2|\hat{z}_2|^2)} \left(1 - e^{-(|\hat{z}_1|^2+|\hat{z}_2|^2)}(1 + (|\hat{z}_1|^2 + |\hat{z}_1|^2))\right)^{(N-1)} \times \frac{1}{\pi^2} e^{-(|\hat{z}_1|^2+|\hat{z}_2|^2)} d\hat{z}_1 d\hat{z}_2 \tag{43}$$

Replacing the complex valued integration variables $\hat{z}_1$ and $\hat{z}_2$ with $\hat{z}_1 = \sigma_1 e^{j\theta_1}$ and $\hat{z}_2 = \sigma_2 e^{j\theta_2}$, we arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 4N \int_0^\infty \int_0^\infty e^{-\frac{\rho}{8}(\lambda_1\sigma_1^2+\lambda_2\sigma_2^2)} \left(1 - e^{-(\sigma_1^2+\sigma_2^2)}(1 + (\sigma_1^2 + \sigma_2^2))\right)^{(N-1)} \times e^{-(\sigma_1^2+\sigma_2^2)} \sigma_1\sigma_2 d\sigma_1 d\sigma_2 \tag{44}$$

since $\int\int d\theta_2 d\theta_2 = 4\pi^2$ and the integrand does not depend on θ_1 or θ_2. By a change of variables in (44) with $\sigma_1 = y\cos\theta$ and $\sigma_2 = y\sin\theta$, we obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 4N \int_0^{\pi/2} \int_0^\infty e^{-\frac{\rho y^2}{8}(\lambda_1\cos^2\theta+\lambda_2\sin^2\theta)} \left(1 - e^{-y^2} - e^{-y^2}y^2\right)^{(N-1)} e^{-y^2} y^2 \cos\theta\sin\theta y dy d\theta \tag{45}$$

Assuming $\lambda_1 \neq \lambda_2$, integrating with respect to θ results in

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 4N \int_0^\infty \left(1 - e^{-y^2} - e^{-y^2}y^2\right)^{(N-1)} e^{-y^2} y \frac{e^{-\rho/8y^2\lambda_1} - e^{-\rho/8y^2\lambda_2}}{\rho/4(\lambda_2 - \lambda_1)} dy. \tag{46}$$

We can further simplify (46) by replacing $e^{-y^2} = x$ to obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{8N}{(\lambda_2 - \lambda_1)\rho} \int_0^1 (1 + x\log(x/e))^{(N-1)} \left(x^{\rho\lambda_1/8} - x^{\rho\lambda_2/8}\right) dx. \tag{47}$$

When the eigenvalues are equal, *i.e.*, $\lambda_1 = \lambda_2 = \lambda$, the upper bound is given by

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N \int_0^1 (1 + x\log(x/e))^{(N-1)} \left(x^{\rho\lambda/8}\right) \log(x) dx \tag{48}$$

We can simplify (47) or (48) by using a binomial expansion and integrating each term separately to obtain the following closed-form expression (see Appendix A)

$$P(S \to \hat{S}) \leq N \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n \left(\frac{1}{a_n b_n} + \frac{n\,(a_n + b_n)}{a_n^2 b_n^2} + \frac{n(n-1)\left(a_n^2 + a_n b_n + b_n^2\right)}{a_n^3 b_n^3} + \cdots + \frac{n!\,(a_n^n + \cdots + b_n^n)}{a_n^{n+1} b_n^{n+1}} \right) \tag{49}$$

where a_n and b_n are defined as in the Appendix. Although (49) is a closed-form expression, it is not very simple. We now present the upper bounds for several specific cases in the following examples.

Example 1: $M = 2$ **and** $N = 2, 3, 4$

It is easy to show that, for $M = 2$ and $N = 2$,

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 2 \left[\frac{a^2 + b^2 + (a-1)(b-1) - 1}{a^2 b^2 (a-1)(b-1)} \right], \tag{50}$$

where $a = 2 + \rho\lambda_1/8$ and $b = 2 + \rho\lambda_2/8$. When the SNR is high, assuming $\lambda_1, \lambda_2 > 0$, we can write

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 2 \left[\frac{\lambda_1^2 + \lambda_2^2 + \lambda_1\lambda_2}{\lambda_1^3 \lambda_2^3} \right] (\rho/8)^{-4}. \tag{51}$$

It is worth comparing this result with the upper bound for the full-complexity system in which all the antennas are used in the decoding process. In [163], the PEP for this system is given

$$P_{e,no\_sel} \leq \frac{1}{\lambda_1^2 \lambda_2^2} (\rho/8)^{-4}. \tag{52}$$

Comparing (52) and (51), we observe that the diversity advantage of both systems are the same. That is, the antenna selection based on the SNR observed does not degrade the diversity advantage of the system. On the other hand, there is some loss in the coding advantage if we perform antenna selection. Comparing the upper bounds, we see that if $M = N = 2$, and $L = 1$, the loss is approximately given by (in dB)

$$C_{sel}/C_{no\_sel} = 10 \log_{10} 2 \left[1 + \frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} \right]. \tag{53}$$

It is clear that (53) reaches its minimum value when the eigenvalues, λ_i, $i = 1, 2$, associated with the codeword difference matrices, are equal to each other. This result can be used as

an additional criterion to optimize the design of space-time codes for systems using antenna selection.

When we have $N = 3$, we obtain the upper bound as:

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 6\frac{(3a+1)b^5+(3a^2+7a+2)b^4+(3a^3+7a^2+2a)b^3+(3a^4+7a^3+2a^2-6a-2)b^2+(3a^5+7a^4+2a^3-6a^2-5a-1)b+a^5+2a^4-2a^2-a}{(a-1)(b-1)a^2b^2(a+1)^3(b+1)^3}, \tag{54}$$

where $a = 2 + \rho\lambda_1/8$ and $b = 2 + \rho\lambda_2/8$. When the SNR is high, the bound is given by

$$P(S \to \hat{S}) \leq 18\frac{\lambda_1^4 + \lambda_1^3\lambda_2 + \cdots + \lambda_2^4}{\lambda_1^5\lambda_2^5}(\rho/8)^{-6}.$$

Similarly for N=4, the upper bound for high SNRs (keeping only the highest order terms) is given as

$$P(S \to \hat{S}) \leq 360\frac{\lambda_1^6 + \lambda_1^5\lambda_2 + \cdots + \lambda_2^6}{\lambda_1^7\lambda_2^7}(\rho/8)^{-8}. \qquad \blacksquare$$

Example 2: $M = 2$**,** $N = 3$ **and** $L = 2$ **antennas selected.**

Assume that there are 2 transmit and 3 receive antennas and 2 are selected based on the SNRs observed. The expression (33) can be rewritten as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 3\int_{\|\mathbf{r}_2\|<\|\mathbf{r}_1\|} e^{-\frac{\rho}{8}\mathrm{tr}\{\tilde{\mathbf{H}}\mathbf{U}\mathbf{\Lambda}(\tilde{\mathbf{H}}\mathbf{U})^H\}}\left[1 - e^{-\|\mathbf{r}_2\|^2}(1 + \|\mathbf{r}_2\|^2)\right]\frac{1}{\pi^4}e^{-(\|\mathbf{r}_1\|^2+\|\mathbf{r}_2\|^2)}d\mathbf{r}_1 d\mathbf{r}_2 \tag{55}$$

where $\mathbf{B}\mathbf{B}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$. Let $\tilde{\mathbf{H}}\mathbf{U} = \begin{bmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{bmatrix} = \begin{bmatrix}\beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22}\end{bmatrix}$. Then $\|\mathbf{r}_i\| = \|\boldsymbol{\beta}_i\|$, $i = 1, 2$. Transforming the integral into polar coordinates with $\beta_{ij} = \sigma_{ij}e^{\theta_{ij}}$, we get

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 3\iiiint e^{-\frac{\rho}{8}(\lambda_1(\sigma_{11}^2+\sigma_{21}^2)+\lambda_2(\sigma_{12}^2+\sigma_{22}^2))}\left[1 - e^{-(\sigma_{21}^2+\sigma_{22}^2)}(1 + \sigma_{21}^2 + \sigma_{22}^2)\right] \times \\ \frac{1}{\pi^4}e^{-(\sigma_{11}^2+\sigma_{12}^2+\sigma_{21}^2+\sigma_{22}^2)}\sigma_{11}\sigma_{12}\sigma_{21}\sigma_{22}d\sigma_{11}d\sigma_{12}d\sigma_{21}d\sigma_{22}\left(\int_0^{2\pi} d\theta\right)^4. \end{aligned} \tag{56}$$

Unfortunately, exact evaluation of the integral over the region

$$\mathcal{R} = \left\{\sigma_{11} > 0, \sigma_{12} > 0, \sigma_{21} > 0, \sigma_{22} > 0 : \sigma_{11}^2 + \sigma_{12}^2 > \sigma_{21}^2 + \sigma_{22}^2\right\}$$

as indicated previously, is quite difficult. However, we can evaluate the integral over the whole space. After some manipulations, we obtain (assuming $\lambda_2 > \lambda_1$)

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{24}{\rho(\lambda_2 - \lambda_1)(1 + \rho\lambda_1/8)(1 + \rho\lambda_2/8)}\int_0^1 (1 + x\log x/e)\left(x^{\rho\lambda_1/8} - x^{\rho\lambda_2/8}\right)dx. \tag{57}$$

Note that the integral in this expression is very similar to the integral obtained for the case $M = N = 2$ and $L = 1$. Using the previous results, we arrive at the following upper bound for high SNRs:

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq 3 \left[\frac{\lambda_1^2 + \lambda_2^2 + \lambda_1 \lambda_2}{\lambda_1^4 \lambda_2^4} \right] (\rho/8)^{-6} . \qquad \blacksquare \qquad (58)$$

2.5 Space-time Code Design With Antenna Selection

For the full-complexity system, two design criteria based on the upper bound on PEP have been proposed in [163]: to maximize the diversity gain, the minimum of the ranks of the codeword difference matrices, $\mathbf{B}$, and to maximize the coding gain, the minimum of the determinants of the matrices $\mathbf{B}\mathbf{B}^H$ should be maximized. The bounds developed in the previous sections can be used to develop similar criteria for the design of space-time codes for use with antenna selection at the receiver. We propose the following:

- To achieve the maximum diversity gain MN, the underlying space-time code should be full-rank, *i.e.*, $\text{rank}(\mathbf{B}) = M$.
- To achieve the maximum coding gain, the value of the coefficient in (23) must be minimized for all codeword pairs.

The rank criterion imposes a significant constraint on the design of space-time codes when they will be used on MIMO systems employing antenna selection. The trade-off between the diversity advantage and the transmission rate becomes more essential for such an application. Hence, the recently proposed full-rank and full-rate space-time code design techniques (*i.e.,* [17–19] [51, 52, 108]) may be viable alternatives for systems with antenna selection.

Clearly, instead of (23), we can also use the tighter bounds developed for specific M, N and L. For example, for the case of double transmit diversity, (51) can be used. In this case, we should generate the code that maximizes the coding gain that is defined as the minimum of

$$Gain = \sqrt{\frac{\lambda_1^3 \lambda_2^3}{2\left(\lambda_1^2 + \lambda_2^2 + \lambda_1 \lambda_2\right)}}$$

Table 2: Comparison of known and new 4-PSK space-time codes with theoretical coding gains. GM: Generator Matrix, FC: full-complexity system, S: system with antenna selection.

States	GM [163]	Gains (FC/S)	GM [12]	Gains (FC/S)	GM New	Gains (FC/S)	Improvement over [163]
	0 2		2 2		1 0		
	0 1	2	0 2	$2\sqrt{2}$	2 1	$2\sqrt{2}$	1.51 dB
4	2 0	1.63	1 0	1.66	2 2	1.67	0.09 dB
	1 0		3 1		0 2		
	0 2		2 2		2 1		
	0 1	$\sqrt{12}$	0 1	$\sqrt{12}$	2 0	$\sqrt{12}$	0 dB
8	2 0	2.56	2 0	2.56	0 2	3.13	0.88 dB
	1 0		1 0		3 2		
	2 2		2 2		3 2		
	0 2		1 2		0 2		
	0 1	$\sqrt{12}$	2 0	$\sqrt{20}$	2 2	$\sqrt{20}$	1.11 dB
16	2 0	4.07	2 1	3.25	3 0	4.12	0.04 dB
	1 2		0 2		0 3		
	0 2		0 2		1 2		
	2 0		2 0		1 0		

over all codeword pairs for the case of $M = N = 2$.

Let us give several examples using a systematic code search technique similar to the method presented in [12]. We first transform the trellis representation of the code into generator matrix form, and then perform an exhaustive search systematically using the generator matrix. For brevity, we refer the reader to [12] for the details of the method. Here, we briefly present our search results. Some of the known codes (the generator matrix forms) together with new codes using 4, 8 and 16 state trellises are shown in Table 2.

In Table 2, we included the coding gains for both the full-complexity system and the system with antenna selection. Two notes are in order: first, the codes designed (by the proposed method) for the system with antenna selection performs better than known codes when they are used with antenna selection. This is because already existing codes are designed for the full-complexity system, not for systems with antenna selection. Second, the proposed codes, which are designed for the system with antenna selection, also provide improved (or, at least the same) coding gains for the full-complexity system compared to the known codes. That is, the new design criteria do not conflict with those of full-complexity systems, they only impose additional constraints.

2.6 Examples

In this section, we present several numerical examples and simulation results to clarify the theoretical analysis performed. We consider three cases, *i.e.*, upper bounds on the PEP, space-time code design with antenna selection, and the case of rank-deficient space-time codes separately.

2.6.1 Theoretical Upper Bounds on PEP

We now evaluate the bounds given in Section 2.3.2.1 and 2.3.2.2 for several codeword pairs that are selected from the codes developed in [163]. We also provide the actual frame error rates (FER) for the space-time codes considered.

In Figures 8 and 9, we present the PEP bounds for the system with $M = 2$, $N = 3$, and $L = 1, 2$. We select the codeword pairs from the 2 bits/sec/Hz 8−state space-time trellis codes using 4−PSK modulation (with $M = 2$). This code provides a diversity advantage of 6 [163], *i.e.,* full spatial diversity. The two codewords considered differ in three consecutive symbols. The following observations are in order: (i) the simulated PEPs and the PEPs obtained numerically exactly match for both the full-complexity system and the one using antenna selection, (ii) for $L = 2$, the Chernoff bound obtained by integration of the bound over the whole space (evaluated analytically) rather than over the support of the actual pdf (evaluated numerically) are very close to each other (differing by 0.8 dB for this example), (iii) for both the cases of $L = 1$ and $L = 2$, the upper bound and the approximations to the Chernoff bound are very tight at high SNR, and finally (iv) the performance of the system with optimal selection (that maximizes capacity) is only slightly superior to the selection we considered, *i.e.,* optimal selection performs only slightly better for $L = 2$.

The actual frame error rate comparisons for $M = N = 2$ antennas are provided in Figure 10 (see the solid lines). In this example, we present the simulation results (as opposed to the theoretical results) for the space-time trellis code considered in the previous example. The channel is assumed to be constant for a period of 130 transmissions. We observe that the slopes of the FER plots are the same, implying that both systems achieve the same diversity advantage. There is about a 2 dB loss in coding gain as a result of not fully exploiting the

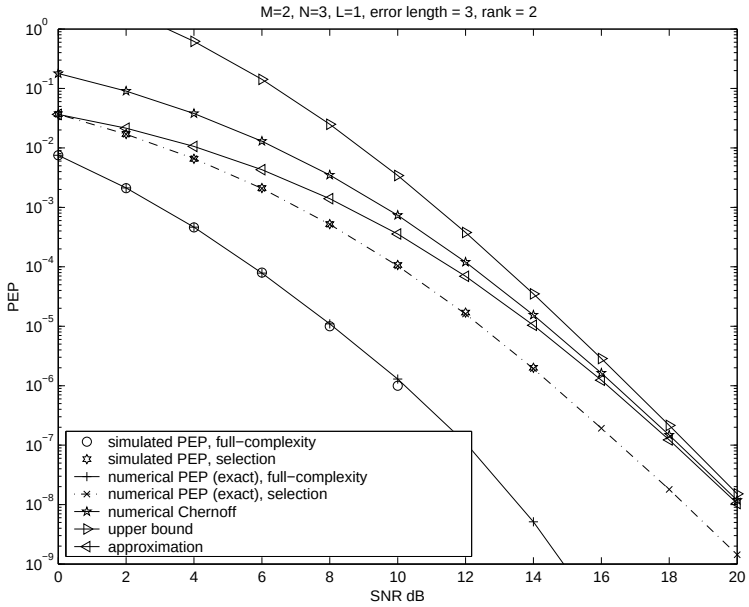


Figure 8: PEP comparison between the full-complexity system and the system using single receive antenna observing maximum average SNR, $M = 2, N = 3, L = 1$. The codeword pairs are from 2 bits/sec/Hz space-time trellis code using 4-PSK, 8-state trellis [3].

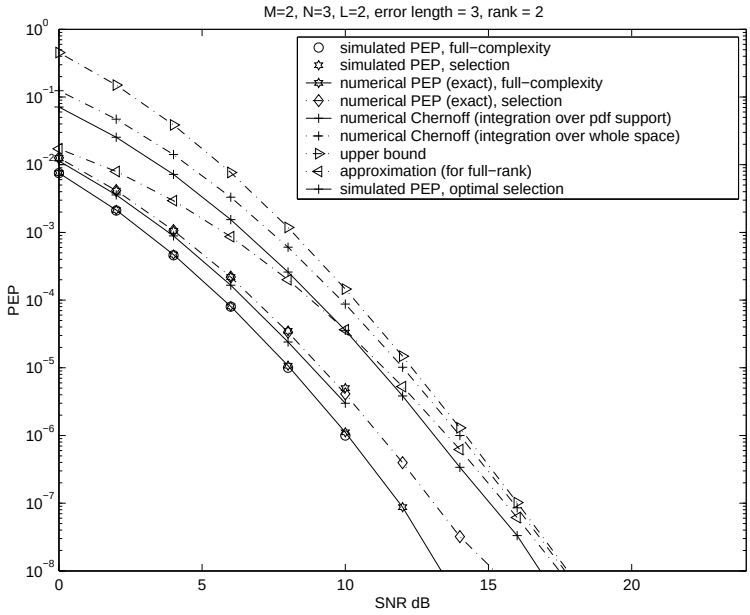


Figure 9: PEP comparison between the full-complexity system and the system using single receive antenna observing maximum average SNR, $M = 2, N = 3, L = 2$. The codeword pairs are from 2 bits/sec/Hz space-time trellis code using 4-PSK, 8-state trellis [3].

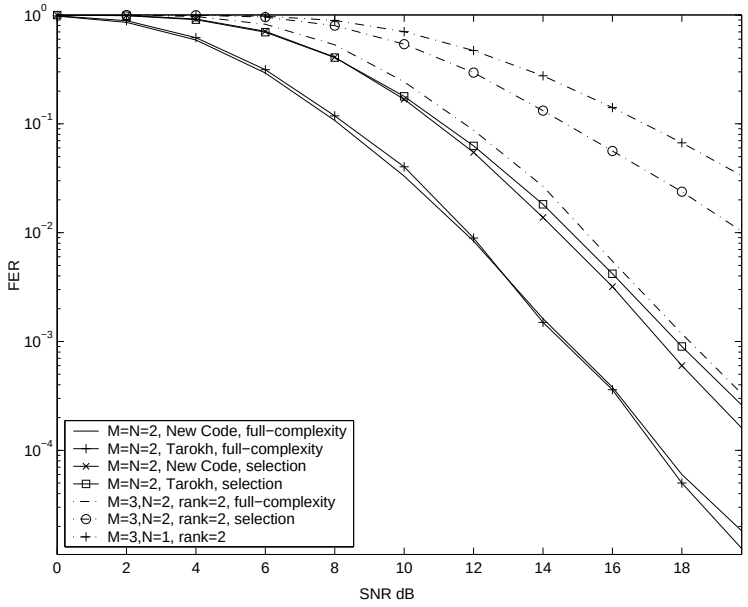


Figure 10: FER comparison between (i) (solid lines) the new code and the known space-time trellis code in [3] when M=2, N=2, the 2 bits/sec/Hz space-time trellis code using 4-PSK, 8-state trellis, and (ii) (dashed lines) the full-complexity system and the system using antenna selection when M=3 and N=2, rank=2.

receive antenna elements. However, the cost is significantly reduced.

2.6.2 Space-Time Codes with Antenna Selection

In Figure 10, we also present a performance comparison between the space-time code designed in [163] and the code we designed for the system with antenna selection. Both space-time codes achieve 2 bits/sec/Hz and they use an 8−state trellis with 4−PSK modulation. The coding gains of both codes for the full-complexity system are the same, *i.e.,* the gains are $\sqrt{12}$, and therefore, they have the same performance. However, for the system with antenna selection, we observe a significant improvement with the new space-time code. The improvement predicted by the theoretical coding gains shown in Table 2 is 0.88 dB which is very close to the 0.7 dB improvement observed by the simulations.

2.6.3 Effect of Rank-deficiency on the Performance

With antenna selection, we showed that the diversity order is preserved provided that the underlying space-time code achieves full spatial diversity. However, the upper bound analysis for rank-deficient space-time codes has revealed that the diversity order is expected

to degrade dramatically. To illustrate this point further, in Figure 11, we depict the bounds on PEP together with the exact values of PEP obtained by simulations when $M = N = 2$ and the space-time code is rank-deficient, *i.e.,* the codeword difference matrix has rank 1. We observe from the slopes of the exact PEP that the diversity gain is only $r = 1$ for antenna selection system while it is $r = 2$ for the full-complexity system. The bounds obtained for the system employing antenna selection also have the same asymptotic slopes as the exact value of the PEP. However, these bounds are not as tight as the ones for the full-rank space-time code presented earlier.

For $M = 3$ and $N = 3$, Figure 12 depicts the PEP bounds for a codeword pair from a rank-deficient space-time code. We present two cases in this figure: (i) selection of a single antenna (ii) selection of 2 antennas. We present exact PEP (obtained through simulations) as well as the bounds we derived for all cases. We observe from these plots that: (i) The diversity order with antenna selection is given by Lr, *i.e.,* the achievable diversity orders are 2 and 4 with $L = 1$ and $L = 2$, respectively, (ii) the optimal selection (that maximizes the capacity) performs slightly better than the selection based on energy for $L = 2$, and finally, (iii) the bounds become tighter as L increases (*i.e.,* as we select more antennas), and become looser as r decreases, (*i.e.,* the rank of the codeword difference matrix reduces).

The frame error rates for a rank-deficient space-time code whose codeword difference matrices of rank-2 are shown in Figure 10 (see dashed lines). In this example, $M = 3$ and the maximum diversity order for the full-complexity system with $N = 2$ receive antennas is $2 \times N = 4$, which can be seen from the solid line in the figure. The dotted-line shows the frame error rate when $L = 1$. We observe that diversity gain in this case is the same as that would be obtained if we had used only one receive antenna, *i.e.,* it is 2. However, there is a 3 dB improvement in the coding gain over the system with only one receive antenna.

2.7 Chapter Summary

We investigated antenna selection for a MIMO wireless system using space-time coding. We considered the case when the fading is iid, only the receiver knows the CSI and antenna

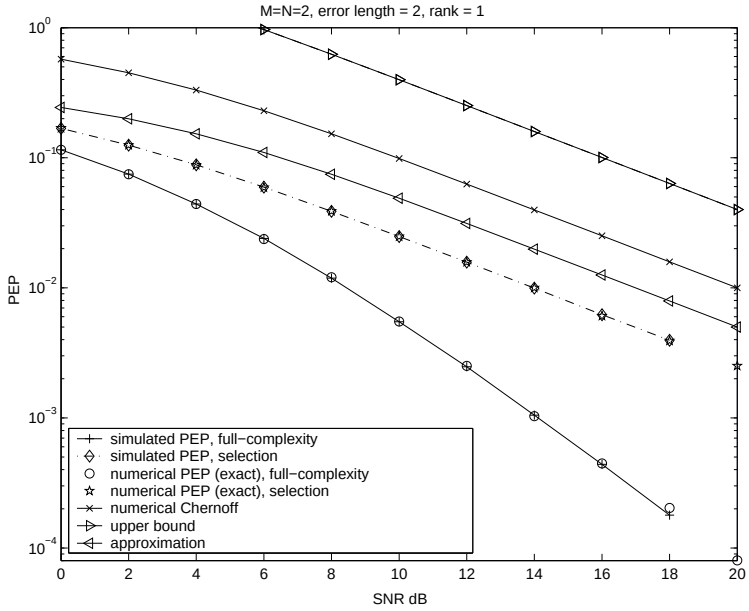


Figure 11: Bounds for rank-deficient space-time code. PEP comparison between the full-complexity system and the system using single receive antenna observing maximum average SNR, $M = 2, N = 2$. For rank-deficient code, $\text{rank}(\mathbf{B}) = 1$.

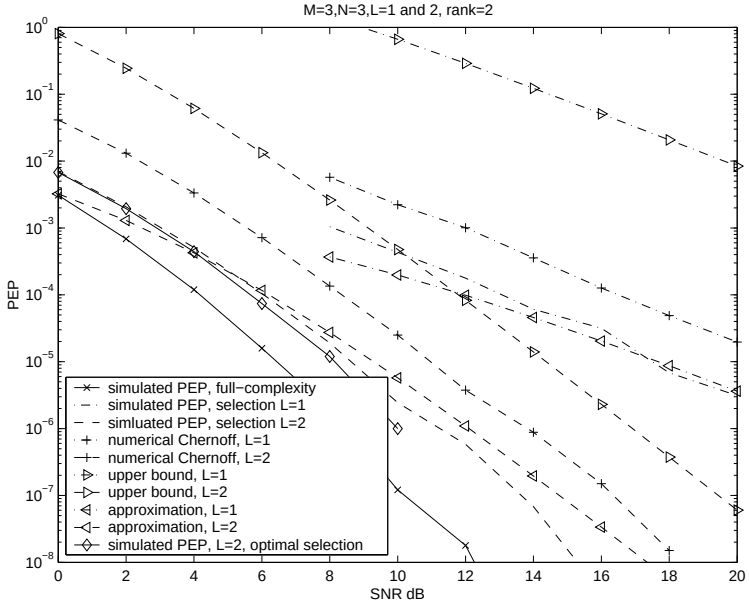


Figure 12: PEP for rank-deficient space-time code: $M = 3$, $N = 3$ and $L = 1$ or 2. For the codeword pairs, $\text{rank}(\mathbf{B}) = 2$.

selection (based on received signal power) is performed only at the receiver. We have analytically shown that the diversity gain does not change and hence, we can exploit the full diversity advantage promised by the MIMO system that uses all available antenna elements, provided that the space-time code employed has full spatial diversity. For rank-deficient systems, we have shown that the diversity gain with antenna selection deteriorates significantly compared to the full-complexity one. Furthermore, for double transmit diversity, we computed tighter upper bounds in closed form. We have also determined guidelines for optimal space-time code design with antenna selection, and presented several simple codes for the case of two transmit antennas. We have also provided extensive numerical examples, and simulation results, and observed that the results are in agreement with the theoretical analysis.

CHAPTER III

PERFORMANCE OF MIMO ANTENNA SELECTION FOR CORRELATED FADING CHANNELS

In Chapter 2, antenna selection for multiple antenna transmission systems has been studied under the assumption that the subchannels between antenna pairs fade independently. In this chapter, we consider the performance of such systems when the subchannels experience correlated fading. We again consider the energy-based antenna selection. We analyze the system performance and quantify the effects of channel correlations on the diversity and coding gain when the receiver selects a subset of the antennas. The organization of the chapter is as follows: In the next section, we briefly summarize related work on receive antenna selection for correlated fading channels. In Section 3.2, we describe the correlation models we consider. In Section 3.3, we derive the PEP the bounds for the system using receive antenna selection. We study the performance in the case of rank-deficient space-time codes in Section 3.4. Section 3.5 summarizes the space-time code design criteria for the correlated fading case. The numerical results validating the analysis are illustrated in Section 3.6. Finally, we summarize our results in Section 3.7.

3.1 Introduction

Most of the work on multiple antenna systems make the assumption that the subchannels between transmit/receive antenna pairs experience independent and identically distributed (iid) fading [3,4,11–13] [49, 73, 95, 151, 163]. A more realistic assumption, however, is that the fades are not independent, because of insufficient spacing between antenna elements, placement of scatterers, etc. [145]. The nonzero correlation between subchannels may significantly reduce the capacity as shown in [30]. The effect of such correlations on the system performance is studied in [22] and [169]. The recent work on correlated fading includes [28, 87, 93, 147]. Hong *et al.* investigate the design and performance of spatial

multiplexing for MIMO correlated fading channels in [87]. In [93], Ivrlac *et al.* study the effects of fading correlations and transmitter channel knowledge on the capacity and cutoff rate for MIMO systems, and in [28], Chiani *et al.* derive closed form expression for the characteristic functions for MIMO system capacity for correlated fading case. Smith *et al.* also study the capacity of MIMO systems, but they focus on semicorrelated flat fading [147].

The effects of subchannel correlation when antenna subset selection is employed have interesting implications. For instance, in [62], Gore *et al.* consider the capacity of MIMO systems with antenna selection when the channel is rank-deficient, and show that a larger capacity can be achieved by using a "good" subset of transmit antennas (*i.e.,* by using those antennas that result in a full-rank channel). In [135], following the work in [62], Sandhu *et al.* propose an efficient method to find the optimal subset of antennas. Another line of work investigates antenna selection based on error probability [61,64]. In [64] and [61], Gore *et al.* assume that the channel statistics change very slowly and that the selected antenna subset remain same over the transmission period. To develop the criteria for selection, the authors derive bounds on average pairwise error probabilities for the full-complexity system over correlated fading that depend on the channel covariance matrix and select the subset of antennas that minimize those bounds. Note that in these studies, error probability for a system using antenna selection is not formulated at all. Only the error probability expressions for the full-complexity system is considered.

In the previous chapter, we studied the performance of the received-power based antenna selection for MIMO systems when the subchannels undergo (independent) quasi-static fading. The case when correlations exist among the subchannels is the topic of this chapter. Our goal is, assuming the presence of transmit or receive correlation, to show whether one can achieve a similar asymptotic performance as the full-complexity system does.

3.2 System Model and Pairwise Error Probability

We consider a system equipped with M transmit and N receive antennas. We denote by $\mathbf{H}$ the $N \times M$ channel transfer matrix

$$\mathbf{H} = \begin{pmatrix} h_{11} & \cdots & h_{1M} \\ \vdots & \ddots & \vdots \\ h_{N1} & \cdots & h_{NM} \end{pmatrix}.$$

The channel is modeled as a flat Rayleigh fading (*i.e.*, $h_{nm} \sim \mathcal{CN}(0,1)$) that remains constant over a block of t symbols and changes independently from one block to the next. The received signal matrix $\mathbf{X}_{N\times t}$ is given by

$$\mathbf{X} = \sqrt{\frac{\rho}{M}}\mathbf{HS} + \mathbf{W} \tag{59}$$

where $\mathbf{S}$ is the $M \times t$ transmitted signal matrix (selected from a space-time codeword alphabet), and $\mathbf{W}$ is the $N \times t$ additive white Gaussian noise matrix. The average energy of the transmitted signal is normalized to unity over M antennas so that ρ is the expected signal-to-noise ratio at each receive antenna. We assume that channel state information (CSI), *i.e.,* $\mathbf{H}$, is known at the receiver, but not at the transmitter.

The assumption of independently and identically distributed (iid) fading is typically made to model the channel [163]. However, in real time propagation, the presence of local scatterers around the transmitter and the receiver induces correlations among subchannels that can be modeled as $\mathbf{H} = \mathbf{R}_{(r)}^{1/2}\mathbf{H}_w\mathbf{R}_{(t)}^{1/2}$ where $\mathbf{R}_{(r)} = \mathbf{R}_{(r)}^{1/2}\mathbf{R}_{(r)}^{H/2}$ is the receive covariance matrix, $\mathbf{R}_{(t)} = \mathbf{R}_{(t)}^{1/2}\mathbf{R}_{(t)}^{H/2}$ is the transmit covariance matrix, and $\mathbf{H}_w$ is a matrix with iid $\mathcal{CN}(0,1)$ entries [145][1]. In general, one can classify four different types of fading correlations: uncorrelated, semicorrelated, semicorrelated type-2, and fully-correlated fading [93]. In this chapter, we consider semicorrelated fading channels for which there exist either receive correlation or transmit correlation. It has been shown that urban radio is well approximated by semicorrelated fading channel models.

Throughout this chapter, we assume the presence of either transmit correlation or receiver correlation. We next describe these correlation models.

[1] $(.)^H$ denote the Hermitian transpose

3.2.1 Transmit Correlation

Let us assume that the receiver is surrounded by many scatterers while the transmitter is placed high enough that there are not many scatterers in its vicinity. Assume also that the antenna spacing at the receiver is sufficiently large so that fading associated with each receive antenna is (almost) independent [145]. In such a set-up, each antenna at the receiver observes correlated fading gains from the transmitter antennas. Explicitly, each row (denoted by $\mathbf{r}_i$) of the channel gain matrix is a circularly symmetric complex Gaussian distributed vector with covariance matrix

$$R_{(t)} = E\{\mathbf{r}_i^H \mathbf{r}_i\}$$

and all the rows are independent and identically distributed, i.e., $\mathbf{R}_{(r)}$ is the $N \times N$ identity matrix. We assume that the complex correlation coefficient between h_{ij} and h_{ik} is r_{jk}. Then

$$R_{(t)} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1M} \\ r_{12}^\star & 1 & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1M}^\star & r_{2M}^\star & \cdots & 1 \end{pmatrix},$$

where $(\cdot)^\star$ denotes the complex-conjugation.

The PEP for a full-rank channel[2], assuming maximum likelihood decoding at the receiver, is given by [61]

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{1}{|R_{(t)}|^N \, |\mathbf{\Lambda}|^N} \left(\frac{\rho}{4M}\right)^{-MN} \tag{60}$$

where λ_m are the nonzero eigenvalues of $\mathbf{\Delta}\mathbf{\Delta}^H$ and $r = \text{rank}(\mathbf{\Delta}\mathbf{\Delta}^H)$ with $\mathbf{\Delta} = \mathbf{S} - \hat{\mathbf{S}}$ denoting the codeword difference matrix. If $\text{rank}(\mathbf{\Delta}) = r < M$, *i.e.*, a rank-deficient space-time code, we can show also that

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{1}{|R_{(t)}|^N \left(\prod_{m=1}^{r} \lambda_m\right)^N} \left(\frac{\rho}{4M}\right)^{-Nr}. \tag{61}$$

From (60) (resp. (61)), we conclude that the diversity of the multiple antenna system over correlated fading is the same as that obtained for the uncorrelated fading channel. The

[2]Low-rank (full-rank) channels refer to channels with singular (non-singular) covariance matrices within this manuscript.

Inequalities (60) and (61) also reveal that there is some loss in coding gain depending on the correlation structure. Note that these results are valid when the channel is full-rank, *i.e.*, the covariance matrix of the channel is nonsingular.

3.2.2 Receive Correlation

Now, we exchange the roles of the transmitter and the receiver, that is, the transmitter is now in a richly scattering environment while the receiver is not. In this case, the correlations exist only among the subchannels from a certain transmit antenna to all receive antennas. Mathematically, the columns of $\mathbf{H}$, $\mathbf{c}_i$, are iid circularly symmetric complex Gaussian with an $N \times N$ correlation matrix

$$R_{(r)} = E\{\mathbf{c}_i \mathbf{c}_i^H\}.$$

Letting r_{ij} be the complex correlation coefficient between h_{ki} and h_{kj}, we have

$$R_{(r)} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1N} \\ r_{12}^\star & 1 & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1N}^\star & r_{2N}^\star & \cdots & 1 \end{pmatrix}.$$

The PEP bound for the this case can be expressed as (as $\rho \to \infty$)

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{1}{|R_{(r)}|^M |\mathbf{\Lambda}|^N} \left(\frac{\rho}{4M}\right)^{-MN} \tag{62}$$

if the space-time code is full-rank. For a rank-deficient space-time code, we have

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{1}{|R_{(r)}|^M \left(\prod_{m=1}^{r} \lambda_m\right)^N} \left(\frac{\rho}{4M}\right)^{-Nr}. \tag{63}$$

The Inequality (62) (resp. ((63)) is the analogous of (60) (resp. (61)) for the case of receive correlation. Once again, we note that the diversity of the multiple antenna system over correlated fading is the same as that obtained for the uncorrelated fading channel.

The Inequalities in (60) and (62) indicate that the fading correlations can not improve the system performance. This can be seen by observing $|R_{(t)}| \leq 1$ or $|R_{(r)}| \leq 1$ that follows from the Hadamard's Inequality [34]. The equality is satisfied only if the correlation

matrices are identity, which is the case for iid fading. Hence, the system performance will be the best for iid fading, and it will get worse as correlations occur among the subchannels.

To explain further, let us consider a MIMO system with $M = 2$ and $N = 2$ antennas. The correlation matrix takes the form[3] $R = \begin{pmatrix} 1 & r_{12} \\ r_{12}^\star & 1 \end{pmatrix}$, and hence $|R| = 1 - |r_{12}|^2$. The coding gain loss (in dB) is then given by $-5\log_{10}(1 - |r_{12}|^2)$. There is no loss if $r_{12} = 0$, *i.e.*, iid fading case, and the loss increases as the correlation coefficient increases, *e.g.*, we observe about 0.6 dB loss for $|r_{12}| = 0.5$ and 3.6 dB loss for $|r_{12}| = 0.9$ due to correlations. We will comment on the effects of fading correlations later in Section 3.6.

3.3 Error Probability Analysis with Antenna Selection

In this section, we present the PEP analysis for systems employing antenna selection. First, we consider the case where a single antenna is selected. Then, we generalize our analysis to the case where more than one antennas are selected.

3.3.1 PEP Analysis with Single Antenna Selection

3.3.1.1 Transmit Correlation

If only one antenna is selected out of the N receive antennas, the selection rule is reduced to choosing the antenna element that observes the largest instantaneous SNR , *i.e.*,

$$i^\star = \text{argmax}_{i=1,\cdots,N} |h_{i1}|^2 + \cdots + |h_{iM}|^2.$$

In this case, the Chernoff bound on the PEP can be expressed as

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) &\le N \int e^{-\frac{\rho}{4M}\|\mathbf{r}\mathbf{\Delta}\|^2} F_Z(\|\mathbf{r}\|^2)^{N-1} f_{\mathbf{R}}(\mathbf{r}) d\mathbf{r} \\ &= N \int e^{-\frac{\rho}{4M}\|\mathbf{r}\mathbf{\Delta}\|^2} F_Z(\|\mathbf{r}\|^2)^{N-1} \frac{1}{\pi^M |R_{(t)}|} e^{-\mathbf{r}R_{(t)}^{-1}\mathbf{r}^H} d\mathbf{r}, \end{aligned} \tag{64}$$

where $F_Z(\cdot)$ is the cumulative distribution function (cdf) of $Z = \mathbf{r}\mathbf{r}^H$. Using the Singular Value Decomposition (SVD) $\mathbf{\Delta}\mathbf{\Delta}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ in (64), and then letting $\boldsymbol{\beta} = \mathbf{r}\mathbf{U}$, we obtain

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) &\le N \int e^{-\frac{\rho}{4M}\boldsymbol{\beta}\mathbf{\Lambda}\boldsymbol{\beta}^H} F_Z(\|\boldsymbol{\beta}\|^2)^{N-1} \frac{1}{\pi^M |R_{(t)}|} e^{-\boldsymbol{\beta}\mathbf{U}^H R_{(t)}^{-1}\mathbf{U}\boldsymbol{\beta}^H} d\boldsymbol{\beta} \\ &= \frac{N}{\pi^M |R_{(t)}|} \int e^{-\boldsymbol{\beta}\left(\frac{\rho}{4M}\mathbf{\Lambda} + \mathbf{U}^H R_{(t)}^{-1}\mathbf{U}\right)\boldsymbol{\beta}^H} F_Z(\|\boldsymbol{\beta}\|^2)^{N-1} d\boldsymbol{\beta}. \end{aligned} \tag{65}$$

[3] We denote by R the transmit or receive correlation matrix, *i.e.* $R_{(t)}$ or $R_{(r)}$

To find the $F_Z(\cdot)$, we need to evaluate the probability

$$F_Z(a) = P\{Z \leq a\} = P\{|h_{i1}|^2 + \cdots + |h_{iM}|^2 \leq a\} \tag{66}$$

$$= \int_{\mathcal{C}} f_{\mathbf{R}_i}(\mathbf{r}) d\mathbf{r} \tag{67}$$

$$= \frac{1}{\pi^M |R_{(t)}|} \int_{\mathcal{C}} e^{-\mathbf{r} R_{(t)}^{-1} \mathbf{r}^H} d\mathbf{r}, \tag{68}$$

where $\mathcal{C}$ is the region $\left\{h_{i1}, \cdots, h_{iM}, \ : \ |h_{i1}|^2 + \cdots + |h_{iM}|^2 \leq a\right\}$. Using the SVD of $R_{(t)} = \mathbf{VMV}^H$ in (68), and the routine integration tools, we obtain

$$F_Z(a) = \frac{1}{|\mathbf{M}|} \int_{u_1 + \cdots + u_M \leq a} e^{-(\mu_1^{-1} u_1 + \cdots + \mu_M^{-1} u_M)} du_1 \cdots du_M. \tag{69}$$

The evaluation of this integral results in

$$F_Z(a) = 1 - \sum_{j=1}^{M} k_j e^{-\mu_j^{-1} a} \tag{70}$$

where

$$k_j = \frac{\mu_j^{M-1}}{\prod_{i=1, i \neq j}^{M} \mu_j - \mu_i}, \ j = 1, \cdots, M. \tag{71}$$

Substituting (70) in (65) and simplifying the resulting expression, we arrive at

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{\prod_{j=1}^{M} \mu_j} \int e^{-(\lambda_1' u_1 + \cdots + \lambda_M' u_M)} \left[1 - \sum_{j=1}^{M} k_j e^{-\frac{u_1 + \cdots + u_M}{\mu_j}}\right]^{(N-1)} du_1 \cdots du_M, \tag{72}$$

where $\lambda_i', \ i = 1, \cdots, M$ are the eigenvalues of $\mathbf{\Lambda}' = \frac{\rho}{4M} \mathbf{\Lambda} + \boldsymbol{\zeta}$ with $\boldsymbol{\zeta} = \mathbf{U}^H R_{(t)}^{-1} \mathbf{U}$. For specific values of N, (72) can easily be evaluated. In fact, a closed form expression for any values of M, N can also be obtained. In terms of $\mathbf{\Lambda}'$ and μ_j, the final result can be expressed as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N}{\prod_{j=1}^{M} \mu_j} \sum_{l=0}^{N-1} C(N-1, l)(-1)^l \left(\sum_{j_1=1}^{M} \cdots \sum_{j_l=1}^{M} k_{j_1} \cdots k_{j_l} \left| \mathbf{\Lambda}' + (\mu_{j_1}^{-1} + \cdots + \mu_{j_l}^{-1}) I_M \right|^{-1} \right). \tag{73}$$

Unfortunately, for the general case, the closed form expression does not give much insight about the effect of the correlation on the system performance. Therefore, in what follows, we will present a few special cases. We will also provide some numerical results later in Section 3.6.

Example 1: For $M = 2$, the cdf in (70) is given by

$$F_Z(a) = 1 - \frac{\mu_2}{\mu_2 - \mu_1} e^{-\mu_2^{-1} a} - \frac{\mu_1}{\mu_1 - \mu_2} e^{-\mu_1^{-1} a}, \text{ for } \mu_1 \neq \mu_2.$$

Using the cdf, the Chernoff bound in (73) for $M = N = 2$ can be simplified to

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2}{\mu_1 \mu_2} \left[\frac{1}{|\mathbf{\Lambda}'|} - \frac{\mu_1/(\mu_1 - \mu_2)}{|\mathbf{\Lambda}' + I/\mu_1|} - \frac{\mu_2/(\mu_2 - \mu_1)}{|\mathbf{\Lambda}' + I/\mu_2|} \right]. \tag{74}$$

For high SNR and full-rank space-time codes, *i.e.*, $\rho \to \infty$, and $\text{rank}(\mathbf{\Delta}) = M$, we can further simplify the expression as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2}{(\mu_1 \mu_2)^2} \frac{1}{(\lambda_1 \lambda_2)^2} \left(\frac{\rho}{4M} \right)^{-4}. \tag{75}$$

Hence, the diversity order remains the same as that of the full-complexity system while there is some loss in the coding gain that depends on the determinant of the correlation matrix.

If the underlying space-time code is rank-deficient, *i.e.*, $\text{rank}(\mathbf{\Delta}) = 1 < M$, then we have

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2}{(\mu_1 \mu_2) \lambda_1} \left(\frac{\rho}{4M} \right)^{-1}. \tag{76}$$

This result implies that for rank-deficient space-time codes, antenna selection degrade the diversity order significantly, *i.e.*, the diversity order with selection is 1 while it is $Nr = 2$ with the full-complexity system. ■

Example 2: When $M = 3$, assuming that $\mu_i \neq \mu_j$, for $i \neq j$, $i, j = 1, 2, 3$, we have

$$F_Z(a) = 1 - \frac{\mu_3^2}{(\mu_3 - \mu_2)(\mu_3 - \mu_1)} e^{-\mu_3^{-1} a} - \frac{\mu_2^2}{(\mu_2 - \mu_3)(\mu_2 - \mu_1)} e^{-\mu_2^{-1} a} - \frac{\mu_1^2}{(\mu_1 - \mu_2)(\mu_1 - \mu_3)} e^{-\mu_1^{-1} a}$$

When the space-time code is full-rank, the bound for $M = 3$ and $N = 2$ is given by

$$\begin{aligned} P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2}{\mu_1 \mu_2 \mu_3} \Bigg[& \frac{1}{|\mathbf{\Lambda}'|} - \frac{\mu_1^2/(\mu_1 - \mu_2)(\mu_1 - \mu_3)}{|\mathbf{\Lambda}' + I/\mu_1|} - \frac{\mu_2^2/(\mu_2 - \mu_1)(\mu_2 - \mu_3)}{|\mathbf{\Lambda}' + I/\mu_2|} \\ & - \frac{\mu_3^2/(\mu_3 - \mu_1)(\mu_3 - \mu_2)}{|\mathbf{\Lambda}' + I/\mu_3|} \Bigg]. \end{aligned} \tag{77}$$

At high SNR, this bound can be approximated by

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{2}{(\mu_1 \mu_2 \mu_3)^2} \frac{g(\lambda_1, \lambda_2, \lambda_3)}{(\lambda_1 \lambda_2 \lambda_3)^2} \left(\frac{\rho}{4M} \right)^{-6}, \tag{78}$$

where $g(\cdot)$ is a function that depends only on $\lambda_i,\ i = 1, 2, 3$. Thus, the diversity order is equal to $MN = 6$, which is equal to the diversity order of the full-complexity system. On the other hand, we can obtain Chernoff bounds, if $\text{rank}(\boldsymbol{\Delta}) = 1 < M$, as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \xi_1(\boldsymbol{\zeta}, R_{(t)}) \left(\frac{\rho}{4M}\right)^{-1}, \tag{79}$$

and if $\text{rank}(\boldsymbol{\Delta}) = 2 < M$, as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \xi_2(\boldsymbol{\zeta}, R_{(t)}) \left(\frac{\rho}{4M}\right)^{-2}, \tag{80}$$

where $\xi_1(\cdot)$ and $\xi_2(\cdot)$ depend only on $\boldsymbol{\zeta}$ and $R_{(t)}$, but not ρ. These bounds indicate the degradation in the diversity order due to antenna selection when we have low-rank space-time codes. ■

We note that if some of the eigenvalues of $R_{(t)}$ are identical, we can evaluate the integral in (69) as well. We also note that the above results are valid only if $R_{(t)}$ is nonsingular.

3.3.1.2 Receive Correlation

For this case, we will need the PEP

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \int_{\mathcal{C}^M} \exp\left(-\frac{\rho}{4M}\mathbf{r}\boldsymbol{\Delta}\boldsymbol{\Delta}^H\mathbf{r}^H\right) f_{\hat{\mathbf{R}}}(\mathbf{r})d\mathbf{r}, \tag{81}$$

where $f_{\hat{\mathbf{R}}}(\mathbf{r})$ is the pdf of the row of $\mathbf{H}$ with the maximum norm. Using a similar analysis as in Chapter 2, we obtain this pdf as

$$f_{\hat{\mathbf{R}}}(\mathbf{r}) = NP_r(\|\mathbf{r}_1\|^2 < \|\mathbf{r}\|^2, \cdots, \|\mathbf{r}_{n-1}\|^2 < \|\mathbf{r}\|^2, \|\mathbf{r}_{n+1}\|^2 < \|\mathbf{r}\|^2, \cdots, \|\mathbf{r}_N\|^2 < \|\mathbf{r}\|^2) f_{\mathbf{R}_n}(\mathbf{r}), \tag{82}$$

where $f_{\mathbf{R}_n}(\mathbf{r})$ is the marginal distribution of any row of $\mathbf{H}$, *i.e.*,

$$f_{\mathbf{R}_n}(\mathbf{r}) = \int f_{\mathbf{H}}(\mathbf{r}_1, \cdots, \mathbf{r}_{n-1}, \mathbf{r}_{n+1}, \cdots, \mathbf{r}_N)d\mathbf{r}_1 \cdots d\mathbf{r}_{n-1}d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N,$$

and $P_r(\cdot)$ is the probability given by

$$P_r(\cdot) = \int_\Omega f_{\mathbf{H}}(\mathbf{r}_1, \cdots, \mathbf{r}_{n-1}, \mathbf{r}_{n+1}, \cdots, \mathbf{r}_N)d\mathbf{r}_1 \cdots d\mathbf{r}_{n-1}d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N$$

where Ω is the region

$$\Omega = \left\{\mathbf{r}_i : \|\mathbf{r}_i\|^2 < \|\mathbf{r}\|^2,\ i = 1, \cdots, n-1, n+1, \cdots, N\right\}.$$

Unlike the case of the transmit correlation where it is possible to obtain simpler expressions for the PEP, further simplification does not seem to be warranted. However, we can perform numerical calculations to evaluate the PEP given by (81), which we will do in Section 3.6.

3.3.2 Selection of More Than One Antenna

For this case, since the receive correlation model does not allow a mathematically tractable analysis, we will study only the case of fading correlation at the transmitter side. For fading correlation at the receiver side, we will resort to numerical simulations.

3.3.2.1 Transmit Correlation

We now obtain the bounds on the PEP for the case when L out of N antennas are selected. Using a similar line of argument as in Chapter 2, we arrive at

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \sum_{l=1}^{L}\int_{\mathcal{R}_l} e^{-\frac{\rho}{4M}\|\tilde{\mathbf{H}}\Delta\|^2}\frac{N!}{(N-L)!L!L}\left[1-\sum_{j=1}^{M}k_j^{-\|\mathbf{r}_l\|^2/\mu_j}\right]^{N-L}\times \frac{1}{\pi^{ML}|R_{(t)}|^L}e^{-(\mathbf{r}_1R_{(t)}^{-1}\mathbf{r}_1^H+\cdots+\mathbf{r}_LR_{(t)}^{-1}\mathbf{r}_L^H)}d\mathbf{r}_1\cdots d\mathbf{r}_L, \tag{83}$$

which can be re-written as

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \sum_{l=1}^{L}\int_{\mathcal{R}_l} e^{-\sum_{j=1}^{L}\mathbf{r}_j\left(\frac{\rho}{4M}\mathbf{\Lambda}+\mathbf{U}^HR_{(t)}^{-1}\mathbf{U}\right)\mathbf{r}_j^H}\frac{N!}{(N-L)!L!L}\left[1-\sum_{j=1}^{M}k_je^{-\|\mathbf{r}_l\|^2/\mu_j}\right]^{N-L}\times \frac{1}{\pi^{ML}|R_{(t)}|^L}d\mathbf{r}_1\cdots d\mathbf{r}_L, \tag{84}$$

where the region $\mathcal{R}_l$ is defined as

$$\mathcal{R}_l = \{\mathbf{r}_1,\cdots,\mathbf{r}_L : \|\mathbf{r}_l\| < \|\mathbf{r}_k\|,\ \ k=1,\cdots,l-1,l+1,\cdots,L\},$$

and $\tilde{\mathbf{H}}$ is the $L\times M$ matrix formed by deleting the rows of $\mathbf{H}$ corresponding to the antennas that are not selected.

Analytical evaluation of this integral over $\mathcal{R}_l$ is a formidable task. Integrating over the whole space, although resulting in a looser bound, yields a mathematically tractable

analysis. In this case, integration over $\mathcal{R}_l$ will not depend on l and the analysis results in the following upper bound

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \frac{N!}{(N-L)!L!}\frac{1}{|R_{(t)}|^L|\Lambda'|^{L-1}}\left(\sum_{l=0}^{N-L} C(N-L,l)(-1)^l \times \left(\sum_{j_1=1}^{M}\cdots\sum_{j_l=1}^{M} k_{j_1}\cdots k_{j_l}\left|\mathbf{\Lambda}' + (\mu_{j_1}^{-1}+\cdots+\mu_{j_l}^{-1})I_M\right|^{-1}\right)\right). \tag{85}$$

Since further simplification of this expression is not analytically tractable, we have to resort to numerical examples to present the system performance. However, it is possible to obtain simpler expressions for special cases as shown in the next example.

Example 3: When $M = N = 3$ and $L = 2$, the bound in (85) can be written as

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \frac{3}{(\mu_1\mu_2\mu_3)^2|\mathbf{\Lambda}'|}\left[\frac{1}{|\mathbf{\Lambda}'|} - \frac{\mu_1^2/(\mu_1-\mu_2)(\mu_1-\mu_3)}{|\mathbf{\Lambda}'+I/\mu_1|} - \frac{\mu_2^2/(\mu_2-\mu_1)(\mu_2-\mu_3)}{|\mathbf{\Lambda}'+I/\mu_2|} - \frac{\mu_3^2/(\mu_3-\mu_1)(\mu_3-\mu_2)}{|\mathbf{\Lambda}'+I/\mu_3|}\right]. \tag{86}$$

At high SNR, we can further simplify this bound using (78) to obtain

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \frac{3}{(\mu_1\mu_2\mu_3)^2|\mathbf{\Lambda}'|}\left[\frac{g(\lambda_1,\lambda_2,\lambda_3)}{\mu_1\mu_2\mu_3(\lambda_1\lambda_2\lambda_3)^2}\left(\frac{\rho}{4M}\right)^{-6}\right]. \tag{87}$$

Finally, letting $\rho\to\infty$, we arrive at

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \frac{3g(\lambda_1,\lambda_2,\lambda_3)}{(\mu_1\mu_2\mu_3)^3(\lambda_1\lambda_2\lambda_3)^3}\left(\frac{\rho}{4M}\right)^{-9}. \tag{88}$$

Hence, the diversity order of the system using antenna selection is $MN = 9$, which is equal to the diversity order achieved by the full-complexity system.

For rank-1 and rank-2 space-time codes, *i.e.*, $\text{rank}(\mathbf{\Delta}) = 1$ or 2, the asymptotic performance is given by, respectively,

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \xi_1'(\boldsymbol{\zeta}, R_{(t)})\left(\frac{\rho}{4M}\right)^{-2}, \tag{89}$$

and

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \xi_2'(\boldsymbol{\zeta}, R_{(t)})\left(\frac{\rho}{4M}\right)^{-4}, \tag{90}$$

where $\xi_1'(\cdot)$ and $\xi_2'(\cdot)$ can be obtained in a similar fashion as $\xi_1(\cdot)$ and $\xi_2(\cdot)$. The expressions in (89) and (90) indicate that the diversity order with antenna selection is degraded significantly when the space-time code is rank-deficient. ■

3.4 Performance of MIMO systems with Antenna Selection over Non-Full-Rank Channels

Note that in the above analysis, we considered a positive definite correlation matrix R for the channel model, that is, all eigenvalues of R are assumed to be positive. Since the covariance matrix was nonsingular, the pdf could easily be written and the analysis in the previous sections followed. However, when we have a singular covariance matrix for a multivariate Gaussian vector, it is not possible to write a density function [4]. Therefore, we are in need of different methods to study the non-full-rank channels.

Let us consider the transmit correlation. We use the SVD of $R_{(t)}$ to pose an equivalent problem,

$$R_{(t)} = \mathbf{V}\mathbf{M}\mathbf{V}^H \tag{91}$$

$$= \mathbf{V}\mathbf{M}^{1/2}\mathbf{M}^{1/2}\mathbf{V}^H \tag{92}$$

$$= \mathbf{Q}\mathbf{Q}^H. \tag{93}$$

Note that $R_{(t)}$ may be singular in which case some of the eigenvalues may be zero. Now, we can model an MIMO system that is statistically equivalent to the one given in (59) as

$$\mathbf{X}' = \sqrt{\rho/M}\mathbf{H}'\mathbf{S}' + \mathbf{W}', \tag{94}$$

where $\mathbf{S}' = \mathbf{Q}^H\mathbf{S}$. $\mathbf{H}'$ and $\mathbf{W}'$ channel transfer matrix and additive white Gaussian noise vector with iid $\mathcal{CN}(0,1)$ entries, respectively. By statistical equivalence, we mean that the error rate performance for the two systems are the same. The Chernoff bound on the PEP for the equivalent full-complexity system can be obtained as [163]

$$P(\mathbf{S}' \to \hat{\mathbf{S}}') \leq \frac{1}{|I + \rho/4M\mathbf{Q}^H\mathbf{\Delta}\mathbf{\Delta}^H\mathbf{Q}|^N}. \tag{95}$$

For high SNR, this bound can be approximated by

$$P(\mathbf{S}' \to \hat{\mathbf{S}}') \leq \frac{1}{|\mathbf{Q}^H\mathbf{\Delta}\mathbf{\Delta}^H\mathbf{Q}|_+^N}\left(\frac{\rho}{4M}\right)^{-Nr} \tag{96}$$

where $r = \text{rank}(\mathbf{Q}^H\mathbf{\Delta}\mathbf{\Delta}^H\mathbf{Q})$, and $|A|_+$ denotes the product of the nonzero eigenvalues of A. Note that if the space-time code is full-rank, *i.e.*, $\text{rank}(\mathbf{\Delta}) = M$, then

$$\text{rank}(\mathbf{Q}^H\mathbf{\Delta}) = \text{rank}(\mathbf{Q}^H) = \text{rank}(\mathbf{Q}).$$

This last equality formally states that when some or all of the subchannels are correlated such that the correlation coefficients have unit magnitude, the diversity gain of the MIMO system is reduced to Nr where r is the rank of the covariance matrix of the channel.

The extension of these arguments to the case of MIMO antenna selection is not trivial. Consider, for example, the selection criteria for the equivalent system

$$\mathbf{r}'^{\star} = \text{argmax}_{\mathbf{r}'_i, i=1,\cdots,N} \mathbf{r}'_i \mathbf{Q}^H \mathbf{Q} \mathbf{r}'^H_i. \tag{97}$$

It is clear that rather than using the power captured by the receive antennas, we use a weighted power-sum metric that takes the structure of the correlation into account. As we mentioned, due to the inherent problems in expressing the pdf of the low-rank multivariate Gaussian, a complete analysis for this case is not analytically tractable. Therefore, we resort to Monte-Carlo simulations to assess the performance of the system over low-rank channels. One might expect to observe a severe reduction in the diversity order as it is the case for the iid fading case (recall that the diversity gain of the MIMO system with selection over independent fading Lr where L is the number of the antennas selected and r is the rank of the underlying space-time code). As we will demonstrate shortly, the selection under the assumption of fully correlated fading achieves the same diversity advantages as that achieved by the full-complexity system, as long as the space-time code is full-rank.

3.5 Space-Time Code Design with Antenna Selection Under Correlated Fading

For the full-complexity system over an iid quasi-static fading channel, Tarokh *et al.* have proposed two design criteria based on the upper bounds on the PEP [163]: *rank* and *determinant* criteria. The space-time code design principles for correlated fading channels, on the other hand, have not been studied in detail yet. Nevertheless, several precoding techniques have been studied for MIMO systems to deal with correlated fading channels [21, 58, 87, 116, 134]. In [87], for example, Hong *et al.* propose several precoding schemes to improve the performance of spatial multiplexing systems, but their approach is not applicable for general space-time coding schemes.

The bounds derived in the previous sections for correlated fading channels can be used

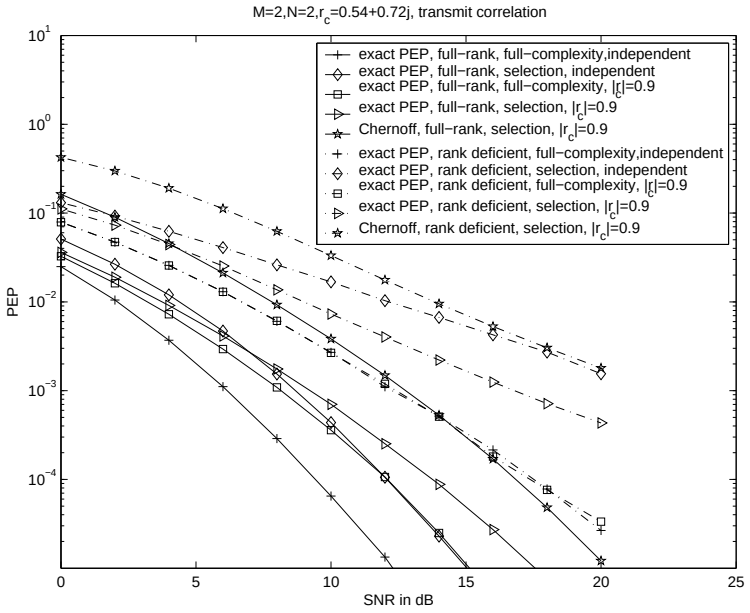


Figure 13: PEP vs. SNR for single antenna selection for MIMO channel with transmit correlation. Simulation parameters are: $M = N = 2$, $r_c = 0.54 + 0.72j$.

to develop space-time codes. The asymptotic analysis for both the full-complexity system and the system with antenna selection has shown that the effect of correlations on the PEP can be abstracted as a multiplicative term in the Chernoff bounds; namely, the asymptotic PEP is scaled by the determinant of the correlation matrix. This implies that the criteria for designing space-time codes for iid MIMO fading channels directly apply to the case of correlated MIMO fading channels. In addition, the PEP bounds derived in Section 3.3.1.1 or 3.3.2.1, *e.g.*, Inequalities (75),(78) and (87), can be used to design the space-time codes that have the largest coding gains for MIMO systems with energy-based antenna selection.

3.6 Examples

In the previous sections, we have theoretically analyzed the performance and derived several bounds on the PEP. We now evaluate those bounds for several codeword pairs that are selected from the codes developed in [163]. We present the effect of fading correlation (including the case of both full-rank nonfull-rank space-time codes), and compare the performance of the system against that of the same system over uncorrelated fading.

In Figures 13–14, we compare the PEP bounds for full-complexity system and the one

using antenna selection for the case of transmit correlation. In these figures, the performance of a full-rank code (solid lines) and a rank-deficient code (dashed lines) is presented. Figure 13 illustrates the results for the case of double-transmit and the double-receive antenna system when $r_c = 0.54 + 0.72j$. The full-rank space-time codeword pairs are selected from the 2 bits/sec/Hz 8−state space-time trellis codes using 4−PSK modulation (with $M = 2$). This code provides a diversity advantage of 6 [163], *i.e.,* full spatial diversity. The two codewords considered differ in three consecutive symbols. For comparison purposes, we also present the PEP bounds when there is no correlation, *i.e.*, iid fading. We observe that even for a high level of correlation, *i.e.*, $|r_c| = 0.9$, although there is some loss in the coding gain, the diversity orders of both the full-complexity system and the system using antenna selection are the same. The Chernoff bound evaluated using (74) is also plotted, and it is about 2 dB away from the exact PEP. From the PEP curves obtained for rank-deficient space-time codeword pairs, we observe that (i) the performance of the full-complexity system under correlated fading and iid fading is very close to each other, while the performance of the system with antenna selection is superior when there is correlated fading, and (ii) the diversity order is reduced when antenna selection is performed.

The PEP for the case of $M = 3$ transmit and $N = 2$ receive antennas is presented in Figure 14. We assume that the channel correlation matrix is given by

$$R_{(t)} = \begin{pmatrix} 1 & 0.6 & 0.4 \\ 0.6 & 1 & 0.45 \\ 0.4 & 0.45 & 1 \end{pmatrix} \tag{98}$$

For this correlation structure, we observe that the loss in the coding gain due to antenna selection is about 4 dB for the case of both iid fading and correlated fading. We note that the Chernoff bound (77) for the system with antenna selection is about 2 dB away from the exact PEP. The bound, on the other hand, is not as strict as this when the space-time code is rank deficient, *i.e.*, it is about 6 dB away from the exact PEP.

In Figure 15, we plot the Chernoff bounds on the PEP for the full-complexity system and the system with antenna selection for various values of the correlation coefficient. We observe an increasing loss in the coding gain as the strength of the correlation increases

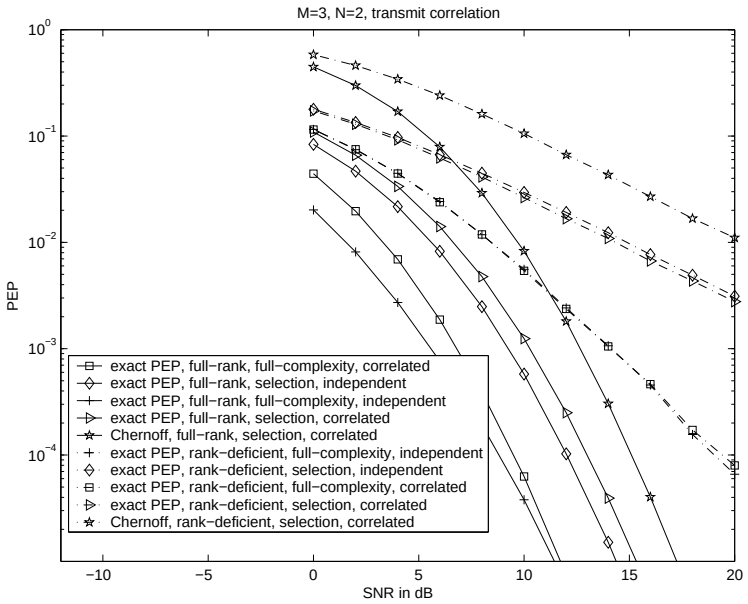


Figure 14: PEP vs. SNR for single antenna selection for MIMO channel with transmit correlation. Simulation parameters are: $M = 3$, $N = 2$.

while the diversity of the system remains the same. The performance of the antenna selection system when $r_c = r_{12} = 0.5$ is very close (within $0.6dB$) to that when $r_c = 0$, *i.e.*, independent fading. However, for channels having higher correlation, we observe degradation of at least 2 dB for $|r_c| \geq 0.8$ at a BER of 10^{-5}. Note from the slopes of the PEP curves that the diversity of the system remains the same for various correlation levels in the channel.

The performance of a system over a fading channel with receive correlation is presented in Figure 16. We use the same correlation matrix given by (98). We use Monte-Carlo integration of (81) to obtain the PEP curves for the system using antenna selection. We present the performance for both a rank-deficient space-time code and a full-rank space-time codes. The conclusions on the diversity orders and the coding gains for this fading model is similar to the previous results. In addition, we note that (i) performing optimal selection that maximizes the instantaneous channel capacity gives very similar performance as that obtained by SNR-based selection provided that the space-time code is full-rank, (ii) and if the space-time code is rank deficient, one can obtain a lower PEP by using capacity-based criterion for selection; however, the diversity order achieved with optimal selection is the

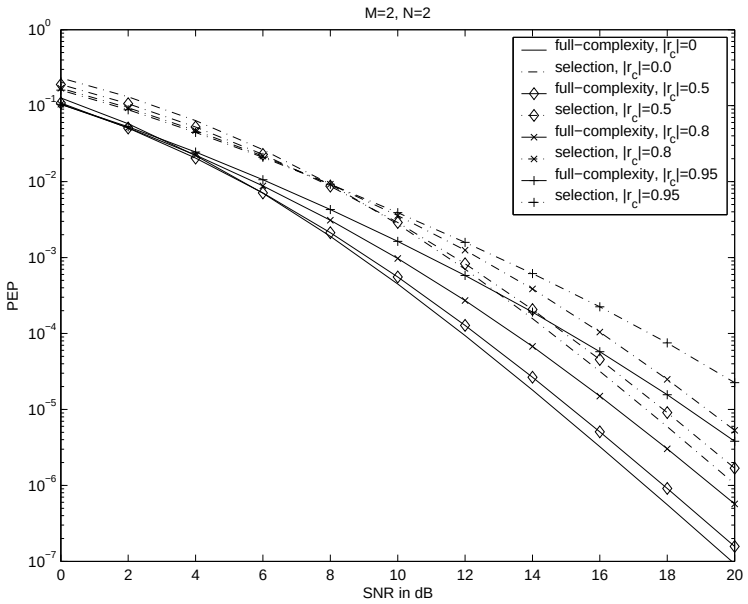


Figure 15: SNR vs. PEP for different r_c values

same as that achieved by SNR-based selection.

In Figure 17, we present the FER results for various r_c when the 2 bits/s/Hz space-time code in [163] that employs 8-state trellis and 4-PSK is used. A performance degradation of about 0.5-dB (resp. 3.8 dB) is observed for this case when the correlation increases from 0 to 0.5 (resp. 3.8 dB), which is almost the same as expected from the theoretical results. The diversity order remains the same with antenna selection even under heavy correlation. However, for low-rank channels, *e.g.*, $|r_c| = 1$, we observe significant loss in the diversity gain.

In Figures 18 and 19, we present the effect of low-rank channels on the performance of MIMO systems with antenna selection. We first consider the transmit fading correlation (See Figure 18). We assume $M = 3$ and $\text{rank}(R_{(t)}) = 1$, *i.e.*, the branches are correlated having a correlation coefficient of magnitude-1. The following observations are in order: (i) The diversity order achieved by the full-complexity system can also be achieved by the system using antenna selection. Recall that if the space-time code were rank-deficient, the diversity order achieved with antenna selection could significantly decrease; however, when the channel is rank-deficient, we can still obtain the same diversity order with antenna selection; (ii) As the number of receive antennas, N, increases, the diversity order increases

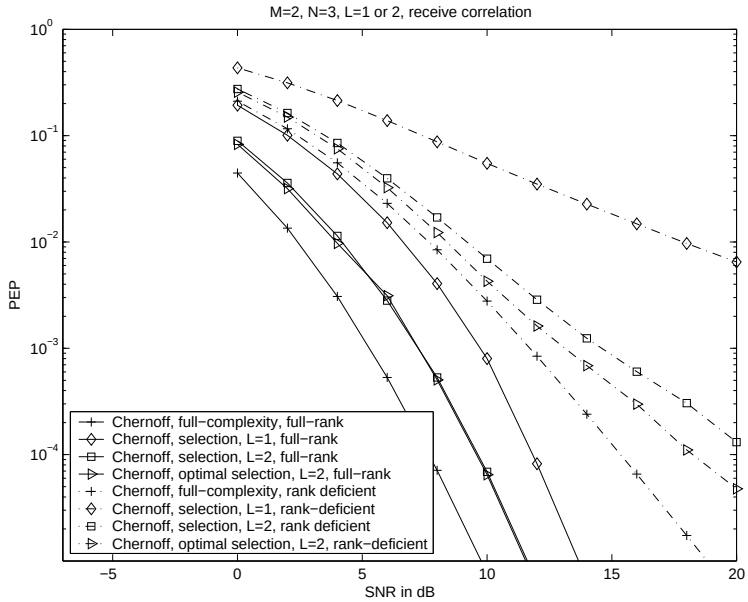


Figure 16: PEP vs. SNR for single and double antenna selection for MIMO channel with receiver correlation. Simulation parameters are: $M = 2$, $N = 3$.

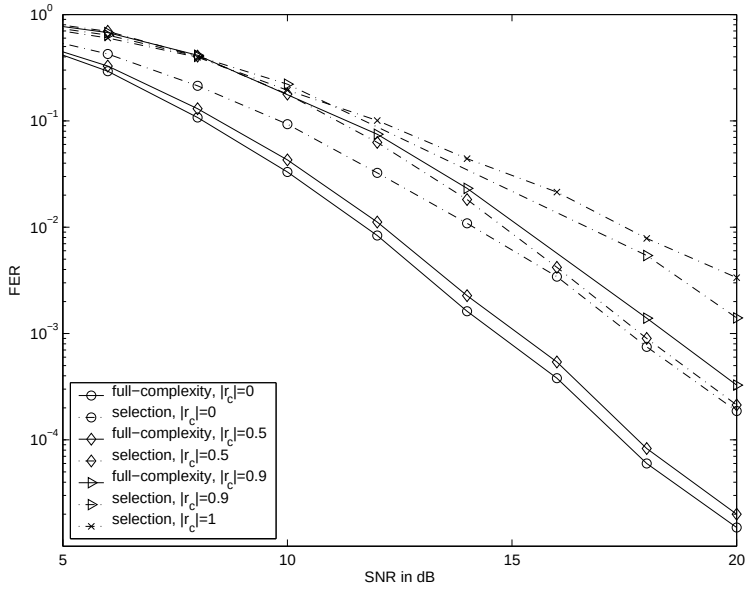


Figure 17: SNR vs. FER for different correlation coefficient values. Simulation paramters are: $M = 2$, $N = 2$, space-time codes from [163] are employed.

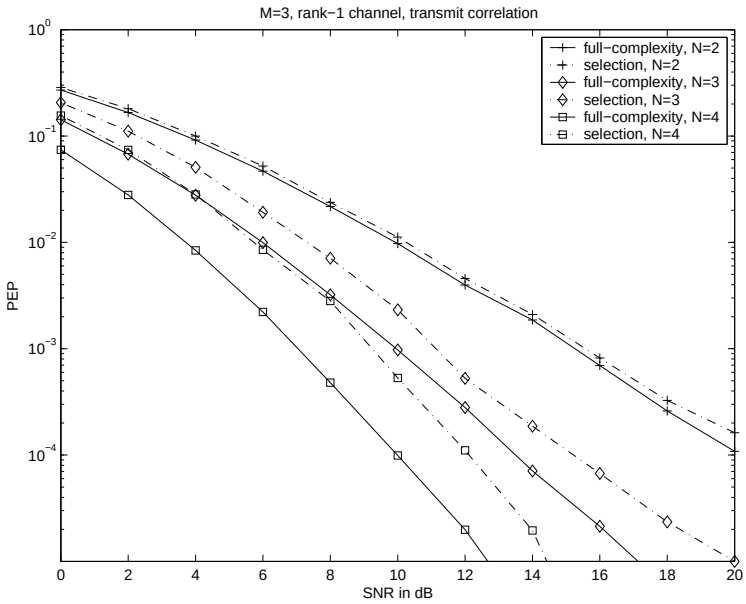


Figure 18: PEP vs. SNR for transmit correlation when the channel-rank = 1, i.e., rank$R_{(t)} = 1$. Simulation parameters are $M = 3$, $N = 2, 3, 4$.

accordingly (for both full-complexity system and the system using antenna selection), but we experience more and more coding gain loss.

Figure 19 shows the Chernoff bounds on the PEP for a system with $M = 2$ transmit antennas and $N = 3$ and $N = 4$ receive antennas. $L = 2$ antennas are selected for systems using antenna selection. We plot the PEPs also for the optimal selection that maximizes the instantaneous channel capacity. To show the effect of rank-deficiency in the space-time code, we provide the PEPs for a rank-1 space-time code as well. From the plots, the following observations are made: (i) For a low-rank channel with a correlation matrix of rank-r, the diversity order remains as Mr, where $r = \text{rank}(R_r)$, as we increase the number of receive antennas; (ii) For the full-rank space-time code, the performance of optimal selection is either very close to that of the SNR-based selection or it is slightly worse, i.e, SNR-based selection performs about 0.7 dB better than optimal selection for $N = 4$; (iii) For rank-deficient space-time codes over low-rank channels, optimal selection achieve the same diversity order as that achieved by the full-complexity system, but SNR-based selection experiences some loss in the diversity gain.

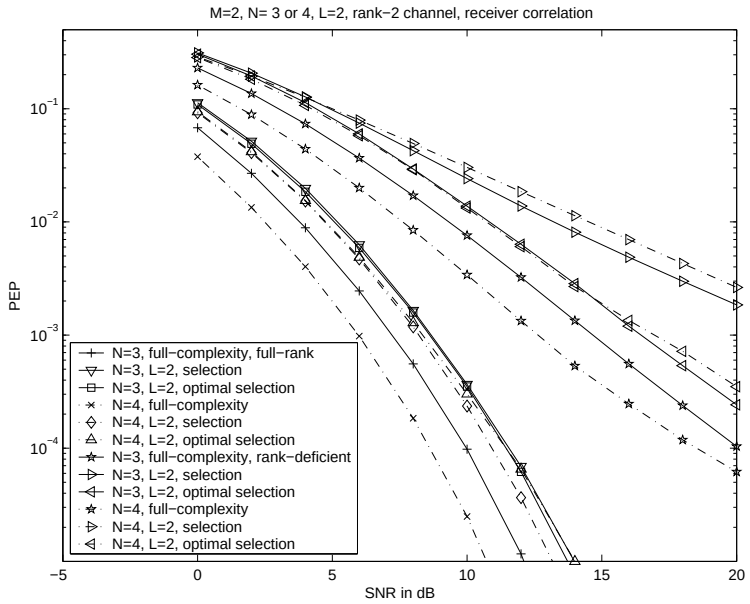


Figure 19: PEP vs. SNR for receive correlation when the channel-rank = 2, i.e., rank$R_{(r)} = 2$. Simulation parameters are $M = 2$, $N = 3, 4$, $L = 2$.

3.7 Chapter Summary

We analyzed the performance of MIMO systems with antenna selection under correlated fading channels. We considered a semicorrelated fading channel model assuming the presence of transmit or receive correlation. We derived closed-form expressions for the Chernoff-bounds on the PEP for the case of transmit correlation while the analysis for receive correlation did not allow for closed form solutions. The analysis and numerical examples for the system employing antenna selection has shown that the correlation between subchannels degrade the coding gain of the system but does not effect the diversity advantage as long as the channel is full-rank. For low-rank space-time codes, however, there may be considerable loss in the diversity order when antenna selection is performed.

CHAPTER IV

ANTENNA SELECTION FOR SPACE-TIME CODED OFDM SYSTEMS

In this chapter, we study the receive antenna selection for space-time coded orthogonal frequency division multiplexing (OFDM) systems that employ multiple transmit and receive antennas. We derive explicit closed-form upper bounds on the pairwise error probability (PEP). The organization of this chapter is as follows: In the next section, we summarize the related work on OFDM systems with MIMO antennas. Section 4.2 introduces the channel model and summarizes the necessary background. In Section 4.3, we present the performance analysis for MIMO-OFDM systems using the antenna selection. The simulation results justifying the theoretical analysis are presented in Section 4.5. Finally, we conclude the chapter in Section 4.6.

4.1 Introduction

Most of the space-time codes have been proposed for flat-fading channels [151,163], however, many wireless channels are frequency-selective, which complicates the system design. This is even the case for single transmit-single receive antenna systems. In an effort to simplify the signaling for frequency-selective fading channels, orthogonal frequency division multiplexing (OFDM) technique, which transforms the frequency-selective fading channel into a flat-fading one, can be used. By suitable incorporation of MIMO signaling (*i.e.*, space-time coding) and OFDM, one can obtain great flexibility to design space-time codes for such channels and achieve very high rates

The first space-time coded OFDM scheme has been proposed in [1] where the authors argue that the diversity and coding gain of the MIMO system is preserved with OFDM signaling compared to the flat fading case. In [20], the capacity of an OFDM-based MIMO system is studied. The performance analysis of a space-time coded OFDM system in terms

of its pairwise error probability is presented in [107] where the diversity or the coding gains are quantified using the Chernoff bound and it is shown that the maximum diversity order is the product of the number of transmit and receive antennas and the order of the channel. Various space-time coded OFDM schemes have also been proposed in [19,104,106,114,115].

The improvement in the performance using MIMO technology comes with an increased cost due to the necessity of multiple RF chains used for implementation, which is also the case for MIMO-OFDM systems. While MIMO antenna selection has been studied extensively in the recent literature [57, 65, 113], most of these work assume flat-fading channels. Recently, antenna selection technique has also been proposed for frequency-selective channels in [72], and in the context of space-time coded OFDM in [47, 143, 166]. The work in [166] studies an adaptive antenna selection scheme for multiple antenna systems using adaptive modulation and OFDM. In [143], Shao *et al.* study the error-rate performance of MIMO-OFDM using a capacity-based antenna selection. Although MIMO-OFDM systems will be an important part in future communication systems, the work on antenna selection for MIMO-OFDM systems is still very limited.

A general error probability analysis for MIMO-OFDM systems using antenna selection is quite difficult. For instance, the selection criteria may be quite complicated due to the large number of the subchannels between antenna pairs and the multitones related to each subchannel. In this chapter, we consider the energy-based antenna selection technique for such systems. We study the performance of the system by explicitly deriving upper bounds on the pairwise error probability. This chapter can be imagined as the extension to results of Chapter 2 where we presented the performance of a *single carrier* MIMO system over a flat-fading channel using a similar criteria. This chapter derives a more general expression for the Chernoff bound on the pairwise error probability, which reduces to the previous result when the order of the channel and the number of carrier frequencies are set to 1. It turns out that it is difficult to make remarks about the diversity order since the expressions are not simple. However, for the special case of double transmit diversity, we explicitly show that (i) with single antenna selection out of $N = 2$ receive antennas, and (ii) with the selection of $N - 1$ antennas out of N receive antennas, one can attain the same diversity

order as that attained by the full-complexity system. For some other cases, we illustrate via numerical results that diversity order is preserved with antenna selection. The proposed antenna selection technique can be used to reduce the implementation cost of the multiple-input multiple-output OFDM systems while resulting in no degradation in the asymptotic performance of the system.

Notation: Throughout the chapter, we denote the Kronecker product by $\otimes$, the Hermitian transpose of A by A^H, the $K \times K$ identity matrix by I_K, and the trace operator by $\mathrm{tr}(\cdot)$.

4.2 Channel and Signal Model

We consider a wireless link equipped with M transmit and N receive antennas (see Figure 20). OFDM is used to transmit symbols output by a space-time code (STC). We adopt the signal model from [107] and briefly summarize the necessary parts here. Each STC codeword spans P time slots, *i.e.*, P OFDM words. Each OFDM symbol is assumed to consist of K subcarriers. Hence, MK symbols are transmitted simultaneously at each channel use. At the receiver, we perform matched filtering, sampling and IDFT to obtain the discrete-time received signal

$$\mathbf{y}[p,k] = \mathbf{H}[p,k]\mathbf{S}[p,k] + \mathbf{n}[p,k]\,, k = 0, \cdots, K-1,\ p = 1, \cdots, P \tag{99}$$

where $\mathbf{H} \in \mathcal{C}^{M \times N}$, $\mathbf{S}[p,k]$ and $\mathbf{y}[p,k]$ are channel gain matrix, transmitted signal vector, and received signal vector, respectively, at the p^{th} time slot and k^{th} subcarrier. $\mathbf{n}[p,k]$ is the additive circularly symmetric complex Gaussian noise with zero mean and unit variance.

We assume that the channel gains are constant during each OFDM word while changing from one OFDM word to another, and that the subchannels between antenna pairs experience independent fading. The complex time domain impulse response from j^{th} transmit to i^{th} receive antenna, h_{ij}, can be described by a tapped-delay line

$$h_{ij}(t,\tau) = \sum_{l=1}^{L_f} \alpha_{i,j}(l,t)\delta\left(\tau - \frac{n_l}{K\Delta_f}\right)$$

where $\alpha_{i,j}(l,t)$ is the complex amplitude of the l^{th} nonzero tap that has a delay of $n_l/(K\Delta_f)$, Δ_f is the frequency spacing between OFDM subcarriers, and n_l is an integer that depends

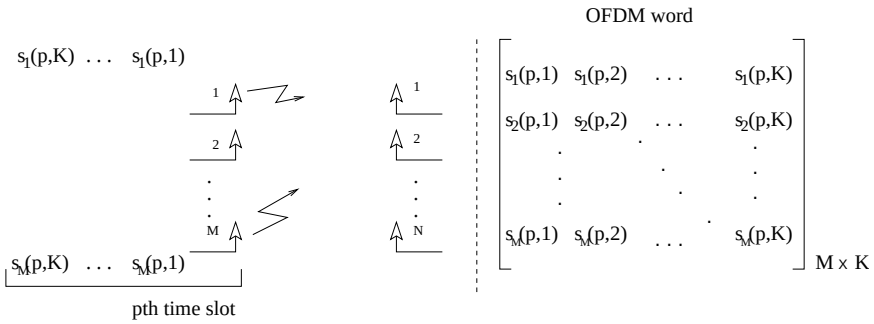


Figure 20: MIMO-OFDM system: We have M transmit antennas and N receive antennas. Each OFDM word consists of MK STC symbols, and they are transmitted simultaneously during one time slot. $s_i(p,k)$ denotes the symbol transmitted from ith transmit antenna at time slot p and subcarrier k. Each STC codeword spans P time slots, and hence consists of PMK symbols.

on the channel delay profile. Considering widesense stationary uncorrelated scattering (WSSUS) with a band-limited Doppler power spectrum, we have the frequency response at time t as

$$H_{ij}(t,f) = \sum_{l=1}^{L_f} \alpha_{i,j}(l,t) \exp\left(-j2\pi f \frac{n_l}{K\Delta_f}\right). \tag{100}$$

For OFDM systems that have proper cyclic extension, proper sample timing and tolerable leakage, the $(i,j)^{th}$ entry of the channel gain matrix in (99) is given by [107]

$$\begin{aligned} H_{i,j}[p,k] &= H_{i,j}(pT, k\Delta_f) \\ &= \sum_{l=1}^{L_f} \alpha_{i,j}(l;pT) e^{-j2\pi k n_l/K} \\ &= \mathbf{h}_{i,j}^H(p)\mathbf{w}_f(k) \end{aligned} \tag{101}$$

where $\mathbf{h}_{i,j}(p) = [\alpha_{i,j}(1;pT), \cdots, \alpha_{i,j}(L_f;pT)]^H$ is $L_f \times 1$ vector for the non-zero taps and $\mathbf{w}_f(k)$ is the corresponding DFT coefficients. Using Karhunen-Loève (KL) expansion, the time- and band-limited random process $\alpha_{i,j}(l;t), t \in [0, PT]$ can be approximated by

$$\alpha_{i,j}(l;t) \simeq \sum_{n=-f_dPT}^{f_dPT} \beta_{i,j}(l,n) e^{j2\pi nt/(PT)} \tag{102}$$

$$= \boldsymbol{\beta}_{i,j}^H(l)\mathbf{w}_t(p) \tag{103}$$

where $\boldsymbol{\beta}_{i,j}(l) = [\beta_{i,j}(l,-f_dPT), \cdots, \beta_{i,j}(l,f_dPT)]_{L_t\times 1}^H$ is a complex vector with $\beta_{i,j}(l,n)$ circularly symmetric complex Gaussian random variables, $L_t = \lceil 2f_dPT+1 \rceil$ is the number of

significant eigenvalues in the KL-expansion, and $\mathbf{w}_t(p)$ is the corresponding DFT coefficient vector. Using (103), we can rewrite (101) as

$$H_{i,j}[p,k] = \mathbf{g}_{i,j}^H \mathbf{W}_t'(p)\mathbf{w}_f(k) \tag{104}$$

where $\mathbf{g}_{i,j} = [\boldsymbol{\beta}_{i,j}^H(1), \cdots, \boldsymbol{\beta}_{i,j}^H(L_f)]_{L_tL_f\times 1}^H$ and $\mathbf{W}_t'(p) = I_{L_f} \otimes \mathbf{w}_t(p)$.

We next consider the pairwise error probability (PEP) for the space-time coded OFDM system described by Equation (99). With perfect CSI at the receiver, the maximum likelihood (ML) decoding rule is given by

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{p=1}^{P}\sum_{k=0}^{K-1} \left| y_i[p,k] - \sum_{j=1}^{M} H_{i,j}[p,k]s_j[p,k] \right|^2 . \tag{105}$$

Using the Chernoff bound, the PEP conditioned on $\mathbf{H}$ can be upper bounded by [107]

$$P(\mathbf{S} \to \hat{\mathbf{S}}|\mathbf{H}) \le \exp\left(-\frac{d^2(\mathbf{S},\hat{\mathbf{S}})\rho}{8M}\right) \tag{106}$$

where ρ is the total transmitted signal power, and $d^2(\mathbf{S},\hat{\mathbf{S}})$ is the Euclidean distance given by

$$d^2(\mathbf{S},\hat{\mathbf{S}}) = \sum_{i=1}^{N} \bar{\mathbf{g}}_i^H \mathbf{D} \bar{\mathbf{g}}_i \tag{107}$$

where

$$\mathbf{D} = \left\{ \sum_{p=1}^{P}\sum_{k=0}^{K-1} \mathbf{W}_t(p)\mathbf{W}_f(k)\mathbf{e}[p,k]\mathbf{e}^H[p,k]\mathbf{W}_f^H(k)\mathbf{W}_t^H(p) \right\}$$

and $\bar{\mathbf{g}}_i = [\mathbf{g}_{i,1}^H, \cdots, \mathbf{g}_{i,M}^H]_{ML\times 1}^H$, $\mathbf{W}_f(k) = I_M \otimes \mathbf{w}_f(k)$, $\mathbf{W}_t(p) = I_M \otimes \mathbf{W}_t'(p)$, $\mathbf{e}[p,k] = \mathbf{S}[p,k] - \hat{\mathbf{S}}[p,k]$, and $L = L_tL_f$. By averaging the conditional PEP in (106) over the channel distribution, one can obtain the average PEP as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \le \left(\frac{1}{\prod_{j=1}^{r}\left(1 + \frac{\lambda_i\rho}{8M}\right)} \right)^N \tag{108}$$

where $r = \operatorname{rank}(\mathbf{D})$ and $\lambda_i,\ i = 1, \cdots, r$, are the eigenvalues of $\mathbf{D}$. Noting that $\min_{\mathbf{S},\hat{\mathbf{S}}} r \le \min\{ML, D_{eff}\}$, where D_{eff} is the effective length of the space-time code [137], the maximum diversity gain using the MIMO-OFDM over an L^{th} order channel is NML.

4.3 Performance Analysis for MIMO-OFDM using Antenna Selection

In this section, we derive performance bounds for the space-time coded OFDM system with receive antenna selection. We assume that the CSI is exactly available at the receiver, and thus the antennas that observe the larger instantaneous SNR are selected for decoding, that is $\xi_i = \sum_{k,p} \sum_{j=1}^{M} |H_{i,j}[p,k]|^2$ are computed for each $i, i = 1, \cdots, N$, and the antennas that correspond to larger ξ_i are selected. Assuming that L_r out of N receive antennas are selected, the Chernoff bound on the average PEP can be expressed as

$$P(\mathbf{S} \rightarrow \hat{\mathbf{S}}) \leq E_{\mathbf{H}} e^{-\frac{\rho}{8M} \sum_{i=1}^{L_r} \bar{\mathbf{g}}_i^H \mathbf{D} \bar{\mathbf{g}}_i}. \tag{109}$$

In order to evaluate (109), we need to evaluate the joint probability density function (pdf) of the channel coefficients corresponding to the selected antennas. Using a technique similar to the one in Chapter 2, we define two auxiliary events:

1. $\mathcal{F} = \{\bar{\mathbf{g}}_{i_1}^H \mathbf{D}_0 \bar{\mathbf{g}}_{i_1}, \cdots, \bar{\mathbf{g}}_{i_{L_r}}^H \mathbf{D}_0 \bar{\mathbf{g}}_{i_{L_r}}$ have the largest values among all antennas$\}$

2. $\mathcal{A}_l = \{\bar{\mathbf{g}}_{i_l}^H \mathbf{D}_0 \bar{\mathbf{g}}_{i_l}$ is the minimum of $\bar{\mathbf{g}}_{i_j}^H \mathbf{D}_0 \bar{\mathbf{g}}_{i_j}$ $j = 1, \cdots, L_r$ $\}$ where

$$\mathbf{D}_0 = \left\{ \sum_{p=1}^{P} \sum_{k=0}^{K-1} \mathbf{W}_t(p) \mathbf{W}_f(k) \mathbf{W}_f^H(k) \mathbf{W}_t^H(p) \right\}_{ML \times ML} \tag{110}$$

$$= I_M \otimes \sum_{p=1}^{P} \sum_{k=0}^{K-1} \mathbf{W}_t'(p) \mathbf{w}_f(k) \mathbf{w}_f^H(k) \mathbf{W}_t'^H(p). \tag{111}$$

Equation (111) follows by using the properties of the Kronecker product. For simplicity, we rewrite (111) as $\mathbf{D}_0 = I_M \otimes \mathbf{D}_0'$. Note that $\bar{\mathbf{g}}_i^H \mathbf{D}_0 \bar{\mathbf{g}}_i$ in these expressions quantifies the amount of received power at the i^{th} receive antenna.

Now let us denote by $\hat{\bar{\mathbf{g}}}_1, \cdots, \hat{\bar{\mathbf{g}}}_{L_r}$ the coefficient vectors associated with the antennas that have the largest SNRs. We can obtain the joint pdf for these vectors as

$$f_{\hat{\bar{\mathbf{g}}}_1,\cdots,\hat{\bar{\mathbf{g}}}_{L_r}}(\mathbf{x}_1,\cdots,\mathbf{x}_{L_r}) = f_{\bar{\mathbf{g}}_{i_1},\cdots,\bar{\mathbf{g}}_{i_{L_r}}}(\mathbf{x}_1,\cdots,\mathbf{x}_{L_r}|\mathcal{F}) \tag{112}$$

$$= \sum_{l=1}^{L_r} f_{\bar{\mathbf{g}}_{i_1},\cdots,\bar{\mathbf{g}}_{i_{L_r}}}(\mathbf{x}_1,\cdots,\mathbf{x}_L|\mathcal{F},\mathcal{A}_l)P(\mathcal{A}_l) \tag{113}$$

$$= \sum_{l=1}^{L_r} \frac{P(\mathcal{F},\mathcal{A}_l|\bar{\mathbf{g}}_{i_1}=\mathbf{x}_1,\cdots,\bar{\mathbf{g}}_{i_{L_r}}=\mathbf{x}_{L_r})f_{\bar{\mathbf{g}}_{i_1},\cdots,\bar{\mathbf{g}}_{i_{L_r}}}(\mathbf{x}_1,\cdots,\mathbf{x}_{L_r})P(\mathcal{A}_l)}{P(\mathcal{F},\mathcal{A}_l)} \tag{114}$$

$$= \frac{\binom{N}{L_r}}{L_r}\sum_{l=1}^{L_r} P(\bar{\mathbf{g}}_{i_{L_r+1}}^H\mathbf{D}_0\bar{\mathbf{g}}_{i_{L_r+1}} < \mathbf{x}_l^H\mathbf{D}_0\mathbf{x}_l,\cdots,\bar{\mathbf{g}}_{i_N}^H\mathbf{D}_0\bar{\mathbf{g}}_{i_N} < \mathbf{x}_l^H\mathbf{D}_0\mathbf{x}_l)\prod_{j=1}^{L_r} f_{\bar{\mathbf{g}}_{i_j}}(\mathbf{x}_j)I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L) \tag{115}$$

$$= \frac{N!}{(N-L_r)!L_r!L_r}\left(\sum_{l=1}^{L_r}\left[F(\mathbf{x}_l^H\mathbf{D}_0\mathbf{x}_l)\right]^{N-L_r} I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L)\right)\frac{e^{-(\|\mathbf{x}_1\|^2+\cdots+\|\mathbf{x}_{L_r}\|^2)}}{\pi^{MLL_r}}. \tag{116}$$

We note that (113) follows using the total probability theorem, (114) follows using the Bayes' rule, and (115) follows because $P(\mathcal{F},\mathcal{A}_l) = 1/\binom{N}{L_r}$ and $P(\mathcal{A}_l) = 1/L_r$, where $\binom{n}{k} = n!/((n-k)!k!)$. $I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L)$ is the indicator function

$$I_{\mathcal{R}_l}(\mathbf{r}_1,\cdots,\mathbf{r}_L) = \begin{cases} 1 \text{ if } (\mathbf{x}_1,\cdots,\mathbf{x}_{L_r}) \in \mathcal{G}_l \\ 0 \qquad\qquad \text{else} \end{cases}$$

where the region $\mathcal{G}_l$ is defined as

$$\mathcal{G}_l = \left\{\mathbf{x}_1,\cdots,\mathbf{x}_{L_r} : \mathbf{x}_l^H\mathbf{D}_0\mathbf{x}_l < \mathbf{x}_k^H\mathbf{D}_0\mathbf{x}_k, \;\; k=1,\cdots,L_r, k\neq l\right\}.$$

$F(\cdot)$ is the cumulative distribution function (cdf) of $Z = \mathbf{z}^H\mathbf{D}_0\mathbf{z}$ where $\mathbf{z}$ is $ML\times 1$ circularly symmetric complex Gaussian random vector with zero mean and unit variance.

By averaging the conditional pairwise error probability (109) over the pdf given in (116), the pairwise error probability can be obtained as

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \sum_{l=1}^{L_r}\int_{\mathcal{G}_l} e^{-\frac{\rho}{8M}\sum_{i=1}^{L_r}\bar{\mathbf{g}}_i^H\mathbf{D}\bar{\mathbf{g}}_i}\frac{N!}{(N-L_r)!L_r!L_r}F(\bar{\mathbf{g}}_l^H\mathbf{D}_0\bar{\mathbf{g}}_l)^{N-L_r}\frac{1}{\pi^{MLL_r}}e^{-(\|\bar{\mathbf{g}}_1\|^2+\cdots+\|\bar{\mathbf{g}}_{L_r}\|^2)}d\bar{\mathbf{g}}_1\cdots d\bar{\mathbf{g}}_{L_r}$$

$$= \int_{\mathcal{G}_1} e^{-\frac{\rho}{8M}\sum_{i=1}^{L_r}\bar{\mathbf{g}}_i^H\mathbf{D}\bar{\mathbf{g}}_i}\frac{N!}{(N-L_r)!L_r!}F(\bar{\mathbf{g}}_1^H\mathbf{D}_0\bar{\mathbf{g}}_1)^{N-L_r}\frac{1}{\pi^{MLL_r}}e^{-(\|\bar{\mathbf{g}}_1\|^2+\cdots+\|\bar{\mathbf{g}}_{L_r}\|^2)}d\bar{\mathbf{g}}_1\cdots d\bar{\mathbf{g}}_{L_r} \tag{117}$$

where (117) follows by observing that the integration over the region $\mathcal{G}_l$ gives the same result for different l. The evaluation of the integral over $\mathcal{G}_1$ is analytically formidable; however, it is possible to compute it by relaxation of the integral region to overall space, which results in a slightly looser upper bound:

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \int_{\mathcal{C}^{L_r}} e^{-\frac{\rho}{8M}\sum_{i=1}^{L_r}\bar{\mathbf{g}}_i^H\mathbf{D}\bar{\mathbf{g}}_i}\frac{N!}{(N-L_r)!L_r!}F(\bar{\mathbf{g}}_1^H\mathbf{D}_0\bar{\mathbf{g}}_1)^{N-L_r}\frac{1}{\pi^{MLL_r}}e^{-(\|\bar{\mathbf{g}}_1\|^2+\cdots+\|\bar{\mathbf{g}}_{L_r}\|^2)}d\bar{\mathbf{g}}_1\cdots d\bar{\mathbf{g}}_{L_r}. \tag{118}$$

By standard integration tools, (118) can be simplified to

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \left| \frac{\rho}{8M}\mathbf{D} + I_{ML} \right|^{-(L_r-1)} \frac{N!}{(N-L_r)!L_r!\pi^{ML}} \mathcal{I}_{\mathcal{C}} \tag{119}$$

where

$$\mathcal{I}_{\mathcal{C}} = \int_{\mathcal{C}} e^{-\bar{\mathbf{g}}_1^H (\frac{\rho}{8M}\mathbf{D} + I_{ML})\bar{\mathbf{g}}_1} F(\bar{\mathbf{g}}_1^H \mathbf{D}_0 \bar{\mathbf{g}}_1)^{N-L_r} d\bar{\mathbf{g}}_1. \tag{120}$$

In order to evaluate (120), we need to compute $F(\cdot)$. Observing that $Z = \mathbf{z}^H(I_M \otimes \mathbf{D}_0')\mathbf{z}$, and using the SVD $\mathbf{D}_0'^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_L\}$, we have $Z = \mathbf{z}^H(I_M \otimes \mathbf{U})(I_M \otimes \mathbf{\Lambda})(I_M \otimes \mathbf{U}^H)\mathbf{z}$. Letting $\mathbf{z}_i = [z_{iL-L+1}, \cdots, z_{iL}]^T,\ i = 1, \cdots, M$ be $L \times 1$ segments from $\mathbf{z}$, and defining $\boldsymbol{\beta}_i^H = \mathbf{z}_i^H \mathbf{U}$, we have

$$Z = \lambda_1|\beta_1|^2 + \cdots + \lambda_L|\beta_L|^2 + \cdots + \lambda_1|\beta_{ML-L+1}|^2 + \cdots + \lambda_L|\beta_{ML}|^2 \tag{121}$$

where β_j is the j^{th} entry of $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^H, \cdots, \boldsymbol{\beta}_M^H]^H$. By regrouping the summation in (121), we finally obtain

$$Z = \lambda_1(|\beta_1|^2 + |\beta_{1+L}|^2 + \cdots + |\beta_{ML-L+1}|^2) + \cdots + \lambda_L(|\beta_L|^2 + |\beta_{2L}|^2 \cdots + \beta_{ML}|^2) \tag{122}$$

Since β_j are iid circularly symmetric complex Gaussian with zero-mean and unit-variance, $\chi_j = |\beta_j|^2 + |\beta_{j+L}|^2 + \cdots + |\beta_{ML-L+j}|^2$ has a Chi-square distribution with $2M$ degrees of freedom, e.g., $\chi_j \sim \chi^2(2M)$. Thus, the characteristic function of Z follows as

$$\phi_Z(s) = \prod_{j=1}^{L} \frac{1}{(1-\lambda_j s)^M} \tag{123}$$

$$= \prod_{j} \frac{1}{(1-\lambda_j s)^{Ml_j}} \tag{124}$$

$$= \sum_{j} \sum_{i=1}^{Ml_j} \frac{A_{j,i}}{(1-\lambda_j s)^i} \tag{125}$$

where in (124) it is assumed that λ_j is repeated l_j times, and in (125), we use partial fraction expansion and A_{ji} can be computed by

$$A_{j,Ml_j-i} = \frac{1}{(-\lambda_j)^i i!} \frac{\partial^i}{\partial s^i} \left[(1-\lambda_j s)^{Ml_j} \prod_k \frac{1}{(1-\lambda_k s)^{Ml_k}} \right] \Bigg|_{s=1/\lambda_j}. \tag{126}$$

Inverse transforming the characteristic function (125) yields the pdf for Z

$$f_Z(z) = \sum_{j} \sum_{i=1}^{Ml_j} A_{j,i} \frac{1}{(i-1)!\lambda_j^i} z^{i-1} e^{-z/\lambda_j} \tag{127}$$

i	$f_i(\cdot)$
1	$\mathrm{tr}(\tilde{\mathbf{B}})$
2	$\mathrm{tr}(\tilde{\mathbf{B}})^2+\mathrm{tr}(\tilde{\mathbf{B}}^2)$
3	$\mathrm{tr}(\tilde{\mathbf{B}})^3+3\mathrm{tr}(\tilde{\mathbf{B}})\mathrm{tr}(\tilde{\mathbf{B}}^2)+2\mathrm{tr}(\tilde{\mathbf{B}}^3)$
4	$\mathrm{tr}(\tilde{\mathbf{B}})^4+8\mathrm{tr}(\tilde{\mathbf{B}})\mathrm{tr}(\tilde{\mathbf{B}}^3)+3\mathrm{tr}(\tilde{\mathbf{B}}^2)^2+6\mathrm{tr}(\tilde{\mathbf{B}})^2\mathrm{tr}(\tilde{\mathbf{B}}^2)-17\mathrm{tr}(\tilde{\mathbf{B}}^4)$

Table 3: The function $f_i(\cdot)$ for $i=1,2,3,$ and 4.

and finally, the cdf can be obtained as

$$F_Z(a)=1-\sum_j\sum_{i=1}^{Ml_j}A_{j,i}e^{-a/\lambda_j}\sum_{k=0}^{i-1}\frac{a^k}{\lambda_j^k k!}. \tag{128}$$

Substituting (128) into (120), then regrouping the terms and using the binomial expansion for the cdf, and then changing the order of integration and summation, we finally arrive at

$$\begin{aligned}\mathcal{I}_{\mathcal{C}}=&\sum_{l=0}^{N-L_r}\binom{N-L_r}{l}(-1)^l\sum_{j_1}\sum_{i_1}^{Ml_{j_1}}\cdots\sum_{j_l}\sum_{i_l}^{Ml_{j_l}}A_{j_1,i_1}\cdots A_{j_l,i_l}\sum_{k_1}^{i_1-1}\cdots\sum_{k_l}^{i_l-1}\\&\frac{1}{k_1!\cdots k_l!\lambda_{j_1}^{k_1}\cdots\lambda_{j_l}^{k_l}}\int e^{-\bar{\mathbf{g}}_1^H\left(\frac{\rho}{8M}\mathbf{D}+\mathbf{D}_0(\lambda_{j_1}^{-1}+\cdots+\lambda_{j_l}^{-1})+I_{ML}\right)\bar{\mathbf{g}}_1}(\bar{\mathbf{g}}_1^H\mathbf{D}_0\bar{\mathbf{g}}_1)^{(k_1+\cdots+k_l)}d\bar{\mathbf{g}}_1.\end{aligned} \tag{129}$$

To evaluate (129), we need the following lemma:

Lemma: For any $K\times K$ Hermitian matrix $\mathbf{A}$, and positive definite Hermitian matrix $\mathbf{B}$, we have

$$\int_{\mathcal{C}}(\mathbf{z}^H\mathbf{A}\mathbf{z})^i e^{-\mathbf{z}^H\mathbf{B}\mathbf{z}}d\mathbf{z}=\frac{\pi^K}{|\mathbf{B}|}f_i(\mathrm{tr}(\tilde{\mathbf{B}}),\cdots,\mathrm{tr}(\tilde{\mathbf{B}}^i)) \tag{130}$$

where $\tilde{\mathbf{B}}=\mathbf{A}\mathbf{B}^{-1}$, and $f_i(\cdot)$ is a function that can be evaluated for any i. Some f_i are presented in Table 3.

Proof: See Appendix B.1.

Letting $\mathbf{B}=\frac{\rho}{8M}\mathbf{D}+\mathbf{D}_0(\lambda_{j_1}^{-1}+\cdots+\lambda_{j_l}^{-1})+I_{ML}$, and $\tilde{\mathbf{B}}=\mathbf{D}_0\mathbf{B}^{-1}$, and using the lemma, we can rewrite (129) as

$$\begin{aligned}\mathcal{I}_{\mathcal{C}}=&\sum_{l=0}^{N-L_r}\binom{N-L_r}{l}(-1)^l\sum_{j_1}\sum_{i_1}^{Ml_{j_1}}\cdots\sum_{j_l}\sum_{i_l}^{Ml_{j_l}}A_{j_1,i_1}\cdots A_{j_l,i_l}\sum_{k_1}^{i_1-1}\cdots\sum_{k_l}^{i_l-1}\\&\frac{1}{k_1!\cdots k_l!\lambda_{j_1}^{k_1}\cdots\lambda_{j_l}^{k_l}}\frac{\pi^{ML}}{|\mathbf{B}|}f_{k_1+\cdots+k_l}(\mathrm{tr}(\tilde{\mathbf{B}}),\cdots,\mathrm{tr}(\tilde{\mathbf{B}}^{k_1+\cdots+k_l}))\end{aligned} \tag{131}$$

Hence, substituting (131) in (119), we finally arrive at the upper bound on the PEP

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le \left|\frac{\rho}{8M}\mathbf{D}+I_{ML}\right|^{-(L_r-1)} \frac{N!}{(N-L_r)!L_r!}\sum_{l=0}^{N-L_r}\binom{N-L_r}{l}(-1)^l \sum_{j_1}\sum_{i_1}^{Ml_{j_1}}\cdots$$
$$\sum_{j_l}\sum_{i_l}^{Ml_{j_l}}\sum_{k_1}^{i_1-1}\cdots\sum_{k_l}^{i_l-1}\frac{A_{j_1,i_1}\cdots A_{j_l,i_l}}{k_1!\cdots k_l!\lambda_{j_1}^{k_1}\cdots\lambda_{j_l}^{k_l}}\frac{1}{|\mathbf{B}|}f_{k_1+\cdots+k_l}(\mathrm{tr}(\tilde{\mathbf{B}}),\cdots,\mathrm{tr}(\tilde{\mathbf{B}}^{k_1+\cdots+k_l})). \tag{132}$$

The bound in (132) is in closed-from in the sense that the functions f_k and the coefficients $A_{j,i}$s in (126) can be evaluated in closed-form. Further simplification of the bound to obtain an insightful expression seems to be formidable except for some specific cases. Therefore in the next section, we will resort to numerical techniques to illustrate the performance of the bound. We note that this expression reduces to the same equation obtained for the case of antenna selection system over a flat-fading channel without OFDM (*i.e.*, $L = K = 1$, see Chapter 2). We illustrate the above result for special cases in the following section.

4.4 Simplified Bounds

Case 1: We consider the case of single antenna selection, $L_r = 1$, when $M = N = L = 2$. Hence, (132) reduces to

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le 2\left[\frac{1}{\left|\frac{\rho}{8M}\mathbf{D}+I_4\right|} - \frac{A_{1,1}+A_{1,2}(1+\mathrm{tr}(\mathbf{D}_0\mathbf{B}_1^{-1})/\lambda_1)}{\left|\frac{\rho}{8M}\mathbf{D}+\mathbf{D}_0/\lambda_1+I_4\right|} - \frac{A_{2,1}+A_{2,2}(1+\mathrm{tr}(\mathbf{D}_0\mathbf{B}_2^{-1})/\lambda_2)}{\left|\frac{\rho}{8M}\mathbf{D}+\mathbf{D}_0/\lambda_2+I_4\right|}\right] \tag{133}$$

where $\mathbf{B}_j = \frac{\rho}{8M}\mathbf{D}+\mathbf{D}_0/\lambda_j+I_4$, $A_{j,1} = (-1)^j\lambda_j\frac{2\lambda_1\lambda_2}{(\lambda_1-\lambda_2)^3}$, $A_{j,2} = \frac{\lambda_j^2}{(\lambda_1-\lambda_2)^2}$, $j\in\{1,2\}$, and λ_j are eigenvalues of $\mathbf{D}_0$. We resorted to symbolic toolbox of MATLAB© in order to simplify (133). For full-rank matrices $\mathbf{D}$ and $\mathbf{D}_0$, explicitly writing down the determinants and simplifying the expressions in (133) and letting $\rho\to\infty$, we obtain the asymptotic performance as

$$P(\mathbf{S}\to\hat{\mathbf{S}}) \le 2G\left(\frac{\rho}{8M}\right)^{-8} \tag{134}$$

where G is a constant that depends only on the eigenvalues of $\mathbf{D}$ and $\mathbf{D}_0$. The expressions obtained in the intermediate steps are very lengthy and therefore omitted here. From

(134), we observe that the maximum achievable diversity order with antenna selection is the same as that achieved by the full-complexity system, *i.e.*, $MNL = 8$. We note that this expression can be obtained when $\mathbf{D}$ and $\mathbf{D}_0$ are full-rank. Recall that $\mathbf{D}$ and $\mathbf{D}_0$ depends on the channel profiles and hence, full-diversity when antenna selection is used can be obtained if the channel exhibits full-rank $\mathbf{D}$ and $\mathbf{D}_0$. ■

Case 2: Assume $M = L = 2$, and $L_r = N - 1$. The full-complexity system with these parameters achieve a diversity order of $MLN = 4N$. The PEP for antenna selection can be computed by using Eqns. (119), (120), and the results of *Example 1*:

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N \left| \frac{\rho}{8M}\mathbf{D} + I_{ML} \right|^{-(N-2)} \cdot G \left(\frac{\rho}{8M}\right)^{-2ML} \tag{135}$$

$$\leq N \left(\frac{\rho}{8M}\right)^{-ML(N-2)} |\mathbf{D}|^{-(N-2)} \left| I_4 + \frac{8M}{\rho}\mathbf{D}^{-1} \right| \cdot G \left(\frac{\rho}{8M}\right)^{-2ML}. \tag{136}$$

For high SNR, using $|I_n + \epsilon\mathbf{A}| \approx 1 + \epsilon \text{tr}\{\mathbf{A}\}$ (See Appendix B.2) and letting $\rho \to \infty$, we finally arrive at the asymptotic PEP bound

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq N |\mathbf{D}|^{-(N-2)} G \left(\frac{\rho}{8M}\right)^{-4N}. \tag{137}$$

Clearly, the diversity is retained with antenna selection in this case as well. ■

Case 3: For our last example, let $M = L = 2$, $N = 3$, and $L_r = 1$. After some algebra, the PEP in (132) can be expressed as

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{N(N-1)}{2} \left\{ \frac{1}{|\mathbf{B}_0|} + \frac{b_1}{|\mathbf{B}_1|} + \frac{b_2}{|\mathbf{B}_2|} - \frac{2c_1}{|\mathbf{C}_1|} - \frac{2c_2}{|\mathbf{C}_2|} + \frac{2e}{|\mathbf{E}|} \right\} \tag{138}$$

where $\mathbf{B}_0 = \frac{\rho}{8M}\mathbf{D} + I_4$, $\mathbf{B}_j = \frac{\rho}{8M}\mathbf{D} + 2/\lambda_j\mathbf{D}_0 + I_4$, $\mathbf{C}_j = \frac{\rho}{8M}\mathbf{D} + 1/\lambda_j\mathbf{D}_0 + I_4$, $\mathbf{E} = \frac{\rho}{8M}\mathbf{D} + (1/\lambda_1 + 1/\lambda_j)\mathbf{D}_0 + I_4$, $b_j = (A_{j,1} + A_{j,2})^2 + 2(A_{j,1} + A_{j,2})A_{j,2}/\lambda_j \text{tr}(\mathbf{D}_0\mathbf{B}_j^{-1}) + A_{j,2}^2/\lambda_j^2(\text{tr}((\mathbf{D}_0\mathbf{B}_j^{-1})^2) + (\text{tr}(\mathbf{D}_0\mathbf{B}_j^{-1}))^2)$, $c_j = (A_{j,1} + A_{j,2}) + A_{j,2}/\lambda_j \text{tr}(\mathbf{D}_0\mathbf{C}_j^{-1})$, and $e = (A_{1,1} + A_{1,2})(A_{2,1} + A_{2,2}) + ((A_{1,1} + A_{1,2})A_{2,2}/\lambda_2 + (A_{2,1} + A_{2,2})A_{1,2}/\lambda_1)\text{tr}(\mathbf{D}_0\mathbf{E}^{-1}) + A_{1,2}A_{2,2}/\lambda_1\lambda_2(\text{tr}((\mathbf{D}_0\mathbf{E}^{-1})^2) + (\text{tr}(\mathbf{D}_0\mathbf{E}^{-1}))^2)$, for $j = 1, 2$. Further simplification of this expression is not analytically tractable, however as shown in the next section, the slopes of the PEP in (138) and that for the full-complexity system are the same indicating similar diversity advantages at high SNR. ■

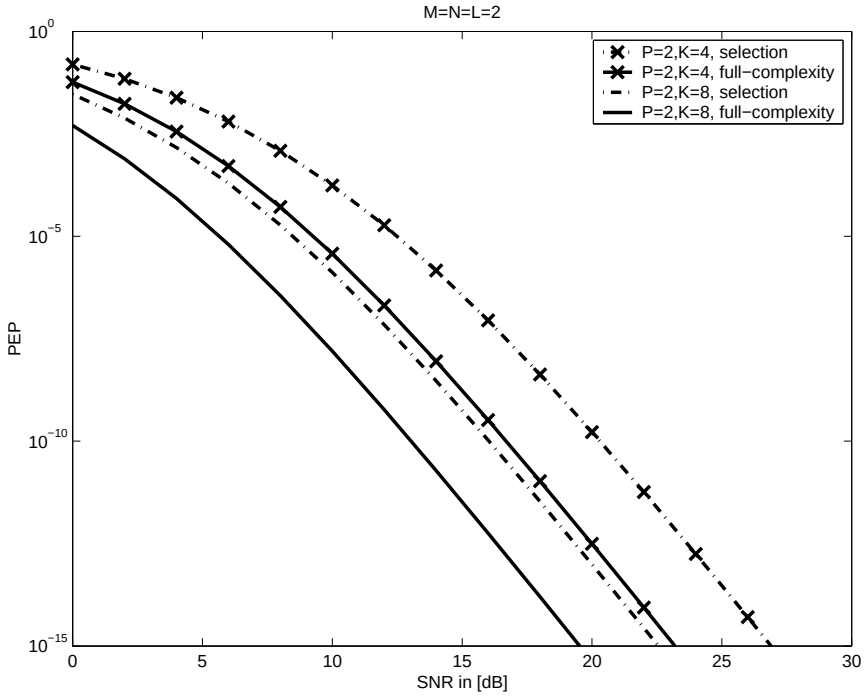


Figure 21: PEP for full-complexity MIMO-OFDM system and MIMO-OFDM with single antenna selection. $K = 4, 8$, $P = 2$, $L = 2$, $M = 2$, $N = 2$.

4.5 Numerical Examples

In this section, we provide several numerical examples to illustrate the performance of the antenna selection for MIMO-OFDM systems. We numerically evaluate the PEP bounds derived in the previous section and compare the performance of system with and without antenna selection.

Figure 21 depicts the PEP curves for both full-complexity MIMO-OFDM system and the one using antenna selection. We consider a frequency-selective channel with two taps, *i.e.*, $L_f = 2$. $K = 4$, or 8 carriers are used and each STC codeword spans $P = 2$ time slots. The PEP results in this plot are for $M = 2$ transmit and $N = 2$ receive antennas. The PEP for system using antenna selection is obtained numerically using (133). The PEP results show that the selection introduces a coding gain loss of about 4 dB when $K = 4$ and 3 dB when $K = 8$ with respect to the full-complexity system, while the diversity gain of 8 is retained. In this example, the space-time codeword pairs are selected such that $\text{rank}(\mathbf{D}_0) = \text{rank}(\mathbf{D}) = 4 = ML$, *i.e.*, full-rank STC-OFDM.

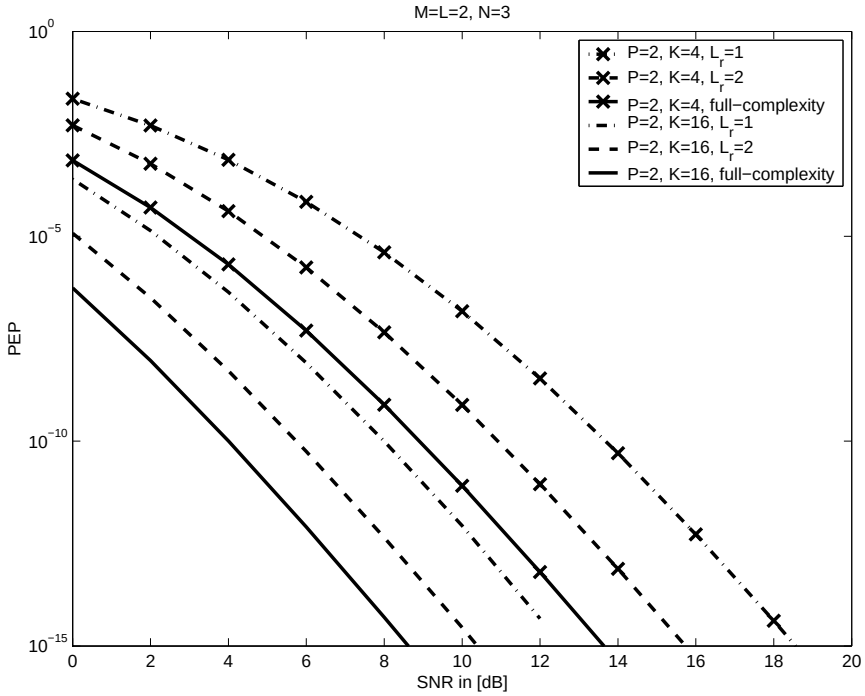


Figure 22: PEP for full-complexity MIMO-OFDM system and MIMO-OFDM with single antenna selection. $K = 4, 16$, $P = 2$, $L = 2$, $M = 2$, $N = 3$, $L_r = 1$ or 2.

Figure 22 illustrates the PEP for a different case. The parameters for this example are $K = 4$, or 16, $P = 2$ and $M = L = 2$ and $N = 3$. We use (135) and (138) to evaluate the PEP for $L_r = 1$ and $L_r = 2$, respectively. The slopes of the curves indicate that the diversity order remains the same as the full-complexity OFDM system when we perform antenna selection. However, the loss in the coding gain is about 5 dB when $L_r = 1$ and 2 dB when $L_r = 2$. As expected, we also observe that the performance of both the full-complexity system and the one with antenna selection improves as the number of the tones in the OFDM system increases.

We next consider a MIMO-OFDM system with $M = 3$ transmit antennas. The PEP can be computed by (132). Further simplification of this expression is formidable, however, the slopes of the PEP plots in log-log scale (See Figure 23) clearly indicates that the diversity gain of the system using antenna selection and the one with full-complexity is the same. These numerical examples hint that the antenna selection is a viable technique to reduce the complexity while retaining the diversity advantage of the MIMO-OFDM system.

We finally present in Figure 24 the word error rate (WER) for the MIMO-OFDM system

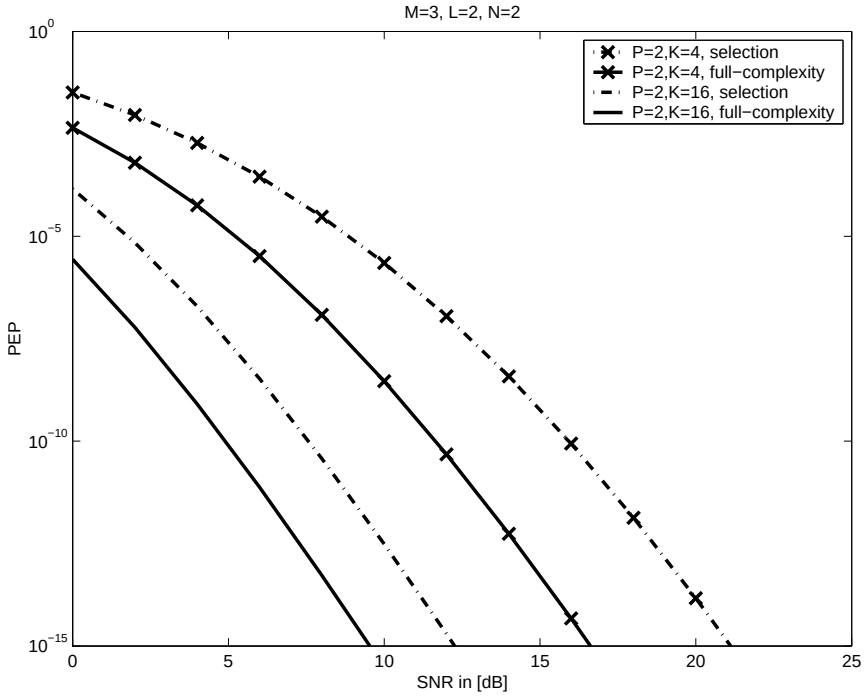


Figure 23: PEP for full-complexity MIMO-OFDM system and MIMO-OFDM with single antenna selection. $K = 4, 16$, $P = 2$, $L = 2$, $M = 3$, $N = 2$.

using space-time trellis codes. The error rates are evaluated via simulation of a MIMO-OFDM system described in [1]. We assume that $K = 6$, $P = 2$, and $L = 2$. The channel state estimation might be unreliable with $P = 2$, however we consider the ideal case where the CSI is exactly known at the receiver; hence the results shown below will indicate the ultimate performance of this MIMO-OFDM system with antenna selection. The codes are selected from 2-space-time codes employing 8, 16, 32, and 64−state trellises [27, 163].In order to achieve full-diversity MIMO-OFDM, one needs to select the codebook such that the effective distance, D_{eff}, of the code is larger than or equal to $ML = 4$. For the code using 8−state trellis, the effective length is $D_{eff} = 2 < ML$, and for 16−state and 32−state trellis codes, $D_{eff} = 3 < ML$, and for 64−state trellis code, $D_{eff} = 4 = ML$. From the slopes of WER curves for the full-complexity system and the system using antenna selection, the following observations are in order: (i) the best diversity order (for both the full-complexity system and the one using antenna selection) can be achieved by the 64−state code, which has $D_{eff} = 4 = ML$, *i.e.*, we have full-rank $\mathbf{D}$, (ii) for the space-time code using 16− and 32−state trellises, the diversity order of the full-complexity system and the

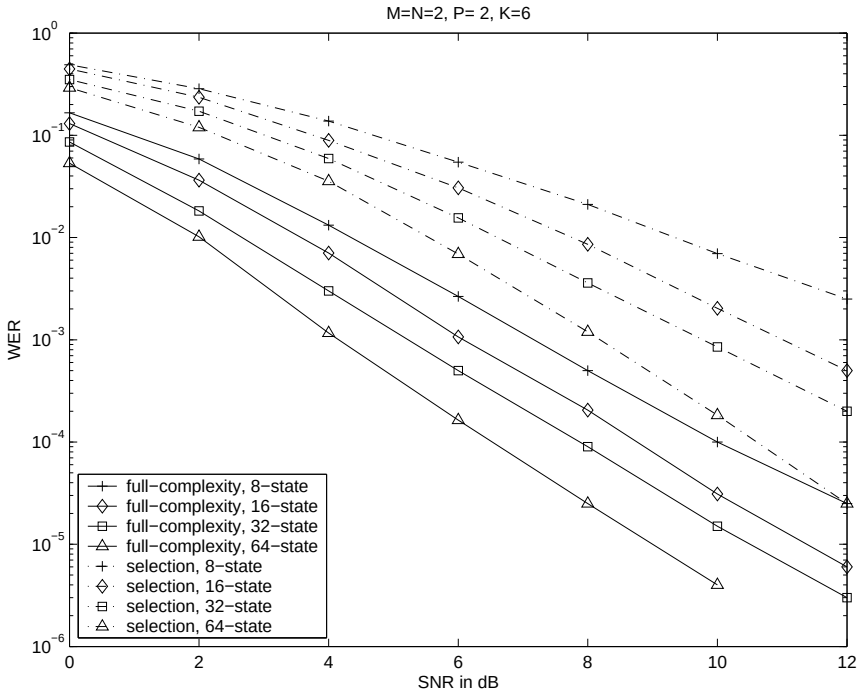


Figure 24: WER for full-complexity MIMO-OFDM system and MIMO-OFDM with antenna selection. $K = 6,\ P = 2,\ M = N = 2,\ L = 2$. 2 b/s/Hz space-time codes with 8, 16, 32 and 64 state trellises are used.

one using antenna selection is slightly less than that of the systems using 64−state trellis code. Also, the slopes of the WER curves for the system using antenna selection and for the full-complexity system are the same although $D_{eff} = 3 < ML$, (iii) and finally, the diversity order of the system using 8−state trellis code is considerably less than that of the others. These observations hint that the diversity order of the system degrade if the underlying space-time code has an effective length that is less than ML. Hence, one needs to use the space-time codes having large effective distances to achieve the best asymptotic performance.

4.6 Chapter Summary

We have studied the MIMO-OFDM system with antenna selection. We derived Chernoff bounds for such systems and used these bounds to quantify the coding gain and the diversity gain of the system using antenna selection. The general expression does not provide useful insight into the asymptotic performance. However, for a system with two transmit antennas,

we show explicitly that with single antenna selection out of $N = 2$ receive antennas, and with the selection of $N - 1$ antennas out of N antennas, one can attain the spatial diversity order of the full-complexity system. For some other cases, numerical results demonstrate that diversity order of the MIMO-OFDM system with antenna selection is retained as well. Simulation results for an actual MIMO-OFDM system using antenna selection indicate that if the effective distance of the underlying space-time code is greater than or equal to ML, the spatial diversity order of the full-complexity system can be achieved with antenna selection, however, if the effective length is less than ML, then there may be some loss in the diversity order with antenna selection.

CHAPTER V

TURBO CODED MULTIPLE DESCRIPTION CODING FOR MIMO ANTENNAS

The previous chapters dealt with the performance of a reduced-complexity MIMO antenna system. In this chapter, we study a different problem: joint-source-channel coding for MIMO antenna systems. In particular, our focus will be on the use of Multiple Description Codes (MDCs) over MIMO antenna systems and we will propose a turbo-coded MDC scheme for such systems. The organization of the chapter is organized as follows: In the next section, we give a brief motivation to use MDC for MIMO antenna systems. In Section 5.2, we describe the transmission system involving MDC, multiple antenna link, and the turbo coded modulation for systems employing antenna diversity. In Section 5.3, we propose a suitable joint source-channel decoding algorithm based on turbo decoding concepts. We present numerical examples in Section 5.4 and summarize our conclusions in Section 5.5.

5.1 Introduction

The traditional design of a communication system makes use of the Shannon's source-channel separation theorem [142]. That is, source and channel coders are designed separately. However, for (practical) multimedia communications over wireless channels, the assumptions held in the separation theorem (*e.g.*, infinite delay and complexity) are not directly applicable, which implies that the joint design of channel and source coding may achieve a better performance. Therefore, joint source-channel codecs have taken considerable attention and various methods have already been developed [45, 46, 67, 76, 136].

A particular source coding method, known as multiple description coding (MDC), can be viewed as a joint source – channel coding technique. MDC generates multiple bitstreams, also called descriptions, of a source so that various quality levels of reconstruction can be

obtained from any subset of the descriptions. The descriptions are transmitted over independent channels with the hope that upon the reception of all or some of the descriptions, a superior or an acceptable quality reconstruction is possible. This can be accomplished by introducing a certain amount of correlation between the individual descriptions.

The MDC has been extensively studied for "on-off" channels, such as the Internet, assuming that there exists multiple independent channels that either provide error-free transmission or experience total failure. However, this assumption is not suited to wireless channels. Such impairments as fading and multipath propagation cannot simply be abstracted as "on-off" type channel. With this motivation, Yang and Vaishampayan [183] show that the performance of an MDSQ-based system dominates that of a channel-code based system for delay-constrained slow-fading channels. Other attempts on the use of multiple description coding for wireless systems are made in [7, 96]. In [7], the authors propose the transmission of multiple descriptions over a wireless link using multiple transmit and receive antennas. Another recent approach makes use of the dependencies in descriptions [146]; in this work, Sirinavasan considers the transmission of multiple descriptions over a noisy channel, and proposes an iterative decoding algorithm based on turbo coding concepts, which exploits the correlation between the descriptions.

Noting that MDC requires multiple independent paths, we naturally consider multiple antenna systems where we can obtain multiple independently fading channels. Deploying multiple antennas has long been known to be an effective technique to provide spatial diversity for combating the destructive effects of multipath fading. It has also recently been shown that using multiple antennas can enhance the capacity of the wireless channels dramatically [164]. To exploit this capacity, various coded modulation techniques, which are known as space-time codes, have been proposed [151,163]. Such coding schemes promise practical high-rate communications over wireless channels. Our main objective is to assess the usability and performance of MDC with these methods, since MDC can be effectively used for transmitting multimedia information such as speech, image and video.

With this motivation, we propose a joint source-channel coding scheme that combines multiple description coding with the space-time turbo coded modulation (TCM). Each

description obtained with MDC is independently turbo coded and transmitted via multiple antennas over the wireless channel. At the decoder, we employ an iterative joint source-channel decoding (JSCD) technique in which we exploit the correlation (induced by the multiple description code) among the descriptions by exchanging the information between the source and channel codes. We compare the performance (in terms of bit error rates and mean square error distortions) obtained by the joint source-channel decoding and the decoding where the correlation between the descriptions are not taken into account. We also compare the performance of the multiple description coding system with the conventional single description system with the same data rate, and we observe that the joint source-channel decoder outperforms the other decoder. Furthermore, the simulations show that the system using the multiple description coding performs significantly better than the one that uses only a single description for moderate levels of signal-to-noise ratios. At high signal to noise ratios, the latter system performs better in terms of the mean square error distortion.

5.2 The Transmission System

Assume that we want to transmit the output of a source (such as speech data or image data) over a wireless link using a joint-source channel encoder using MDC. The block diagram of the encoder, the components of which will be explained in more detail in the following subsections, is given in Figure 25.

5.2.1 Multiple Description Coding Schemes: MDSQ and MDTC

In our turbo coded system, we may use a variety of multiple description encoding schemes [69, 170] to produce two descriptions of the source samples. In this work, we will consider the use of multiple description scalar quantization (MDSQ) and Multiple Description Coding using pairwise correlating transforms (MDTC). These schemes are illustrated in some details in Chapter 1.

It is noteworthy to emphasize the distinct ways that the correlation is created between the descriptions for MDSQ and MDTC. In the case of MDSQ, the correlation between the descriptions are induced by the index assignment and number of diagonals that are

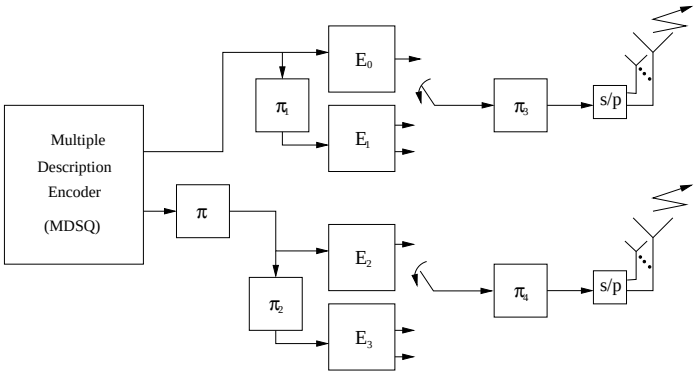


Figure 25: Block diagram for the encoder.

occupied. With MDTC, one can obtain a *well-defined* correlation between the descriptions by using correlation matrices. It is also easier to provide a smooth tradeoff between the correlation introduced by MDC and the performance improvement by JSCD that uses this correlation. MDTC is a technique that enables one to control the amount of correlation as desired through linear transformation.

5.2.2 Space-Time Turbo Coded Modulation with MDC

As shown in Figure 25, the output of the MDC (MDSQ or MDTC) is channel coded using a turbo code and the encoded bits are then interleaved, passed through a serial to parallel converter, and mapped to a signal constellation. The signal sequences corresponding to different descriptions are transmitted from different antenna elements.

The interleavers prior to signal mapping are used to decorrelate the log-likelihoods of the consecutive bits being transmitted and hence, to distribute the burst errors due to a deeply faded block over the entire frame. We note that there are a total of five different interleavers in the encoder: two for scrambling the encoded bits from each description (π_3 and π_4), two for the turbo encoders (π_1 and π_2) and one acting over one of the descriptions from MDC (π).

5.2.3 Multiple Antenna Link Model

We assume that there are m antennas at the transmitter and n antennas at the receiver. Each propagation coefficient between antenna pairs is supposed to be a zero mean complex Gaussian random variable, *i.e.*, Rayleigh fading. Also, we assume that the channel is block

fading for which propagation coefficients remain constant during the coherence time of the channel, *i.e.*, during L symbol periods, and they are independent from one block to the next. At time t, the received signal at the j^{th} receive antenna, $x_j(t)$, can be written as

$$x_j(t) = \sqrt{\frac{\rho}{M}} \sum_{i=1}^{m} h_{ij} c_i(t) + w_j(t), t = 1, ..., L, n = 1, ..., r$$

where the path gains h_{ij} and the noise samples $w_j(t)$ are independent, zero-mean and unit variance circularly symmetric complex Gaussian random variables. The transmitted signals $c_i(t)$ can be chosen from an arbitrary signal constellation. The average energy of transmitted signal at time t is normalized to unity over m antennas so that ρ is the expected signal to noise ratio at each receiver antenna.

5.3 Joint Source-Channel Decoding of the MDC with TCM

In this section, we will present a suboptimal joint source-channel decoding algorithm for the above system adopted from [53]. The correlation introduced by the multiple description encoder will be exploited via iterative decoding process that passes extrinsic information between the channel decoder and source decoder.

The block diagram of the receiver is given in Figure 26. There are two turbo decoding blocks associated with the two turbo encoders. Decoding of any of the turbo coded modulation schemes is similar to the decoding algorithm given in [41,60,151]. We first compute the log-likelihoods of the transmitted bits, and then use them in the standard iterative turbo decoding as if they are the log-likelihoods of the observation from a BPSK modulation over an additive white Gaussian noise channel. The decoding of the individual turbo codes is performed as usual. That is, the two maximum a posteriori (MAP) decoders exchange the extrinsic information between each other at each step of the iterations. However, additional extrinsic information is also exchanged between the turbo decoders, which will be made clear shortly.

Assume that the size of the constellation is 2^{R_c}, and denote the set of constellation points with c_i, $i = 1, \cdots, 2^{R_c}$. Then each constellation point corresponds to R_c bits and the received signal vector at time t, x_{tj}, $j = 1, 2, \cdots, n$, corresponds to mR_c bits, $\boldsymbol{e} = (e_1, \cdots, e_m, e_{m+1}, \cdots, e_{mR_c})$. The bits $(e_1, \cdots, e_{mR_c/2})$ are the encoded bits of one

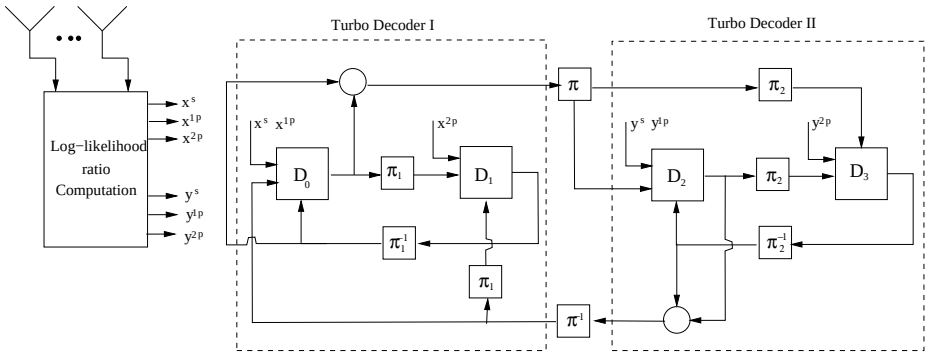


Figure 26: Block diagram for the iterative joint source-channel decoding of combined multiple description coding and space-time turbo coded modulation scheme.

description and the bits $(e_{mR_c/2+1}, \cdots, e_{mR_c})$ are encoded bits of the other description. Based on the bit block $(e_{(i-1)R_c+1}, \cdots, e_{iR_c})$, the signal c_i is selected and transmitted from i^{th} transmit antenna . Using the similar derivation in [151], we obtain the log-likelihood of the l^{th} bit in $\boldsymbol{e}$

$$\Lambda(e_l) = \log \frac{Pr(e_l = 1|X)}{Pr(e_l = 1|X)}$$

The log-likelihood values $\Lambda(e_1), \cdots, \Lambda(e_{\frac{mR_c}{2}})$ are used by the turbo decoder I and $\Lambda(e_{\frac{mR_c}{2}+1})$, $\cdots, \Lambda(e_{mR_c})$ are used by the turbo decoder II. Both turbo decoders are run in alteration and each of them generates the extrinsic information about the input bits, $\{b_l^i\}$, $l = 1, \cdots, R$, $i \in \{1, 2\}$, corresponding to associated description.

For each of the turbo decoders in Figure 26, if the correlation between descriptions is not considered, the iterative decoding of each turbo code is performed using standard forward and backward recursion relations described in [127], and the extrinsic information for both the "present" constituent decoder, D_p, and the "former" constituent decoder, D_f, denoted by $P_e(d_k = z|D_p)$ and $P_e(d_k = z|D_f)$, respectively, are computed. The bit d_k is the information bit for the transition from step $(k-1)$ to step k and its *a priori* probability, which is exchanged between the constituent decoders, is the key to the success of the turbo decoding. At each iteration, the extrinsic information $P_e(d_k = z|D_f)$ from one of the constituent decoders is used as the *a priori* probability of d_k in the current MAP decoder.

Assuming that the extrinsic probabilities from the constituent convolutional decoders are independent, we can write the extrinsic probabilities about the bits in the first description

as [54]

$$P_e\{b_l^1 = z|TD_1\} = KP_e\{b_l^1 = z|D_0\}P_e\{b_l^1 = z|D_1\}$$

where $z \in \{0, 1\}$ and TD_1, D_0, D_1 implies the extrinsic probabilities are from turbo decoder I, MAP decoders 0 and 1, respectively. K is a normalization factor. Similarly, for the second turbo decoder,

$$P_e\{b_l^2 = z|TD_2\} = KP_e\{b_2^1 = z|D_2\}P_e\{b_l^2 = z|D_3\}$$

When the descriptions are considered independently, both turbo decoders are run as described above and the extrinsic probabilities due to the constituent decoders are the only information that is exchanged between the MAP decoders D_0/D_1 or D_2/D_3.

In order to exploit the correlation between the descriptions of the MDC, we need to exchange the additional extrinsic information obtained using the source *a priori* probabilities. We now focus on the transfer of information between the turbo decoders I and II. Since the correlation between descriptions is introduced in the index level, we should perform this transfer in the index level. The extrinsic probabilities of the MDC indices v^i, $i = 1, 2$, for the first or second description can be computed from the extrinsic probabilities of the bits constituting that description as:

$$P_e(v^i = I) = \prod_{l=1}^{R} P_e(b_l^i|TD_i), \quad i \in \{1, 2\}, \;\; I = 1, 2, \cdots, M_1$$

where b_l^i is the l^{th} bit in the binary representation of the index for the i^{th} description and we assume that the bits $b_1^i, \cdots, b_R^i$ are independent. Given the probabilities for one description, the extrinsic probabilities of the other description can be calculated using the *a priori* statistics of the source as

$$\begin{aligned} P_e(v^i = J|TD_{i'}) = \sum_{k=1}^{M_1} P(v^i = J|v^{i'} = k)P_e(v^{i'} = k|TD_{i'}), \\ i, i' \in \{1, 2\}, \;\; J = 1, 2, \cdots, M_1 \end{aligned} \tag{139}$$

where the conditional probabilities of the indices $P(v^i = I|v^{i'} = J)$ can easily be computed using the source p.d.f., threshold vector and the index assignment. We can now compute

the extrinsic probabilities for the information bits constituting this description (index) as

$$P_e(b_r^i = z|TD_p) = \sum_{k:b_r^i(k)=z} P_e(v^i = k|TD_{i'}), \quad z \in \{0,1\}, \\ k = 1, \cdots, M_1, \quad i \in \{1,2\}, \quad r = 1, \cdots, R \tag{140}$$

where $P_e(\cdot|TD_p)$ is extrinsic probability obtained using the extrinsic information from the former turbo decoder $TD_{i'}$ and the *a priori* source statistics, and $b_r^i(k)$ is the r^{th} bit of the index k in the i^{th} description. This probability (i.e., information from the former turbo decoder) is used as an additional extrinsic information in the current turbo decoder. To achieve this, reverse and backward recursion relations of the MAP algorithm for the constituent convolutional decoders are modified as shown in [54] such that the *a priori* information for d_k is obtained using both the extrinsic information from the constituent MAP decoder and the additional information from the other turbo decoder.

We are now able to describe how the individual components of the joint source-channel decoder are connected together. As shown in Figure 26, the component decoders D_0 or D_1 in turbo decoder I (D_2 or D_3 in turbo decoder II) takes the log-likelihood values of the corresponding systematic and parity bits computed by the log-likelihood computation block, the extrinsic information from the other decoder D_1 or D_0 (D_3 or D_2) within the same turbo decoder, and also, takes the additional extrinsic information from turbo decoder II (turbo decoder I) as computed using (140) , and then uses the MAP decoding algorithm with the modified decoding to compute the new extrinsic information about the bits in the corresponding description. At each iteration, the related extrinsic information is passed between the individual MAP decoders, and between the turbo decoders I and II. The iterations proceed as the extrinsic information is exchanged between the decoders until a desired performance is achieved at which point a final decision is made based on the final log-likelihood ratio of information bits. Note that in all these operations, all the likelihood ratio sequences and received sequences are suitably reordered and delayed.

After stopping the iterations, the final decision for the transmitted indices is made using

$$(i^*, j^*) = \arg \max_{(i,j)\in\mathcal{J}} P(i,j) = \arg \max_{(i,j)\in\mathcal{J}} P(i|j)P(j) \\ = \arg \max_{(i,j)\in\mathcal{J}} P(j|i)P(i) \tag{141}$$

where $P(i)$ and $P(j)$ are the probabilities of the first and second indices of the MDSQ, respectively, which are computed using the final log-likelihoods of the constituting bit sequences. Hence, we select the indices that maximize the joint probability of the possible index pairs in the MDSQ scheme and based on the decoded index pairs, the transmitted source sample is reconstructed at the receiver.

5.4 Examples

In this section, we present the performance of the proposed scheme using several simulations. The information bits are generated either using the MDSQ or MDTC with natural binary code assignment where we use the binary representations of the indices. The turbo codes in the system consists of two recursive systematic convolutional codes described by the feedforward and feedback generating polynomials (g_n, g_d). We assume that the interleaver is pseudorandom, and choose g_n and g_d to be 5_{octal} and 7_{octal}, respectively, for both turbo codes. We obtain a rate 1/2 code by puncturing the parity bits periodically. The encoded bits are multiplexed and then interleaved using pseudo random interleavers. In our examples, we use QPSK modulation at each transmit antenna, and we employ two transmit antennas, each of which is used to transmit the sequences due to one of the descriptions. We present the results for a receiver with a single antenna element. We use the iterative decoding algorithm with 8 iterations.

In order to show the effectiveness of the joint source-channel decoding algorithm for the proposed scheme, we first present the iterative decoding results obtained by using two correlated binary sources, u_k^i, $i \in 1, 2$ and $k = 1, 2, \cdots$, where the correlation is created in the following way:

- Generate the i.i.d. bit sequence u_k^1 such that $P(u_k^1 = 0) = P(u_k^1 = 1) = 1/2$.
- Construct the sequence u_k^2 using $u_k^2 = u_k^1 \oplus e_k$ where $P(e_k = 0) = 1 - p$ and $P(e_k = 1) = p$, and $\oplus$ is modulo 2 addition.

A measure of correlation between u_1 and u_2 can be defined in terms of entropy as [34]

$$\rho = 1 - \frac{H(u_2|u_1)}{H(u_1)} = 1 - plog_2(p) - (1-p)log_2(1-p)$$

From this expression, it is clear that, if $p < 1/2$ decreases, the correlation between the two sequence increases.

For this case, the extrinsic information exchanged between the turbo decoders is simply

$$\Lambda_e(u_k^p|TD_p) = \log \frac{\Big\{(1-p)P_e(u_k^f = 1|TD_f) + pP_e(u_k^f = 0|TD_f)\Big\}}{\Big\{(1-p)P_e(u_k^f = 0|TD_f) + pP_e(u_k^f = 1|TD_f)\Big\}}$$

where the subscripts and superscripts p and f are used to denote the "former" or "present" iteration steps at which the extrinsic probabilities $P_e(u_k|TD)$ from the constituent turbo decoders are obtained.

In Figure 27, we present the bit error rate results for the input scenario described above. The size of the interleavers in the turbo codes are 1300. The path is assumed to be constant for a period of 130 symbols. We observe that the joint source-channel decoding algorithm taking the source correlation into account outperforms the standard iterative decoding where we do not exchange between the turbo decoders. At a BER of 10^{-4}, the joint source-channel decoding provides about 2.7 dB gain when $p = 0.1$ and about 0.5 dB gain when $p = 0.2$ over the standard decoding. These results imply that when the amount of correlation between the source sequences is large, then using the extrinsic information due to the correlation improves the performance dramatically. If the correlation between the sequences is reduced, the performance improvement is also reduced, however, we can still expect some gain in exploiting this correlation.

Next, we present the performance results for the combined multiple description coding and space-time turbo coded modulation scheme. First, we will consider the use of MDSQ.

5.4.1 Case 1: MDSQ

In the following examples, we will introduce the correlation between the transmitted information sequences using various MDSQ schemes. We will also compare the results of iterative decoding of multiple descriptions to the conventional single description schemes where we employ a single scalar quantization (with a rate equal to the total rate of the MDSQ scheme) to generate our information sequence and transmit them using the same space-time turbo coded modulation scheme. We note that the source sample delay for the

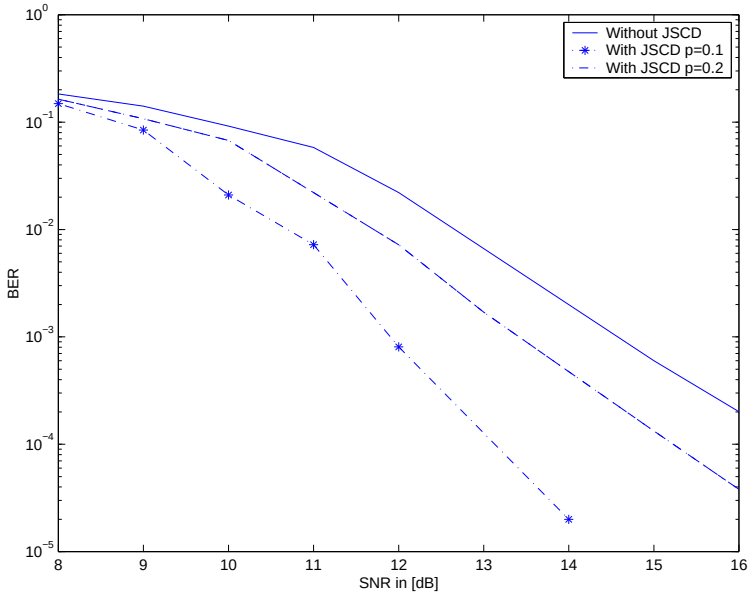


Figure 27: Bit error rate vs. SNR for the joint source-channel turbo decoding of correlated sources. Turbo code block size is 1300, m=2 transmit and n=1 receive antennas.

case of the single description based system equals half of the sample delay for the case of MDC based system since we want to keep the interleaver sizes (and hence, complexities of the turbo codes) in both cases almost similar. We also include performance comparisons with a single description scheme using TCM with the same delay but employing single antenna. For this case, the signals are selected from a larger constellation to provide the same spectral efficiency.

In Figure 28, we present the bit error rate comparison between the joint source-channel decoding algorithm with the standard turbo decoding. In this example, we use MDSQ with $M_1 = M_2 = 8$ (i.e., $R = 3$ bpss/description) and using the index assignment scheme given in Table 1.b (Chapter 1). We also plot the bit error rate for a single description quantization scheme using 64-level uniform quantization (i.e., the rate is $2R$ bpss). The channel is a Rayleigh block fading channel where the path gains are constant for a period of 150 transmissions. The interleaver size in turbo codes is 1500 bits. We employ one receive and two transmit antennas. We see that applying the joint source channel-decoding algorithm improves the performance by about 0.5 dB at a bit error rate of 10^{-4} compared to the case of no joint decoding. We further note that the system using a single description

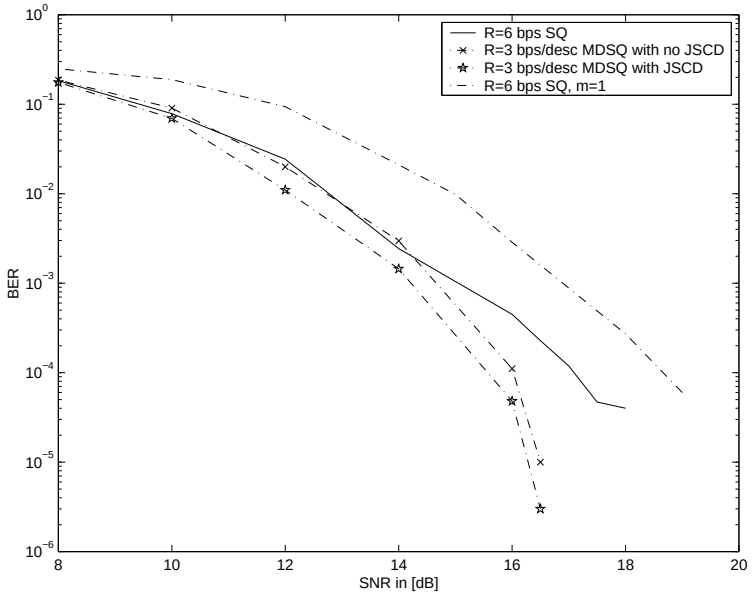


Figure 28: BER vs. SNR for the MDSQ with index assignment shown in Table 1. Turbo code block size is 1500, m=2 transmit and m=1 receive antennas. The block size for $m = 1$ is 3000.

quantizer at a rate of 6 bpss and the same space-time turbo coded modulation scheme is inferior to the system employing MDSQ. We obtain about 1.5 dB performance improvement at a bit error rate of 10^{-4} by using MDSQ system and the joint source-channel decoding. In this figure, we also plot the bit error rate of a single transmit antenna system ($m = 1$) using TCM with an interleaver size of 3000 bits, so that the delays are the same with the MDC based scheme. The transmitted signals c_i are selected from 16−PSK constellation. We observe that the system using the MDC scheme with 2 transmit antenna outperforms the single description scheme with a single transmit antenna, e.g., by about 3 dB at a bit error rate of 10^{-4}.

The performance in terms of square error distortion for the MDC system with joint source-channel decoding and the single description systems (with 1 or 2 transmit antennas) is presented in Figure 29. For this case, we present the mean square error distortions for two different MDSQ index assignments, which are provided in Table 1 (Chapter 1). We apply the joint source-channel decoding algorithm. We observe that the MSE distortion obtained using the MDSQ scheme is less than the systems with a single description finer quantizer for a wide range of SNR levels, i.e., from 8 dB to about 15 dB. For larger SNRs,

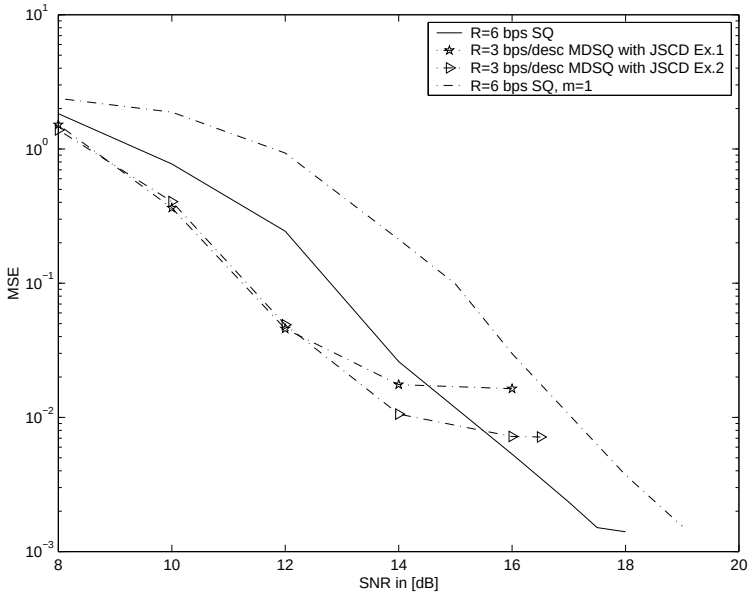


Figure 29: MSE vs. SNR for the MDSQ with index assignment shown in Table 1. Turbo code block size is 1500, m=2 transmit and n=1 receive antennas. The block size for $m = 1$ antenna is 3000.

we observe this is the other way around. We conclude that most of the errors during the transmission are corrected by the iterative decoding at the high SNR values and the MSE distortion at these SNRs is mainly due to the quantization error, which is, as expected, less for the single description system employing a finer quantizer.

In Figures 30 and 31, we present the similar performance comparison for MSE distortions for the systems with $R = 4$ and $R = 5$ bps, respectively. For both cases, the MDC based system with/wothout joint source-channel decoding outperforms the single description based systems for SNR values less than about 17 dB. For larger SNR values, the single description quantization based system achieves smaller values of MSE distortion. However, we observe that increasing the rate from $R = 3$ bpss to $R = 5$ bpss reduces the performance gap (in terms of MSE distortion) between the MDC based system and single description based system.

In Figure 32, we present the comparison between a turbo coded MDC scheme and an uncoded MDC scheme with multiple antennas. For the uncoded scheme, the descriptions are mapped to constellations without any explicit channel coding. In order to make a fair comparison in terms of spectral efficiency, we use a higher rate MDC for the uncoded case.

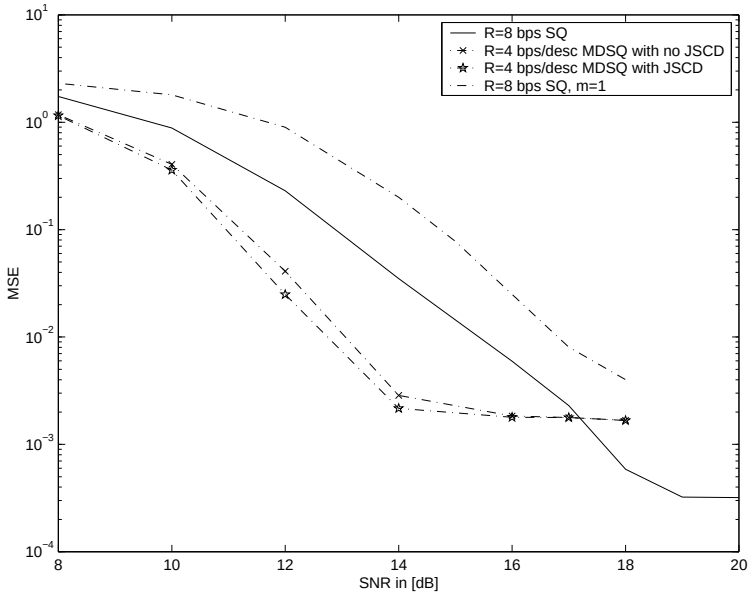


Figure 30: MSE distortion vs. SNR or the MDSQ with $M_1 = M_2 = 16$. Turbo code block size is 1500, n=1 receive and m=2 transmit antennas. The block size for $m = 1$ antenna is 3000.

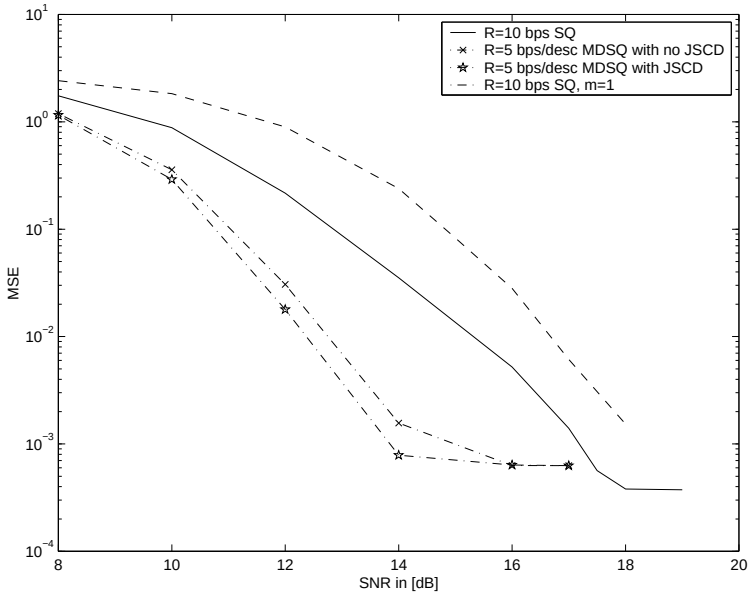


Figure 31: MSE distortion vs. SNR for the MDSQ with $M_1 = M_2 = 32$. Turbo code block size is 1500, m=2 transmit and n=1 receive antennas. The block size for $m = 1$ antenna is 3000.

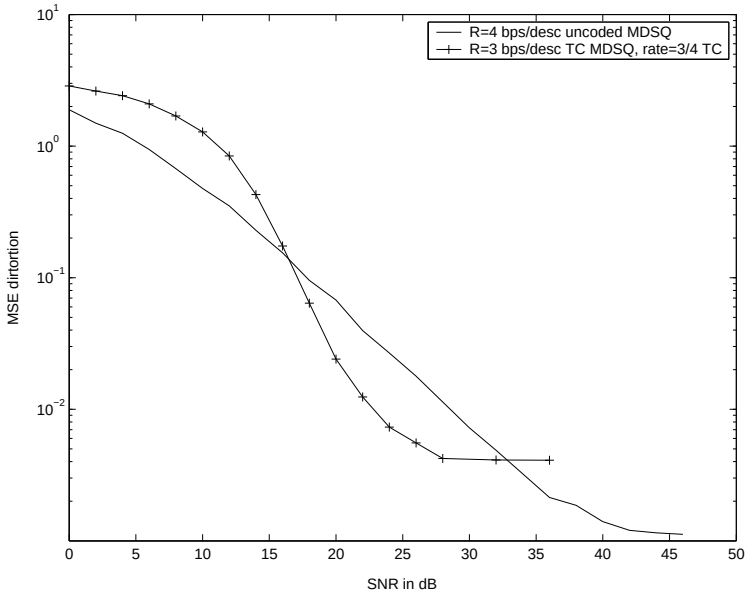


Figure 32: MSE vs. SNR for the turbo coded MDSQ and uncoded MDSQ. Turbo code block size is 1500 and $r = 3/4$, $m = 2$ transmit and $n = 1$ receive antennas. For fair comparison, uncoded MDSQ uses $R = 4$ bps/desc while TC MDSQ uses $R = 3$ bps/desc.

We also use a symbol interleaver for the uncoded MDC such that both systems will have equal interleaving delay. The coded case uses $R = 3$ bps/desc and a rate 3/4 turbo code while the uncoded case uses $R = 4$ bps/desc. In this example, we see that the the turbo coded MDC outperforms the uncoded system for SNR values between $16 - 34$ dB. For example, at a MSE distortion of 10^{-2}, the performance improvement is about 6 dB. For larger SNR values, the uncoded MDSQ converges to a smaller MSE distortion since the source code rate is higher than that for the coded case.

5.4.2 Case 2: MDTC

In this part, we will present the turbo coded MDC results when MDTC is used as the multiple description encoder. In the following graphs, we show the mean square error distortion. In our simulations, we use transforms that give balanced rates, i.e., T is such that the descriptions have equal rates. A balanced-rate transform has the form, for any nonzero α

$$T_\alpha = \begin{pmatrix} \alpha & (2\alpha)^{-1} \\ -\alpha & (2\alpha)^{-1} \end{pmatrix}$$

In Figure 33 and 34 are shown the MSE distortion plots for $\alpha = 1.2$ and $\alpha = 1.79$ for which the redundancy, ρ, is 0.78 and 1.34 bps, respectively. The basic rate is approximately 4 bps for each component of the input pairs and the pairs have independent components with $\sigma_1^2 = 1$ and $\sigma_2^2 = 0.25$. Fixed length codewords are used to represent the MDTC outputs. The distortion for MDTC plots are evaluated using an equally weighted averaging operation over the individual distortions for each component. An initial observation from the plots is that the MSE distortion graphs have similar patterns as in the case of MDSQ based system, that is, for low and high SNR region, the MDTC based systems with JSCD and without joint decoding shows the same performance characteristics. Within the intermediate values of SNR, i.e. from 10 to 17 dBs, JSCD improves the performance by about 0.7 dB at both values of α: although we introduce more redundancy, or correlation, the performance improvement with JSCD is not augmented. However, the improvement in the case of MDTC is slightly more than that in the case of MDSQ, which is mainly due to well-define correlation structure induced by the linear transform.

To this end, we presented the results for a Gaussian source. In a variety of multimedia applications, we encounter source samples that are Laplacian distributed. In order to assess the performance of our system for a Laplacian source, we conducted similar experiments and observed that the system performs similarly for Laplacian distributed inputs.

The amount of correlation between the descriptions is an effective factor on the performance improvement in the iterative JSCD. It is noteworthy to emphasize the distinct ways that the correlation is created between the descriptions for MDSQ and MDTC. Recall that in the case of MDSQ, the correlation between the descriptions are induced by the index assignment and number of diagonals that are occupied. With MDTC, one can obtain a more well-defined correlation between the descriptions. It is also easier to provide a smooth tradeoff between the correlation introduced by MDC and the performance improvement by JSCD that uses this correlation. MDTC is a technique that enables one to control the amount of correlation as desired through linear transformation.

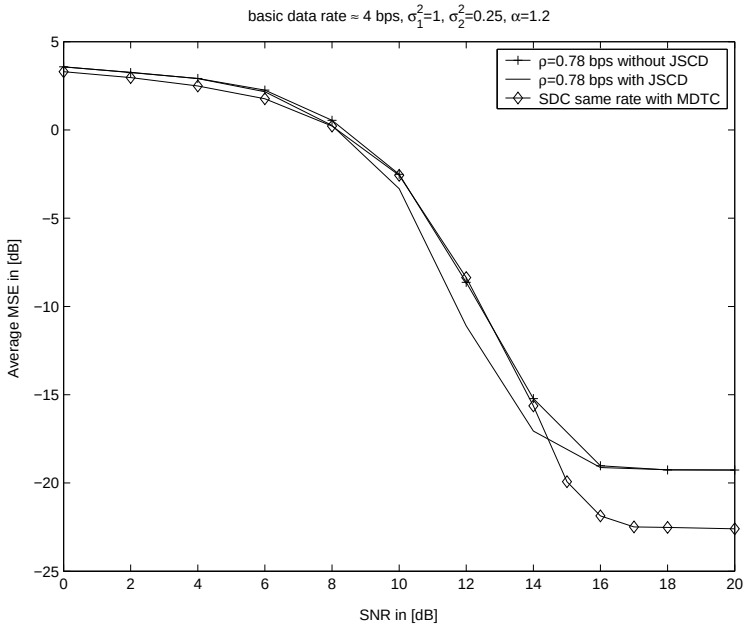


Figure 33: MSE vs. SNR for the MDTC with $\alpha = 1.2$. Turbo code block size is 1500, m=2 transmit and m=1 receive antennas. The block size for $m = 1$ is 3000.

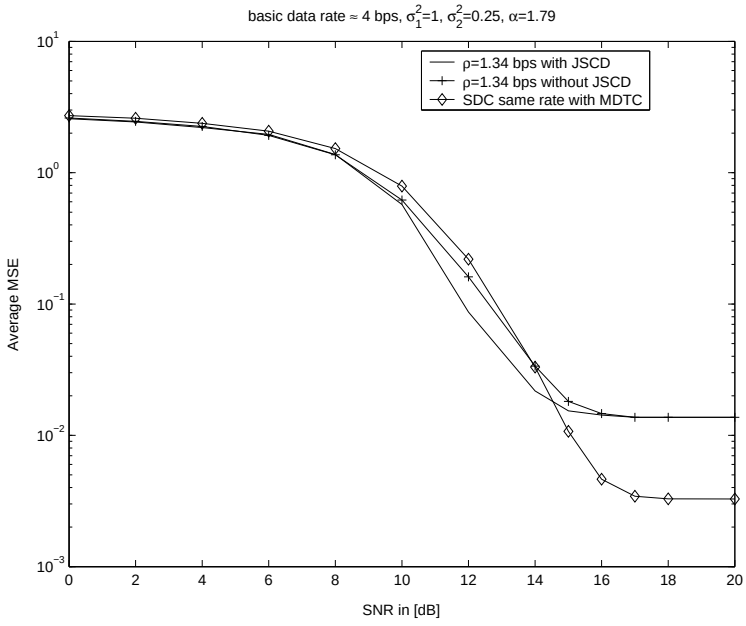


Figure 34: MSE vs. SNR for the MDTC with $\alpha = 1.79$. Turbo code block size is 1500, m=2 transmit and m=1 receive antennas. The block size for $m = 1$ is 3000.

5.5 Chapter Summary

We proposed a turbo coded multiple description coding scheme suitable for multimedia transmission over a wireless link with multiple transmit and receive antennas. We described a suitable iterative joint source-channel decoding algorithm that exploits the correlation between these descriptions. The simulation results show that the performance of the turbo coded MDC system with joint source-channel decoding is superior to the performance of the decoding without joint decoding, and the MDC system provides less distortion than a conventional single description system having the same code rate for moderate SNR values.

CHAPTER VI

TURBO CODES FOR WIRELESS RELAY CHANNELS

In the previous chapters, we studied MIMO systems with multiple transmit and receive antennas. In this chapter, we will deal with another form of MIMO system due to the presence of multiple users that communicate and cooperate with each other. We restrict our work work on the three-terminal wireless relay channel and the case where each node is equipped with a single antenna. We propose a turbo code-based distributed channel code for this system. The organization of the chapter is as follows: In the next section, we provide a summary of previous work related to channel coding for multiterminal networks. In Section 6.2, we describe the relay channel being considered. The turbo code for the relay channel is described in Section 6.3. The iterative decoding methods are explained in Section 6.4. In Section 6.5, we illustrate the system performance via simulations, and finally, in Section 6.6, we conclude the chapter.

6.1 Introduction

The relay channel, which is proposed by Van Der Meulen in 1971 [172], is a channel with three-terminals. The block diagram of such a channel is depicted in Figure 35. This channel can be represented using four finite sets: $\mathcal{X}, \mathcal{X}_R, \mathcal{Y}, \mathcal{Y}_R$, and a set of conditional probabilities $p(\cdot, \cdot|x, x_R)$ on $\mathcal{Y} \times \mathcal{Y}_R$ for each $(x, x_R) \in \mathcal{X} \times \mathcal{X}_R$. Here, x is the original information to be transmitted to the destination. The relay captures the signal y_R and transmits a causal signal depending on the current and previously received signals. The receiver observes the superposition of the signals transmitted from the source and the relay. This channel combines a broadcast channel $(X \rightarrow Y, Y_1)$ and a multiaccess channel $(X, X_1 \rightarrow Y)$.

El Gamal and Cover investigate the capacity of the relay channel in [33], and derive the capacity expressions for Gaussian relay channel and some discrete relay channels, and also compute lower and upper bounds for achievable rates for the general relay channel. With

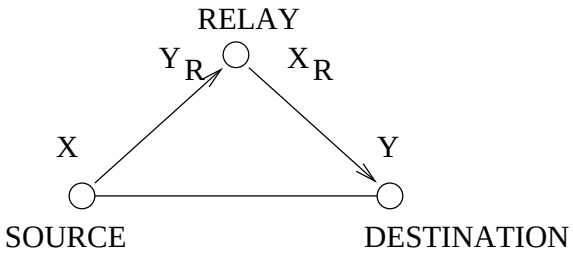


Figure 35: Three-terminal relay channel

the recent advances in wireless networking, an increased interest arose for relay systems [79, 100, 171]. The capacity of wireless relaying is studied in [56, 74, 75, 118, 126]. In [74], Madsen evaluates the capacity of relay channel assuming that the nodes operate in either full duplex or time/frequency division duplex modes. Madsen and Zhang later consolidate these results in [75], and propose that relay channel signaling is superior to traditional multihop protocols, and one can achieve a larger capacity by using power allocation. Reznik *et al.* determines the capacity of a degraded Gaussian relay channel with multiple relay stages. The capacity of the wireless relay networks is investigated in [56].

In addition to the information theoretical results, there has also been some research on signaling techniques and coding for the relay channel. Typical strategies for relaying include *amplify-and-forward*(AF) or *decode-and-forward*(DF) techniques [100, 102]. Any coding scheme may use these strategies in the relay node. In general, DF technique outperforms the other [102]. A simplistic method for relay channel coding is the one for which we assume no direct transmission from source to destination. In such cases, one can use multi-hop transmissions in which AF or DF is employed at the relays. The performance of multi-hop systems is studied in [79–81]. From capacity perspective, however, this strategy is inferior to the one that also utilizes the direct transmission from source to destination [74].

In the classical relay channel, the nodes are assumed to be capable of transmitting and receiving simultaneously at the same time slot and in the same frequency band. However, for some systems, this assumption is not practical, therefore, the source and relay nodes may use orthogonal subchannels for transmission and reception, *e.g.*, time sharing. This model avoids the interference between the source and the relay nodes during the transmission of relay, but it can not achieve the capacity promised by the relay channel [74]: One needs to

transmit simultaneously from the source and relay.

Noting that the transmissions for wireless relaying occur at different spatial locations, one can obtain a *distributed spatial diversity* using relays [101]. *Cooperative diversity* in [99], for example, can achieve spatial diversity with multiple relays by using the techniques described therein. A similar system is also studied in [171] where *macro diversity* can be obtained using relay terminals. Another cooperative diversity technique based on relaying is the *user cooperative diversity* [90, 138–140, 152]. Based on the channel conditions, each node cooperates with other nodes to either act as a relay for them or to transmit its own data.

An interesting coding strategy for relay channels is proposed by Zhao and Valenti [184]. Both the source and the relay nodes use a recursive systematic convolutional (RSC) code so that the destination receives both codes in parallel, which results in a *distributed turbo code*. The relay node uses decode-and-forward technique along with time sharing, however, it first interleaves the decoded stream, hence seeks interleaver gain just like conventional turbo codes. Essentially, the turbo code is embedded in the relay channel. The critical assumption here is that the link between the source and relay is reliable. Although this might be correct for some cases, it might not be a realistic assumption for many other cases.

Most coding techniques mentioned above for relay systems assume that the source and the relay nodes uses time sharing to transmit their data. However, from information theory, we know that the capacity can be achieved by simultaneously transmitting from the source and the relay. In this chapter, we propose turbo-based coding schemes for the relay channel where the source and relay transmissions are performed simultaneously. The source broadcasts a turbo coded block to the relay and destination. The relay decodes and re-encodes the block from the source and forwards them to the destination node. Meanwhile, the source transmits its fresh information block. Hence at the receiver, we observe the superposition of fresh information from the source node and the estimation of the previous block from the relay node. We employ iterative decoding techniques that exploit the information from both the source and relay. Simulation results indicate that we can achieve the relay capacity within 1 dB at a BER of 10^{-5}.

Table 4: Signal and link parameters for the wireless relay channel at time n.

$s[n]$	the signal transmitted from source node
$s_R[n]$	the signal transmitted from relay node
$y_R[n]$	received signal at the relay
$y[n]$	received signal at the destination
$h_{ij}[n]$	the complex channel gain from node j to node i
P	the power of the signals from the source node
P_R	the power of the signals from relay node

6.2 System Model

The block diagram of a wireless relay channel is depicted in Figure 36. To represent the input-output relation, we define the signal and link parameters in Table 4. Then, for the input-output relation at the relay node, we have

$$y_R[n] = \sqrt{P}h_{21}s[n] + w_R[n] \tag{142}$$

while at the destination node, we have

$$y[n] = \sqrt{P}h_{31}[n]s[n] + \sqrt{P_R}h_{32}[n]s_R[n] + w[n] \tag{143}$$

where $w_R[n]$ and $w[n]$ denote the zero mean additive white Gaussian noise at the relay node and at the destination node, respectively. We assume that the channels are independently and identically distributed Rayleigh fading, *i.e.*, h_{ji} are zero mean circularly symmetric complex Gaussian random variables and are independent for different (j, i) pairs. We define signal-to-noise ratio as $SNR = P/N_0$, where N_0 is the one-sided power spectral density of the additive Gaussian noise.

Note that because of simultaneous transmission from the source node and the relay node, the destination observes a noisy superposition of $s[n]$ and $s_R[n]$. We assume that these signals are selected from a BPSK constellation. The transmission from the relay and destination nodes take place in blocks, and we assume that the relay node waits until all the bits (or symbols) in each block is received after which it decodes-and-forwards the estimated block to the destination node.

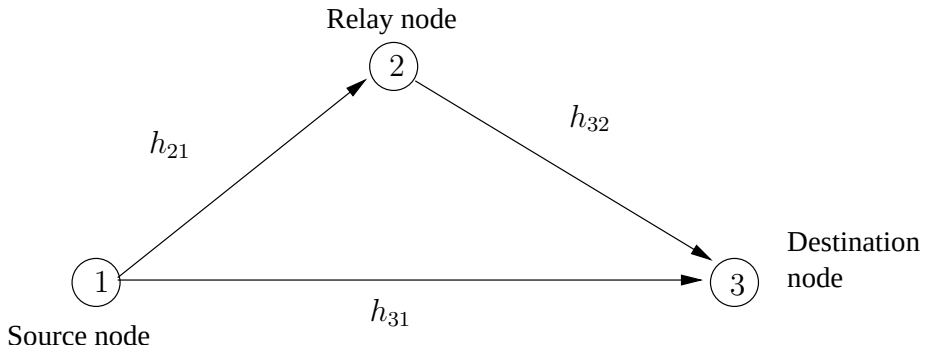


Figure 36: Block diagram for the wireless relay channel

6.3 Turbo Codes for Relay Channels

Consider a block of N message bits, $\mathbf{u} = [u_0, \ldots, u_{N-1}]$, to be transmitted to the destination node. We employ identical block codes of rate-R_c at both relay node and the source node, hence, there are totally N/R_c bits being transmitted at each block from both nodes. The transmission take place as follows (for the l^{th} block): The message bits are turbo coded and mapped to BPSK symbols, and then broadcast to the relay node and the destination node. The relay node receives the bits in the l^{th} block. Meanwhile, it transmits its estimate for the $(l-1)^{th}$ block after decoding and re-encoding. Note that at the first block, the relay node does not transmit any signal, therefore, for this block, the destination node observes only a direct transmission from the source node. Similarly, after the last block is transmitted from the source node, the relay will forward the estimate for this block at the next block slot at which the source does not transmit. We note that this transmission scheme can be imagined as an instance of the block Markov-encoding scheme.

It is straightforward to generate variations of the proposed scheme. For example, one can use different turbo codes at the relay and source nodes. Instead of turbo codes, one can employ convolutional (recursive or nonrecursive) codes as well. Various puncturing schemes can be applied at the relay and source nodes to change the code rate. In this chapter, since turbo codes are powerful codes, we will focus on the performance of the basic turbo coded scheme described above.

We note that a related scheme is proposed by Zhao and Valenti in [184] where they

develop a distributed turbo code for relay channels. Such a code is generated by replacing one of the RSC encoder to the source and the other to the relay node. They assume a half-duplex relay channel, so the transmission takes place in two phases. In the first time slot, the source broadcasts to the relay and destination the systematic and one parity bit sequence. In the second slot, the relay decodes the information bits, interleaves and then re-encodes with its RSC encoder, and finally transmits only the parity bit sequence. The overall encoder behaves like a rate-1/3 turbo code. The receiver uses a standard iterative decoding to decode the turbo code. Note that this scheme divides the time slot into two phases which significantly reduces the spectral efficiency. To benefit from the capacity increase with the relay node, one needs to make simultaneous transmission from the source and he relay nodes as we propose here.

6.4 Decoding

We next describe an iterative decoding method for the destination node. Since the relay node uses a standard iterative turbo decoder which is well-known in the literature, [16,128] (See also Chapter 1), we focus on the details of the decoding at the destination node.

Upon the reception of l^{th} block, $\mathbf{y}^{(l)} = [y^{(l)}[0], \ldots, y^{(l)}[N/R_c - 1]]$, the decoder captures information about two turbo code blocks: one which is the current transmitted block due to fresh information block, and the other which is sent from the relay node and is an estimate for the previous information block. Hence, the decoder can generate the soft information for both of these blocks. For two consecutive blocks, this implies that, at the first block, one can obtain some information about that block (due to direct transmission from the source node to the destination node), and at the next block, it can update the information about that block by processing additional information received from the relay node. One can use different methods to combine the two information. For the case of identical turbo codes at the source and relay nodes, one can simply add the associated extrinsic information obtained using the different nodes, as will be made clear shortly.

The decoding is performed in two steps: Consider the l^{th} block. First, we evaluate the log-likelihoods of the message bits in the current block and the estimated bits of the

previous block received from the relay node. Next, we obtain the soft information on the bits in the $(l-1)^{\text{th}}$ block by combining the soft information due to the transmission (i) from the relay node at the l^{th} block and (ii) from the source node at the $(l-1)^{\text{th}}$ block. Hard decision is made using this final soft output.

After the l^{th} block is received, the log-likelihoods of the i^{th} bit, $b^{(l)}[i]$, from the source node is computed as

$$\Lambda(b^{(l)}[i]) = \frac{P(b^{(l)}[i]=1|y^{(l)}[i])}{P(b^{(l)}[i]=0|y^{(l)}[i])} \tag{144}$$

$$= \frac{\sum_{s[i],s_R[i]:b[i]=1} P(s[i],s_R[i]|y[i])}{\sum_{s[i],s_R[i]:b[i]=0} P(s[i],s_R[i]|y[i])} \tag{145}$$

$$= \frac{\sum_{s[i],s_R[i]:b[i]=1} f(y[i]|s[i],s_R[i])P(s[i],s_R[i])}{\sum_{s[i],s_R[i]:b[i]=0} f(y[i]|s[i],s_R[i])P(s[i],s_R[i])} \tag{146}$$

$$= \frac{\sum_{s[i],s_R[i]:b[i]=1} f(y[i]|s[i],s_R[i])}{\sum_{s[i],s_R[i]:b[i]=0} f(y[i]|s[i],s_R[i])} \tag{147}$$

$$= \frac{\sum_{s_R[i]} e^{-|y[i]-\sqrt{P}h_{31}-\sqrt{P_R}h_{32}s_R[i]|^2/N_0}}{\sum_{s_R[i]} e^{-|y[i]+\sqrt{P}h_{31}-\sqrt{P_R}h_{32}s_R[i]|^2/N_0}} \tag{148}$$

where in (146), we assume that the signals $s[i]$ and $s_R[i]$ are equally likely and independent from each other. Note that we drop the superscript $((\cdot)^{(l)})$ for brevity. In a similar way, the log-likelihood ratio for the i^{th} (estimated) bit from the relay node, $\hat{b}^{(l-1)}$, can be obtained as

$$\Lambda(\hat{b}_R^{(l-1)}[i]) = \frac{P(\hat{b}_R^{(l-1)}[i]=1|y^{(l)}[i])}{P(\hat{b}_R^{(l-1)}[i]=0|y^{(l)}[i])} \tag{149}$$

$$= \frac{\sum_{s[i]} e^{-|y[i]-\sqrt{P}h_{31}s[i]-\sqrt{P_R}h_{32}|^2/N_0}}{\sum_{s[i]} e^{-|y[i]-\sqrt{P}h_{31}s[i]+\sqrt{P_R}h_{32}s_R[i]|^2/N_0}} \tag{150}$$

Assuming that the log-likelihood ratios $\Lambda(\hat{b}^{(l-1)})$ and $\Lambda(b^{(l-1)})$ are independent, we can combine the two information simply by adding them up

$$\Lambda_f(b^{(l-1)}) = \Lambda(\hat{b}^{(l-1)}) + \Lambda(b^{(l-1)})$$

Finally, we feed this final log-likelihood ratio to the turbo decoder assuming that they are the log-likelihoods of the encoded bits as if they are obtained for a system using BPSK modulation over AWGN channel. This iterative decoding method is clearly suboptimal,

however, as shown by the simulations, we can attain very close to the capacity limits with this decoding method.

6.4.1 Improved Decoder

The iterative decoding method described above can be improved by a simple modification. Recall that in (146), we assumed $s_R[i]$ and $s[i]$ were transmitted equally likely. However, using previous transmitted block, we already obtain the log-likelihood ratios for these signals. Hence, we can make use of those likelihood values in the computation of $\Lambda(b^l)$ in (146). In this case, the likelihood for $b^{(l)}[i]$ becomes

$$\Lambda(b^{(l)}[i]) = \frac{\sum_{s_R[i]} P(s_R^{(l)}[i]) e^{-|y[i]-\sqrt{P}h_{31}-\sqrt{P_R}h_{32}s_R[i]|^2/N_0}}{\sum_{s_R[i]} P(s_R^{(l)}[i]) e^{-|y[i]+\sqrt{P}h_{31}-\sqrt{P_R}h_{32}s_R[i]|^2/N_0}} \tag{151}$$

$P(s_R^{(l)}) = P(b^{(l-1)}[i])$, and the likelihood for $b_R^{(l)}[i]$ becomes

$$\Lambda(b_R^{(l-1)}[i]) = \frac{\sum_{s[i]} P(s^{(l)}[i]) e^{-|y[i]-\sqrt{P}h_{31}s[i]-\sqrt{P_R}h_{32}|^2/N_0}}{\sum_{s[i]} P(s^{(l)}[i]) e^{-|y[i]-\sqrt{P}h_{31}s[i]+\sqrt{P_R}h_{32}|^2/N_0}} \tag{152}$$

where $P(s^{(l)}[i]) = P(b^{(l)}[i])$ can be computed using (151). Note that the likelihood computation of the bits in the fresh information block proceeds that of the bits coming from the relay nodes. This is because we have access to the probability of the bit transmitted from the relay node since it is an estimate of the bit in the previous fresh information block. After evaluating the likelihood of the bit in the current block using (151), we can use this soft information to improve the likelihood value of the bit arriving from the relay node.

6.4.2 Relation to Backward Decoding

We can obtain a better decoder if we assume that all transmitted blocks are available at the destination node. If we wait until all the blocks are received, we can use a backward decoding algorithm based on the successive cancelation method. Let us explain this in an abstract form: Consider three blocks A, B, C, each having N/R_c bits, are transmitted by the source and $\hat{A}, \hat{B}, \hat{C}$ are transmitted by the relay. The receiver observes the sequence of blocks $A, B + \hat{A}, C + \hat{B}, \hat{C}$ (one needs 4 slots). With backward decoding, we first decode $C_d = \hat{C}$, then subtract the effect of C_d from $C + \hat{B}$, and then decode $B_d = \hat{B}$, and continue

in this way until decoding all the blocks down to the first one. Of course, this decoding is much less practical since the delay will be much higher (*e.g.*, for the example above, we have to wait until we receive all 4 blocks), and the decoding complexity and the memory requirements will be huge. The decoding proposed in the Sections 6.4 and 6.4.1 require a delay of only 2 blocks for any number of total transmitted turbo code blocks, and hence incur much less complexity.

The decoding method we propose is in fact resembling to the backward decoding. The difference is that we start to decode just after we receive the second block from the source node, and exploit the information coming from the consecutive blocks in an iterative manner, starting from the very first block. Hence, we expect the performance to be very close the optimal one.

6.5 Simulations

In this section, we will illustrate the performance of the system using numerical simulations. We employ a turbo code consisting of two rate-1/2 recursive systematic convolutional codes with $(g_n, g_d) = (37, 21)_{octal}$, therefore the overall code rate at each node is $R_c = 1/3$. We assume a block length of $N = 3000$ bits. We consider two scenarios: (i) The link from the source node to the relay node is ideal, so the decoding at the relay node is error-free, while the relay-destination link and the direct link have equal variance. (ii) The relay node is located such that the source-relay link and the relay-destination link are 6 dB better than the direct link. In all cases, we assume that the channel is a fast Rayleigh fading channel. We set the transmission power at the source and relay nodes equal to each other, *e.g.*, $P = P_R$.

We first study the Scenario 1. Figures 37 and 38 illustrate the bit error rates for the proposed scheme and compare the performance with the direct transmission approach. For the direct transmission, we assume that the same turbo code is employed, and for a fair comparison, we set the transmission power $P_{direct} = 2P$. For the proposed scheme, we depict the performance for both decoding scheme described earlier. The simulation results show that at both code rates, the use of a relay node significantly improves the performance.

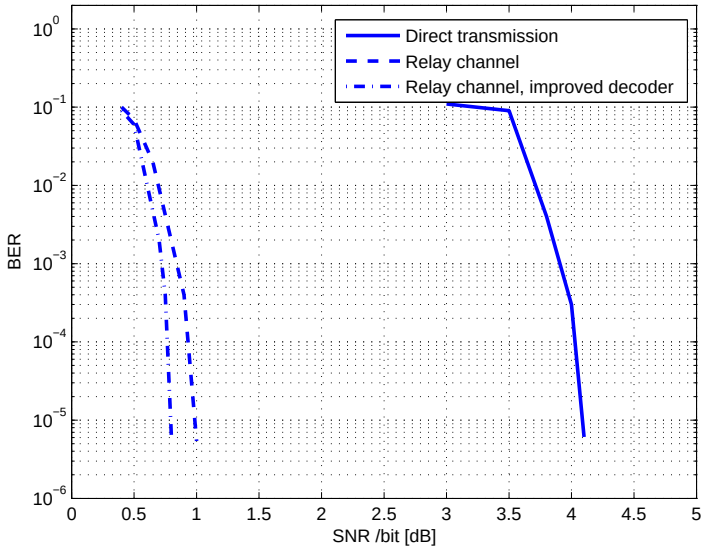


Figure 37: BER vs. SNR when the source-relay link is ideal and the relay-destination link has the same quality as the direct link. Simulation parameters: Turbo code block size, N= 3000, $R_c = 1/3$.

For example, for $R_c = 1/3$, we have about 3 dB gain using the basic iterative decoder, and 3.5 dB gain using the improved decoder at a bit error rate of 10^{-5} compared to the direct transmission.

In Figure 39, we plot the bit error rate when the turbo code block size is 64000 bits. We observe that as the block size increases, the performance improves significantly. For example, with the improved decoder, we can attain a bit error rate of 10^{-5} at a signal-to-noise ratio of -2.3 dB. We note that the minimum SNR required to attain 1/2 bits/channel use for this case is about -3.5 dB. Hence, we are only 1.2 dB away from this limit. This clearly shows that the proposed turbo coding scheme for the relay channel is very promising.

In Figures 40 and 41, we study the second scenario, *i.e.*, both the source-relay and the relay-destination links are fast fading and have 6 dB more power than the direct link. For this case, we also plot the bit error rate obtained with a multi hop transmission scheme. For these set of curves, the following comments are in order: (i) The worst performance is obtained using a direct transmission link. This emphasizes the significance of a relay node. (ii) Multi-hop transmission is about $0.5 - 1$ dB better than the direct transmission. (iii)

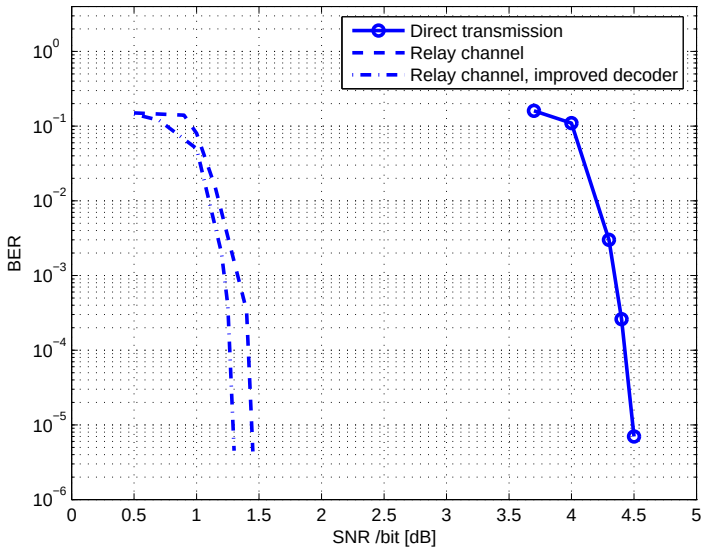


Figure 38: BER vs. SNR when the source-relay link is ideal and the relay-destination link has the same quality as the direct link. Simulation parameters: Turbo code block size, N= 3000, $R_c = 1/2$.

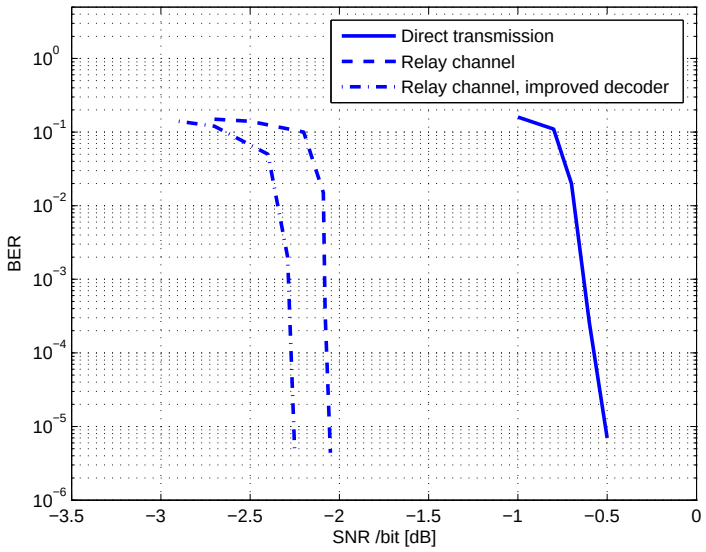


Figure 39: BER vs. SNR when the source-relay link is ideal and the relay-destination link has the same quality as the direct link. Simulation parameters: Turbo code block size, N= 64000, $R_c = 1/2$.

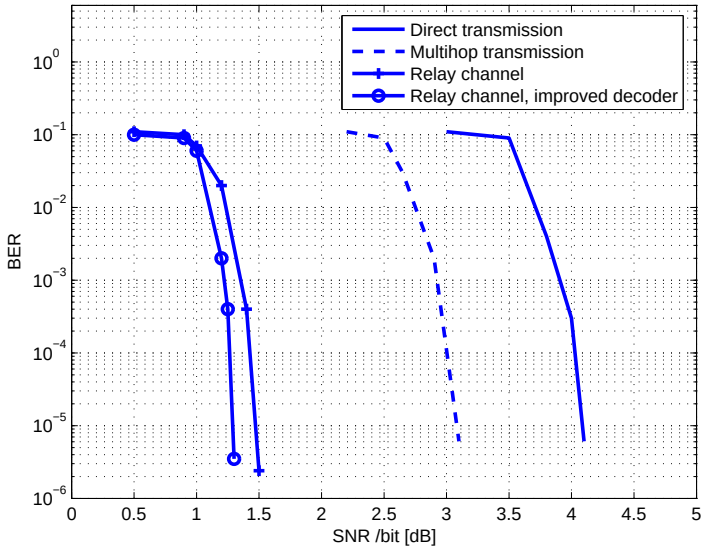


Figure 40: BER vs. SNR when the source-relay link the relay-destination link are both 6 dB better than the direct link. Simulation parameters: Turbo code block size, N= 3000, $R_c = 1/3$.

The best performance is obtained by the turbo coding scheme we proposed. It is clear that in order to achieve better performance, the relay and the source node should cooperate.

While the performance of the proposed scheme is significantly superior to the alternative schemes, we are still $2-3$ dB away from the capacity. We believe that this gap can be reduced by suitable design of the component codes at the source and relay nodes, and in addition to that, by using the more complex decoding methods such as the backward decoding algorithm. The disadvantage in these scheme is the increase in the delay and complexity. This is in accordance with the Shannon coding theorem [34], since in order to achieve the capacity, one generally requires infinite code length and delay.

6.6 Chapter Summary

We presented a turbo coding scheme for wireless relay channels in which the relay node has full-duplex capability. We described suitable decoding algorithms and showed that with the proposed decoding, the system performs within $1-2$ dB of the capacity limits. The turbo code based scheme performs superior to its alternatives of direct transmission and the multihop transmission. While we considered a single-relay node, the scheme can readily

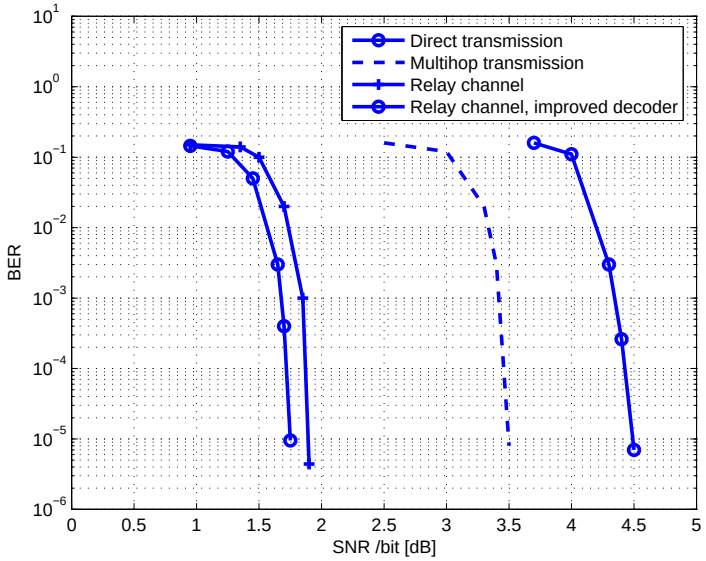


Figure 41: BER vs. SNR when the source-relay link the relay-destination link are both 6 dB better than the direct link. Simulation parameters: Turbo code block size, N= 3000, $R_c = 1/2$.

be generalized to the case of multiple relay nodes. In this case, the resulting scheme will be similar to a multiple component turbo codes.

CHAPTER VII

PARALLEL AND SERIAL DISTRIBUTED DETECTION FOR WIRELESS SENSOR NETWORKS

In this chapter, we study several distributed detection strategies in wireless sensor networks. Our contributions will be in three folds. First, we consider the distributed detection in fading environments, that is, we assume noisy channels between sensor nodes, which is quite different from the traditional distributed detection schemes that assume error-free reception of locally transmitted signals. Second, for various strategies, we derive the false alarm and detection probabilities which are then used for performance assessment and decision-rule optimization. Third, we introduce an analog approach to distributed detection problem and analyze its performance. The organization of this chapter is as follows: In the next section, we give a summary of distributed detection in wireless sensor networks. In Section 7.2, we analyze the performance of distributed detection when the local sensors perform a single-bit local decision. We consider two main topologies: a parallel network and a serial network. Section 7.3 deals with the case that multiple-bit local decisions are transmitted to the fusion center. In Section 7.4, we study an analog scheme where local decision statistics are transmitted without any local processing. Section 7.5 proposes the use of sequential detection in WSNs over noisy fading channels. Numerical examples are illustrated in Section 7.6. Finally, the we sumamrize our results in Section 7.7.

7.1 Introduction

Wireless sensor networks (WSNs) is an emerging technology that experiences a pervasive trend in many application areas including environment monitoring, health, security and surveillance, and robotic exploration [92]. Networks of sensor systems allow for many distributed processing and cooperative communication techniques including distributed data compression [148], tracking and classification [37], and distributed detection [91,130][4,5]. In

this chapter, we focus on distributed detection that are specially tailored for wireless sensor networks. In distributed detection, each sensor sends its observation to the fusion center where a global decision is made. Because of the bandwidth and energy limitations, instead of transmitting the raw data, each sensor generally performs a local detection process and sends its decision (possibly consisting of a few bits) to the fusion center. The fusion center collects all decisions from all sensors and performs a final decision on the hypothesis under investigation.

The research studies on distributed detection generally assume error-free transmission between the sensor nodes and the fusion center [122,130]. This assumption might be viable in traditional radar or sonar networks. However, in sensor networks which are constructed by battery-limited low-power tiny devices, the assumption of error-free transmission is not wise since such transmissions will require powerful error correction coding and very complex signal processing algorithms.

There are three major topologies that are considered in distributed detection: parallel, serial, and tree configuration [130], and several distributed detection algorithms have been investigated for such configurations [91, 119, 129, 130]. Optimal distributed detection algorithms have been focused on optimality under the Neyman-Pearson and Bayesian detection criteria. Under the assumption of conditionally independent observations, the optimal fusion rules are given by likelihood ratio (LR) tests at the individual sensors and at the fusion center [122]. If the observations at different sensor nodes are correlated (dependent), the optimal fusion rules become intractable: they do not reduce to LR tests [39,103]. Distributed detection algorithms have also been investigated under several communication-constraints [94, 111, 148]. Chamberland and Veeravalli [94] showed that under certain conditions, for an N-sensor network with a capacity constraint of N bits per time unit, having each sensor transmitting one bit is optimum. Thomopoulos and Zhang investigates the distributed detection in the case of non-ideal channels [165]. In [42], Duman and Salehi consider the distributed detection over multi access channels where the fusion center gathers the decisions from local sensors via a multi-access channel.

Recently, Chen et al. proposed a channel-based decision fusion for a parallel network

of sensors linked with fading channels [24, 25, 119]. Assuming parallel configuration, the authors incorporate the effect of fading in the detection process, and derive optimal fusion rules and some alternative fusion rules based on diversity combining techniques. In [105], a similar decision fusion for a multihop transmission is considered. While the performance of the decision fusion based on some suboptimal methods are evaluated in these work, the optimality of the decision rules at local sensors and at the fusion center, and optimal designs are not considered. Recently, Chen and Willet have sown that optimal local decisions that minimize the error probability at the fusion center becomes a likelihood-ratio test (LRT) under some particular constraints on the fusion rule [26].

Because of the hard-energy limitations in a sensor network, it is generally preferable to perform a local processing on the raw observed data and transmit the compressed data to the fusion center. For point-to-point communications, however, it is well-known that an uncoded analog transmission is the optimal choice to transmit the data over an additive Gaussian noise channel [50, 55]. In fact, such a scheme is very desirable in a delay-sensitive transmission since as soon as the data is observed it is transmitted to the destination. In a wireless network with multiple nodes, it is not clear which way is a better choice. For a distributed estimation scheme, the performance of analog and digital approaches are compared in [35]. To the best of author's knowledge, the performance of distributed detection with analog data gathering has not been investigated yet.

In this chapter, we present a comprehensive work that addresses the problem of energy-efficient distributed detection in wireless sensor networks in which the links between the sensor nodes are subject to fading. We analyze the performance of different detection strategies and develop design tools required for optimizing their detection performance. We classify two main strategies: (i) the digital approach where the sensors make their local decisions first and then forward these decisions to the associated destination nodes, and (ii) the analog approach where the locally observed samples are directly transmitted by means of analog waveforms. For each scheme, we study the performance of serial and parallel networks. For the former approach, we first study the distributed detection with single-bit local decisions and then look at the problem of multiple-bit sensor decisions. We

observe that multiple-bit sensor decisions provide better detection performance. In the analog approach, it is seen that the fusion of received analog samples gives results that are very similar to those obtained by a centralized decoder. For this case, we develop the optimal power allocation scheme that provides the best detection performance according to NP-lemma. While our analysis models the local samples as Gaussian random variables, the results for any other source distribution can be obtained with our formulation either analytically or numerically. Numerical results indicate that the detection performance with the analog approach is superior to the digital approach for a given power-budget. Hence, the analog approach is more efficient for distributed detection over fading channels. The simulation results also indicate that a two-bit decision strategy in a parallel network provide better detection capability compared to its single-bit counterpart.

7.2 Distributed Detection with Single-Bit Local Decisions

Consider the detection of a binary event by the help of a network of distributed sensors. We initially assume that the local sensors are allowed to produce a single-bit decision, i.e., one level quantization, based on their observations. We will later generalize this scheme to the case of multiple bit decisions and to the case where the observed data itself or a sufficient statistics is transmitted using analog waveforms. We study two different network configurations: (i) a serial network, and (ii) a parallel network. The block diagrams for these configurations are depicted in Figures 42 and 43. Let us first consider the serial network.

7.2.1 Serial Configuration

In the serial network of N sensors shown in Figure 42, let y_j denote the observation or a sufficient statistics at the j^{th} sensor node S_j. Also denote by $u_j \in \{0, 1\}$ the binary decision at S_j. A binary modulation scheme is used for transmitting the decisions from S_{j-1} to S_j. Assuming frequency flat fading and additive noise at each link, the received signal at S_j, denoted by r_{j-1}, is given by

$$r_{j-1} = \sqrt{\rho_{j-1}} g_{j-1} s_{j-1} + n_{j-1} \tag{153}$$

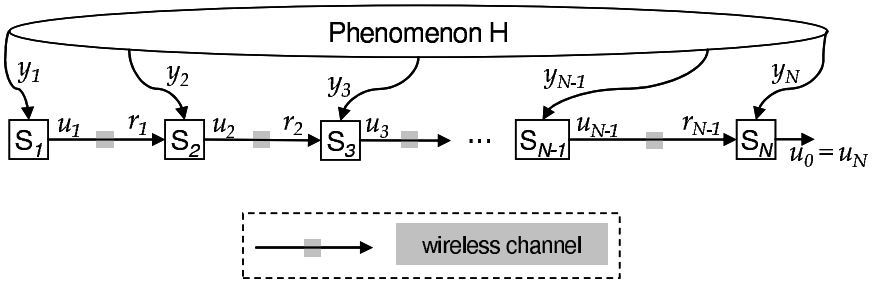


Figure 42: Serial configuration for distributed detection for wireless sensor networks with fading channels

where g_{j-1} is the complex-valued channel gain between S_{j-1} and S_j, n_{j-1} is the additive white Gaussian noise at S_j, and ρ_{j-1} is the transmitter power gain at S_{j-1}. We assume $n_j \sim \mathcal{CN}(0,1)$[1], and they are independent and identically distributed (iid) for $j = 1, \cdots, N-1$. We assume a quasi-static fading channel and so the channel state information (CSI), *i.e.,* g_{j-1}, can be estimated at S_j and can be forwarded to the fusion center with a control channel. It is possible to perform power allocation according to some optimality criterion, which becomes computationally complex for the digital approach considered in this section. Instead, we assume that each sensor transmits at the same power level, and hence, we set $\rho_{j-1} = \rho$, for $j = 2, \ldots, N$. Thus, we derive the optimal detection rules for a uniform power allocation scheme.

The decision at the j^{th} stage is based on the observation, y_j, and the received signal r_{j-1}. We assume that the observations and the received signals at the sensors are statistically independent conditioned on the hypothesis. That is, y_j and r_{j-1} are conditionally independent.

We define the false alarm and detection probabilities at S_j as $P_{F,j} = \Pr(u_j = 1|H_0)$, $P_{D,j} = \Pr(u_j = 1|H_1)$. Our goal is to derive fusion rules based on the Neyman-Pearson lemma, that is, for a prescribed bound on the global false alarm rate, $P_{F,N}$, we wish to find the decision rules that maximize the global detection rate, $P_{D,N}$.

[1] $\mathcal{CN}(0,1)$ denotes the circularly symmetric zero mean and unit variance Complex Gaussian random variable whose density is given by $p_n(z) = \frac{1}{\pi} e^{-|z|^2}$

7.2.1.1 Decision Fusion Rule and Error Probabilities

According to Neyman-Pearson lemma, the optimal decision rules at each stage reduces to LR tests, where the LR at the j^{th} stage can be computed using the received signal r_{j-1} and the observation y_j:

$$\Gamma(y_j, r_{j-1}) = \frac{L(y_j, r_{j-1}|H_1, g_{j-1})}{L(y_j, r_{j-1}|H_0, g_{j-1})} = \frac{p(y_j|H_1)p(r_{j-1}|H_1, g_{j-1})}{p(y_j|H_0)p(r_{j-1}|H_0, g_{j-1})} \tag{154}$$

where

$$p(r_{j-1}|H_1, g_{j-1}) = P_{D,j-1}p_n(r_{j-1} - \sqrt{\rho}g_{j-1}s^1) + (1 - P_{D,j-1})p_n(r_{j-1} - \sqrt{\rho}g_{j-1}s^0), \tag{155}$$

$$p(r_{j-1}|H_0, g_{j-1}) = P_{F,j-1}p_n(r_{j-1} - \sqrt{\rho}g_{j-1}s^1) + (1 - P_{F,j-1})p_n(r_{j-1} - \sqrt{\rho}g_{j-1}s^0) \tag{156}$$

Let $\Lambda(y_j) = p(y_j|H_1)/p(y_j|H_0)$, and $\Upsilon(r_{j-1}) = p(r_{j-1}|H_1, g_{j-1})/p(r_{j-1}|H_0, g_{j-1})$. Assuming binary phase shift keying (BPSK) modulation, we can rewrite $\Upsilon(r_{j-1})$ as

$$\Upsilon(r_{j-1}) = \frac{P_{D,j-1}\xi_{j-1} + 1 - P_{D,j-1}}{P_{F,j-1}\xi_{j-1} + 1 - P_{F,j-1}} \tag{157}$$

where $\xi_{j-1} = \exp\left(4\sqrt{\rho}\Re\{r_{j-1}g_{j-1}^*\}\right)$. Using (154) and (157), the LR test at the j^{th} node is given by

$$\Lambda(y_j)\Upsilon(r_{j-1}) \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{158}$$

where t is a threshold to be determined. For simplicity, it is convenient to use the log-likelihood ratios, $\Gamma^*(y_j, r_{j-1}) = \log(\Gamma(y_j, r_{j-1}))$, $\Lambda^*(y_j) = \log(\Lambda(y_j))$ and $\Upsilon^*(r_{j-1}) = \log(\Upsilon(r_{j-1}))$, and hence, we can rewrite the LR test (158) as

$$\Gamma^*(y_j, r_{j-1}) = \Lambda^*(y_j) + \Upsilon^*(r_{j-1}) \underset{H_0}{\overset{H_1}{\gtrless}} t^* \tag{159}$$

where $t^* = \log(t)$. For the first stage, we have $\Upsilon^*(r_{j-1}) = 0$. Although it is straightforward to implement the fusion rule described by (159), note that it requires the exact knowledge of the channel gain g_{j-1} and the false alarm & detection probabilities at the previous stage.

False Alarm and Detection Probabilities

We next derive the false alarm and detection probabilities to evaluate the performance of the decision fusion rule in (159). At the j^{th} stage, the false alarm probability is given by

$$P_{F,j} = P_{F,j-1}\Pr\left(\Lambda^*(y_j) + \Upsilon_1^*(r_{j-1}) > t^*|H_0\right) + (1-P_{F,j-1})\Pr\left(\Lambda^*(y_j) + \Upsilon_0^*(r_{j-1}) > t^*|H_0\right) \tag{160}$$

where

$$\Upsilon_1^*(r_{j-1}) = \log\frac{P_{D,j-1}e^{4\rho|g_{j-1}|^2+4\sqrt{\rho}n'_{j-1}} + 1 - P_{D,j-1}}{P_{F,j-1}e^{4\rho|g_{j-1}|^2+4\sqrt{\rho}n'_{j-1}} + 1 - P_{F,j-1}},$$

$$\Upsilon_0^*(r_{j-1}) = \log\frac{P_{D,j-1}e^{-4\rho|g_{j-1}|^2+4\sqrt{\rho}n'_{j-1}} + 1 - P_{D,j-1}}{P_{F,j-1}e^{-4\rho|g_{j-1}|^2+4\sqrt{\rho}n'_{j-1}} + 1 - P_{F,j-1}},$$

with $n'_{j-1} \sim \mathcal{C}(0, |g_{j-1}|^2/2)$. Let $\Gamma_i^* = \Lambda^* + \Upsilon_i^*$ (For brevity, we drop y_j and r_{j-1}). Denote the cumulative distributions of Γ_i^*, Λ^* under H_1 and H_0 as $F_{\Gamma_{i,1}^*}(\cdot)$, $F_{\Lambda_1^*}(\cdot)$ and $F_{\Gamma_{i,0}^*}(\cdot)$, $F_{\Lambda_0^*}(\cdot)$, respectively. Also denote the density functions of Υ_1^* and Υ_0^* as $f_{\Upsilon_1^*}(\cdot)$ and $f_{\Upsilon_0^*}(\cdot)$, respectively. Using probability theory [5], we can show that

$$\begin{aligned} f_{\Upsilon_k^*}(y) =& \frac{(P_{D,j-1}-P_{F,j-1})e^y}{(P_{D,j-1}-P_{F,j-1}e^y)((1-P_{F,j-1})e^y-(1-P_{D,j-1}))}\frac{1}{\sqrt{16\pi\rho|g_{j-1}|^2}}\times \\ & \exp\left(-\left(\log\frac{(1-P_{F,j-1})e^y-(1-P_{D,j-1})}{P_{D,j-1}-P_{F,j-1}e^y} - (2k-1)4\rho|g_{j-1}|^2\right)^2/16\rho|g_{j-1}|^2\right) \end{aligned} \tag{161}$$

for $\log\frac{1-P_{D,j-1}}{1-P_{F,j-1}} < y < \log\frac{P_{D,j-1}}{P_{F,j-1}}$. Due to the assumption of conditional independence of y_j and r_{j-1}, $F_{\Gamma_{i,k}^*}$, $i,k \in \{0,1\}$, can be expressed as

$$F_{\Gamma_{i,k}^*}(a) = \int_{\log\frac{1-P_{D,j-1}}{1-P_{F,j-1}}}^{\log\frac{P_{D,j-1}}{P_{F,j-1}}} f_{\Upsilon_i^*}(y)F_{\Lambda_k^*}(a-y)dy. \tag{162}$$

Using (162) in (160), we finally obtain

$$P_{F,j} = 1 - \int_{\log\frac{1-P_{D,j-1}}{1-P_{F,j-1}}}^{\log\frac{P_{D,j-1}}{P_{F,j-1}}} (P_{F,j-1}f_{\Upsilon_1^*}(y) + (1-P_{F,j-1})f_{\Upsilon_0^*}(y))F_{\Lambda_0^*}(t^*-y)dy. \tag{163}$$

Similarly, the detection probability can be computed using

$$P_{D,j} = 1 - \int_{\log\frac{1-P_{D,j-1}}{1-P_{F,j-1}}}^{\log\frac{P_{D,j-1}}{P_{F,j-1}}} (P_{D,j-1}f_{\Upsilon_1^*}(y) + (1-P_{D,j-1})f_{\Upsilon_0^*}(y))F_{\Lambda_1^*}(t^*-y)dy. \tag{164}$$

Hence, if the distribution of the observations y_j is known, using Equations (161), (163) and (164), we can compute the $P_{D,j}$ recursively, provided that the $P_{F,j-1}$ are specified. A

simplistic approach is to set the false alarm rates $P_{F,j}$ at all stages the same, however, in that case, one can not guarantee the maximization of $P_{D,N}$, the global detection probability. According to Neyman-Pearson lemma, for a given upper bound on $P_{F,N}$, we need to make an exhaustive search over all $P_{F,j}$, $j = 1, \cdots, N-1$ in order to find those that maximize the global detection probability $P_{D,N}$. Since an analytical solution is not feasible; we resort to numerical search procedures to determine the decision fusion rules, which is usually the case for distributed detection problems.

7.2.1.2 Optimality of LR-based Decision Fusion

So far, we used the LR-based decision fusion rule without considering its optimality. If the channels between the consecutive sensors are error-free, that is, each sensor node can pass its decision to the next one without error, Viswanathan and Thomopoulos have shown that the optimality can be achieved using Neyman-Pearson test at each stage [131]. Here, we investigate the optimality under fading channels for the proposed fusion rule.

Consider the decision fusion at the last two stages. At the final node S_N, we have the log-likelihood ratio $\Gamma^*(y_N, r_{N-1}) = \Lambda^*(y_N) + \Upsilon^*(r_{N-1})$. Let $P_D = P_{D,N-1}$, $P_F = P_{F,N-1}$, and $\Lambda^* = \Lambda^*(y_N)$. We can rearrange (163) and (164) to obtain

$$1 - P_{F,N} = \int_{\log \frac{1-P_D}{1-P_F}}^{\log \frac{P_D}{P_F}} (P_F f_{\Upsilon_1^*}(y) + (1-P_F) f_{\Upsilon_0^*}(y)) F_{\Lambda_0^*}(t^* - y) dy, \tag{165}$$

$$1 - P_{D,N} = \int_{\log \frac{1-P_D}{1-P_F}}^{\log \frac{P_D}{P_F}} (P_D f_{\Upsilon_1^*}(y) + (1-P_D) f_{\Upsilon_0^*}(y)) F_{\Lambda_1^*}(t^* - y) dy. \tag{166}$$

Integration of (165) and (166) by parts gives

$$1 - P_{F,N} = F_{\Lambda_0^*}(t^* - \log \frac{P_D}{P_F}) - \int_{\log \frac{1-P_D}{1-P_F}}^{\log \frac{P_D}{P_F}} (P_F F_{\Upsilon_1^*}(y) + (1-P_F) F_{\Upsilon_0^*}(y)) f_{\Lambda_0^*}(t^* - y) dy, \tag{167}$$

$$1 - P_{D,N} = F_{\Lambda_1^*}(t^* - \log \frac{P_D}{P_F}) - \int_{\log \frac{1-P_D}{1-P_F}}^{\log \frac{P_D}{P_F}} (P_D F_{\Upsilon_1^*}(y) + (1-P_D) F_{\Upsilon_0^*}(y)) f_{\Lambda_1^*}(t^* - y) dy, \tag{168}$$

where $F_{\Upsilon_k^*}$ is the cumulative distribution of Υ_k^*, $k = 0, 1$. It is required for some fixed $P_{F,N}$ and $P_{F,N-1}$ that the $P_{D,N}$ be a monotonic increasing function of $P_{D,N-1}$ so that the global detection probability takes larger values as $P_{D,N-1}$ is increased. The necessary conditions satisfying this requirement can be obtained by taking the derivative of (167) and (168) with

respect to P_D:

$$\frac{d(1-P_{F,N})}{P_D} = \left(\frac{dt^*}{dP_D} - \frac{1}{P_D}\right) f_{\Lambda_0^*}(t^* - \log\frac{P_D}{P_F}) - P_F\frac{dU_{1,0}}{P_D} - (1-P_F)\frac{dU_{0,0}}{P_D} \tag{169}$$

$$\frac{d(1-P_{D,N})}{P_D} = \left(\frac{dt^*}{dP_D} - \frac{1}{P_D}\right) f_{\Lambda_1^*}(t^* - \log\frac{P_D}{P_F}) + U_{1,1} - U_{0,1} - P_D\frac{dU_{1,1}}{P_D} - (1-P_D)\frac{dU_{0,1}}{P_D} \tag{170}$$

where $U_{k,j} = \int_{\log\frac{1-P_D}{1-P_F}}^{\log\frac{P_D}{P_F}} F_{\Upsilon_k^*}(y) f_{\Lambda_j^*}(t^* - y)dy$. Note that if $P_{D,N-1}$ is changed, to keep $P_{F,N}$ at some fixed value, the threshold t^* at S_N needs to be changed as well. The required expression for t^* can be obtained by equating (169) to 0,

$$\left(\frac{dt^*}{dP_D} - \frac{1}{P_D}\right) f_{\Lambda_0^*}(t^* - \log\frac{P_D}{P_F}) = P_F\frac{dU_{1,0}}{P_D} + (1-P_F)\frac{dU_{0,0}}{P_D}. \tag{171}$$

Substituting (171) in (170), we finally arrive at

$$\begin{aligned}\frac{d(1-P_{D,N})}{P_D} = \left(P_F\frac{dU_{1,0}}{P_D} + (1-P_F)\frac{dU_{0,0}}{P_D}\right)\frac{f_{\Lambda_1^*}(t^* - \log\frac{P_D}{P_F})}{f_{\Lambda_0^*}(t^* - \log\frac{P_D}{P_F})} + U_{1,1} - U_{0,1} \\ - P_D\frac{dU_{1,1}}{P_D} - (1-P_D)\frac{dU_{0,1}}{P_D}.\end{aligned} \tag{172}$$

It is usually required that $P_D > P_F$, which implies that $U_{1,1} - U_{0,1} < 0$. To have $\frac{dP_{D,N}}{P_D} > 0$, a sufficient condition is then given by

$$\frac{f_{\Lambda_1^*}(t^* - \log\frac{P_D}{P_F})}{f_{\Lambda_0^*}(t^* - \log\frac{P_D}{P_F})} \leq \frac{P_D\frac{dU_{1,1}}{P_D} + (1-P_D)\frac{dU_{0,1}}{P_D}}{P_F\frac{dU_{1,0}}{P_D} + (1-P_F)\frac{dU_{0,0}}{P_D}} \tag{173}$$

We observe that the left hand side of (173) is the likelihood ratio of the likelihood ratio. In [173], it is shown that the likelihood ratio of a likelihood ratio is the likelihood ratio itself. Hence, the condition in (173) can be reduced to

$$t^* - \log\frac{P_D}{P_F} \leq \log\frac{P_D\frac{dU_{1,1}}{P_D} + (1-P_D)\frac{dU_{0,1}}{P_D}}{P_F\frac{dU_{1,0}}{P_D} + (1-P_F)\frac{dU_{0,0}}{P_D}} \tag{174}$$

We can evaluate $\frac{dU_{k,j}}{P_D}$ using the Leibniz's' formula [71], however, the resulting expression does not allow for a closed form expression for the sufficiency condition in (174). Nevertheless, it is clear that at as long as the threshold t^* satisfies (174), we guarantee that $P_{D,N}$ is an increasing function of P_D and hence, global optimality is achieved by Neyman-Pearson test at each stage. In Section 7.6, we present several numerical results that show that LR-based decision fusion is optimal in the sense that the detection probability increases as the number of sensors in the serial detection process increases.

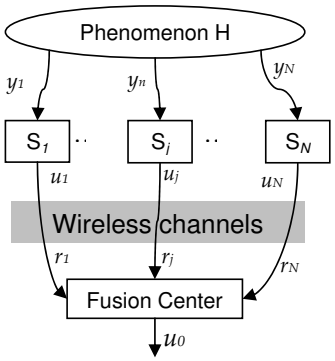


Figure 43: Parallel configuration of a sensor system for distributed detection over fading channels

7.2.2 Parallel Configuration

In Figure 43, we depict the block diagram of a parallel network of N sensors. In a similar manner to serial network, the local sensor observation y_j is first mapped to a local binary decision denoted by $u_j \in \{0, 1\}$, which is then transmitted to the fusion center, S_0 using a binary modulation scheme. Again assuming frequency flat fading and additive noise channel between the local sensor and the fusion center, the signal received from S_j is given by

$$r_j = \sqrt{\rho_j} g_j s_j + n_j \tag{175}$$

where the signal and channel parameters are as described in Section 7.2.1.

The decision at the fusion center is based on the received signals, $\mathbf{r} = [r_1, r_2, \cdots, r_N]$. We assume that both the observations u_j and the received signals r_j are statistically independent for $j = 1, \cdots, N$ conditioned on the hypothesis. The false alarm and detection probabilities at $S_j,\ j = 0, \cdots, N$, are defined as $P_{F,j} = \Pr(u_j = 1|H_0)$, $P_{D,j} = \Pr(u_j = 1|H_1)$. In this case, the fusion rule based on the Neyman-Pearson lemma can be stated as follows: for a prescribed bound on the global false alarm rate, $P_{F,0}$, find the decision rules both at the local sensors and at the fusion center that maximize the global detection rate, $P_{D,0}$.

7.2.2.1 Decision Fusion Rule and Error Probabilities

According to Neyman-Pearson lemma, the optimal decision rules at S_0 reduces to the likelihood ratio test

$$\Gamma(\mathbf{r}) = \frac{L(\mathbf{r}|H_1, \mathbf{g})}{L(\mathbf{r}|H_0, \mathbf{g})} = \prod_{j=1}^{N} \frac{p(r_j|H_1, g_j)}{p(r_j|H_0, g_j)} \underset{H_0}{\overset{H_1}{\gtrless}} t_0 \tag{176}$$

where

$$p(r_j|H_1, g_j) = P_{D,j} p_n(r_j^{(1)}) + (1 - P_{D,j}) p_n(r_j^{(0)}),$$

and

$$p(r_j|H_0, g_j) = P_{F,j} p_n(r_j^{(1)}) + (1 - P_{F,j}) p_n(r_j^{(0)}),$$

with $r_j^{(k)} = r_j - \sqrt{\rho} g_j s^k$. Let $\Upsilon(r_j) = p(r_j|H_1, g_j)/p(r_j|H_0, g_j)$. With binary phase shift keying (BPSK) modulation, we can simplify $\Upsilon(r_j)$ to

$$\Upsilon(r_j) = \frac{P_{D,j}\xi_j + 1 - P_{D,j}}{P_{F,j}\xi_j + 1 - P_{F,j}} \tag{177}$$

where $\xi_j = \exp\left(4\sqrt{\rho}\Re\{r_j g_j^*\}\right)$. Substituting (177) into (176), the LR test at the fusion center can be obtained as

$$\prod_{j=1}^{N} \Upsilon(r_j) \underset{H_0}{\overset{H_1}{\gtrless}} t_0 \tag{178}$$

where t_0 is a threshold to be determined. Taking the logarithm of both sides, we equivalently have

$$\sum_{j=1}^{N} \Upsilon^*(r_j) \underset{H_0}{\overset{H_1}{\gtrless}} t_0^* \tag{179}$$

where $\Upsilon^*(r_j) = \log(\Upsilon(r_j))$, and $t_0^* = \log(t_0)$. Note that the fusion rule described by (179) requires (i) the exact knowledge of the channel gain g_j and (ii) the false alarm & detection probabilities at the local sensors. We assume quasistatic fading where the channel remains constant for a long period. In that case, the CSI can be estimated at the fusion center using a training sequence.

False Alarm and Detection Probabilities

In order to assess the performance of the decision fusion rule in (179), and to develop optimal thresholds, we need to derive the false alarm and detection probabilities. At the fusion center, the false alarm probability is given by

$$P_{F,0} = \Pr\left(\sum_{j=1}^{N} \Upsilon^*(r_j) > t_0^* | H_0\right). \tag{180}$$

Noting that we have 2^N different possible decision vectors $\mathbf{u} = [u_1, \cdots, u_N]$, and using the total probability theorem [5], we have

$$P_{F,0} = \sum_{k=0}^{2^N-1} \Pr(\mathbf{u} = \mathbf{u}_k | H_0) \Pr\left(\sum_{j=1}^{N} \Upsilon^*(r_j) > t_0^* | \mathbf{u} = \mathbf{u}_k, H_0\right) \tag{181}$$

where $\Pr(\mathbf{u} = \mathbf{u}_k | H_0) = \prod_{i=1,\mathbf{u}=\mathbf{u}_k}^{N} P_{F,i}^{u_i}(1 - P_{F,i})^{1-u_i}$, $\mathbf{u}_k = \text{bin}(k, N)$, and $\text{bin}(k, N)$ is the binary vector representation of k using N bits, *e.g.*, $[0, 0, 1, 1] = \text{bin}(3, 4)$. Substituting (177) into (181), using $r_j = (2u_j - 1)\sqrt{\rho} g_j + n_j$, and observing that $\Pr(\cdot \,|\mathbf{u} = \mathbf{u}_k, H_0)$ in (181) is independent of the underlying hypothesis H_0, we can express (181) as

$$P_{F,0} = \sum_{k=0}^{2^N-1} \Pr(\mathbf{u} = \mathbf{u}_k | H_0) \Pr\left(\sum_{j=1}^{N} \Upsilon_{u_j}^*(n_j') > t_0^* | \mathbf{u} = \mathbf{u}_k\right) \tag{182}$$

where

$$\Upsilon_{u_j}^*(n_j') = \log \frac{P_{D,j} e^{(2u_j-1)4\rho|g_j|^2 + 4\sqrt{\rho} n_j'} + 1 - P_{D,j}}{P_{F,j} e^{(2u_j-1)4\rho|g_j|^2 + 4\sqrt{\rho} n_j'} + 1 - P_{F,j}},$$

with $n_j' \sim \mathcal{C}(0, |g_j|^2/2)$[2]. Let $\Gamma_{\mathbf{u}_k}^* = \sum_{j=1}^{N} \Upsilon_{u_j}^*(n_j')$, and denote the cumulative distribution of $\Gamma_{\mathbf{u}_k}^*$ by $F_{\Gamma_{\mathbf{u}_k}^*}(\cdot)$. Then we can rewrite (182) as

$$P_{F,0} = 1 - \sum_{k=0}^{2^N-1} \Pr(\mathbf{u} = \mathbf{u}_k | H_0) F_{\Gamma_{\mathbf{u}_k}^*}(t_0^*). \tag{183}$$

We can show that the probability density of $\Upsilon_{u_j}^*(n_j')$ is given by (161) with $k = u_j$. In a similar fashion, the cumulative distribution can be obtained as

$$F_{\Upsilon_{u_j}^*}(a) = \begin{cases} 0 & a < \log \frac{1-P_{D,j}}{1-P_{F,j}} \\ Q\left(\left((2u_j - 1)4\rho|g_j|^2 - \frac{(1-P_{F,j})e^a - (1-P_{D,j})}{P_{D,j} - P_{F,j}e^a}\right)/4|g_j|\sqrt{\rho}\right) & \log \frac{1-P_{D,j}}{1-P_{F,j}} < a < \log \frac{P_{D,j}}{P_{F,j}} \\ 1 & a > \log \frac{P_{D,j}}{P_{F,j}} \end{cases} \tag{184}$$

[2] $\mathcal{C}(0, \sigma^2)$ denotes the zero-mean Gaussian distribution with variance σ^2

Hence, using (161), (184), and the fact that $\Upsilon^*_{u_j}(n'_j)$ are independent for $j = 1, \cdots, N$, we obtain $F_{\Upsilon^*_{u_j}}(.)$ as

$$F_{\Gamma^*_{\mathbf{u}_k}}(t_0^*) = \int_{A_{N-1}}^{B_{N-1}} \cdots \int_{A_1}^{B_1} f_{\Upsilon^*_{u_1}}(y_1) \prod_{j=2}^{N-1} f_{\Upsilon^*_{u_1}}(y_j - y_{j-1}) F_{\Upsilon^*_{u_N}}(t_0^* - y_{N-1}) dy_1 \cdots dy_{N-1} \tag{185}$$

where $A_k = \sum_{j=1}^{k} \log \frac{1-P_{D,j}}{1-P_{F,j}}$ and $B_k = \sum_{j=1}^{k} \log \frac{P_{D,j}}{P_{F,j}}$. Using (185) and (183), with the knowledge of the detection and false alarm probabilities at the local sensors, we can finally evaluate the global false alarm probability, $P_{F,0}$. Similarly, the detection probability can be computed as

$$P_{D,0} = 1 - \sum_{k=0}^{2^N-1} \Pr(\mathbf{u} = \mathbf{u}_k | H_1) F_{\Gamma^*_{\mathbf{u}_k}}(t_0^*). \tag{186}$$

where $\Pr(\mathbf{u} = \mathbf{u}_k | H_1) = \prod_{i=1, \mathbf{u}=\mathbf{u}_k}^{N} P_{D,i}^{u_i} (1 - P_{D,i})^{1-u_i}$.

Although the fusion center uses a likelihood ratio test, determining the actual parameters for the best detection is a formidable task. This is because, according to NP-lemma, one needs to make an exhaustive search over all $P_{F,i}$ and $P_{D,i}$ to determine t_0^* so that for some $P_{F,0} \leq \alpha$, $P_{D,0}$ is maximized. Since analytical solution is not tractable, the optimal solution can be found for only small values of N. From (183) and (186) it is clear that the complexity increases with N.

7.2.2.2 Decision Rule Design for Parallel Detection

So far, we have not specified the detection rule at the local sensors. In the absence of errors in the transmissions from local sensors to the fusion center, it can be shown that the optimal tests at the local sensors are also LR tests. If the links are subject to fading and additive noise, while the optimal fusion rule according to NP-lemma reduces to a likelihood-ratio test at the fusion center, the optimal tests at the local sensors are not easy to derive. However, one still needs some detection rule, and so, without claiming any optimality, we propose the use of LR test also at the local sensors. The numerical examples in Section 7.6 will justify that the detection in this fashion performs remarkably well.

The LR test at the local sensors is given by

$$\Lambda^*(y_j) = \log \frac{p(y_j|H_1)}{p(y_j|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_j^* \tag{187}$$

Assume that the distribution of the observations y_j is known, and let $F_i(\cdot)$ denote the cumulative distribution of $\Lambda(y_j)^*$ under the hypothesis $H_i,\ i = 0, 1$. Then, we can express the false alarm and detection probabilities at S_j as $P_{F,j} = 1 - F_0(t_j^*)$ and $P_{D,j} = 1 - F_1(t_j^*)$, respectively. Hence, the design of the detection rules reduces to the determination of the thresholds t_j^* that maximize the $P_{D,0}$ for some fixed $P_{F,0}$.

As usual in distributed detection problems, an analytical solution is not feasible; therefore, we resort to numerical search procedures to determine the decision fusion rules. A simplistic approach is to set the thresholds at all sensor nodes the same, however, in that case, one can not guarantee the maximization of the global detection probability.

7.3 Distributed Detection with Multiple-Bit Decisions Transmitted over Fading Channels

Thus far, we allowed transmission of only one bit per sensor to the fusion center. In fact, making single-bit decisions is optimal under certain scenarios whenever they can be available at the fusion center without any error [94]. Nevertheless, under fading and noisy channel assumptions, making multiple-bit decisions and sending more than one bit to the fusion center, although sacrificing from bandwidth and power, might significantly improve the detection probability. In this section, we will investigate the analysis and performance of multiple-bit sensor decisions. We first discuss the two-bit decisions and then generalize the scheme to the multiple-bit decision rules. *The ultimate goal is to determine the best (possibly) multiple-bit decision rules that provide the optimal power/performance tradeoff.*

7.3.1 Local Decisions with two bits

Consider the parallel configuration in Figure 43. When local sensors select one of the hypotheses using Neyman-Pearson lemma, each of them compares the likelihood ratio with a single threshold and makes a positive or negative decision which is the only statistics

transferred to fusion center. This is similar to a two-level quantization problem. Increasing the number of levels at local sensors certainly improves the performance at the fusion center because it provides more information about the likelihood ratio. Assume now that we have two more regions to make the decision in addition to those originally used in single-bit decision. Then we generate two bits using the following likelihood ratio test

$$\Lambda_i(y_i) = \frac{p(y_i|H_1)}{p(y_i|H_0)} \in \begin{cases} (\tau_i^{(1)}, \infty) & \text{decide } H_1, \text{ or set } u_i = 11 \\ (\tau_i^{(2)}, \tau_i^{(1)}] & \text{set } u_i = 10 \\ (\tau_i^{(3)}, \tau_i^{(2)}] & \text{set } u_i = 01 \\ (-\infty, \tau_i^{(3)}] & \text{decide } H_0, \text{ or set } u_i = 00 \end{cases} \tag{188}$$

That is, we make a decision only if the likelihood ratio of the observed signal is sufficiently large (or small). If the likelihood ratio is within the interval $(\tau_i^{(3)}, \tau_i^{(1)})$, then we avoid making a certain decision and, instead, we inform the fusion center with the most likely decision. For example, when the likelihood ratio is within the interval $(\tau_i^{(2)}, \tau_i^{(1)})$, then H_1 is more likely to be observed but the value of the likelihood ratio is not in the range for making a clear choice. This uncertainty is delivered to the fusion center by sending two bits (10). Similarly, if $\Lambda_i(y_i) \in (\tau_i^{(3)}, \tau_i^{(2)}]$, H_0 is more likely but with some uncertainty and so we send (01) to the fusion center. This approach certainly prevents the propagation of ambiguous or less reliable decisions to the fusion center.

The received signal vector $\mathbf{r}$, with binary modulation, is a $2 \times N$ matrix whose entries are independently and identically distributed conditioned on the channel state information. Its conditional density given H_1 is given by

$$p(\mathbf{r}|H_1, \mathbf{h}) = \prod_{i=1}^{N} p(r_{i0}, r_{i1}|H_1, h_{i0}, h_{i1}) \tag{189}$$

$$= \prod_{i=1}^{N} \big[P_{D,i} p_n(r_{i0} - h_{i0}s_i^1) p_n(r_{i1} - h_{i1}s_i^1) + P_{M,i} p_n(r_{i0} - h_{i0}s_i^0) p_n(r_{i1} - h_{i1}s_i^0) + P_{10,i} p_n(r_{i0} - h_{i0}s_i^1) p_n(r_{i1} - h_{i1}s_i^0) + P_{01,i} p_n(r_{i0} - h_{i0}s_i^0) p_n(r_{i1} - h_{i1}s_i^1)\big] \tag{190}$$

A similar expression can be written for the pdf under H_0. The likelihood ratio is given by

$$\Lambda(\mathbf{r}|\mathbf{h}) = \frac{p(\mathbf{r}|H_1, \mathbf{h})}{p(\mathbf{r}|H_0, \mathbf{h})} \tag{191}$$

Taking the logarithm of both sides, the right hand side of (191) can be expressed as the sum of log-likelihood ratios due to the received signal from individual local sensors, and hence, it is possible to obtain the distribution function of the global likelihood ratio test, whereby we determine the global detection and false alarm probabilities. We can then utilize the Neyman-Pearson lemma to make a search for the optimal thresholds. The search space for multiple-bit sensor decision is larger than that for single-bit decision; nevertheless, the detection probability can be improved significantly as shown by the simulation results in Section 7.6.

It is also possible to use a higher order modulation scheme to minimize the delay. For the example above, two-bit decisions can be transmitted by a 4-QAM (Quadrature Amplitude Modulation) or 4-PSK modulation scheme. The analysis is very similar to the case of single-bit decisions transmitted by BPSK modulation.

7.3.2 A General Multi-bit Decision Method

The above technique can readily be generalized to other cases. In the sequel, we study a more general problem to find optimal number of the decision bits (i.e., number of quantization levels) at each sensor. Assume, for example, the k^{th} sensor node generates a b_k-bit decision and employs 2^{b_k}-ary modulation to transmit its decision to the fusion center. Let $\tau_1^{(k)} < \tau_2^{(k)} \ldots < \tau_{2^{b_k}-1}$ be thresholds at node k, and let the observation samples, y_k, $k = 1, \ldots, N$ are confined to the interval $[\tau_{\min}, \tau_{\max}]$. The signal $s_l^{(k)}$ is transmitted whenever $y_k \in (\tau_{l-1}^{(k)}, \tau_l^{(k)}]$, $l = 1, \ldots, 2^{b_k}$. Note that $\tau_0^{(k)} = \tau_{\min}$ and $\tau_{2^{b_k}}^{(k)} = \tau_{\max}$. For signals that take any value in the real line, we have $\tau_{\min} = -\infty$ and $\tau_{\max} = \infty$.

The likelihood ratio test can be expressed as

$$\Lambda(\mathbf{r}) = \sum_{k=1}^{N} \log \frac{\sum_{l=1}^{2^{b_k}} P_{k,l}^{(1)} p_n(r_k - \sqrt{\rho} g_k s_l^{(k)})}{\sum_{l=1}^{2^{b_k}} P_{k,l}^{(0)} p_n(r_k - \sqrt{\rho} g_k s_l^{(k)})} \tag{192}$$

where $P_{k,l}^{(m)} = \Pr(y_k \in (\tau_{l-1}^{(k)}, \tau_l^{(k)}] | H_m)$. The false alarm and detection rates follow as

$$P_{F,0} = \Pr(\Gamma(\mathbf{r}) > t_0^* | H_0) \tag{193}$$

$$P_{D,0} = \Pr(\Gamma(\mathbf{r}) > t_0^* | H_1). \tag{194}$$

Finally, we can write the optimization problem for the best multiple bit decision rules as

$$\begin{aligned} \max \quad & P_{D,0} \\ \text{s.t.} \quad & P_{F,0} \leq \alpha \\ & \sum_{k=1}^{N} b_k = N_{total} \\ & \tau_{\min} \leq \tau_l^{(k)} \leq \tau_{\max},\ k = 1, \ldots, N, l = 1,\ \ldots, 2^{b_k}. \end{aligned} \tag{195}$$

The optimization in (195) is not a convex problem over the set of thresholds and the number of decision bits. Therefore, it is very difficult to solve the problem using standard optimization tools. Instead, one can employ exhaustive search methods to determine the optimal number of bits and the thresholds, which is a formidable task. A suboptimal method is to restrict the search over only the set of b_k values satisfying the total bid budget constraint and using locally most optimal decision thresholds at each node for that b_k value. In Section 7.6, we will present some examples to illustrate the performance of distributed detection with this approach.

7.4 Distributed Detection with Analog Data Gathering

In the previous sections, we investigated the optimal threshold design in Neyman-Pearson sense, and show that the search for the optimal designs does not allow for tractable solutions even in the case of conditionally independent observations. In this section, we will follow a promising approach that (i) incorporates the channel statistics with the detection scheme and (ii) requires much less complexity than those that solely rely on the costly error correction mechanisms for improved reliability. We propose a distributed detection technique based on analog transmission of the local observations (or the sufficient statistics in case of multiple observations) to the fusion center. In this method, each sensor node first generates a sufficient statistics, and then employs a linear analog modulation scheme, such as Double Side Band (DSB), to transmit its real-valued data to the fusion center. That is, the local data is simply amplified-and-forwarded; there is no local processing, quantization or coding performed at the local sensor. We will refer to this method as *analog approach*. This approach has several advantages over its counterpart the *digital approach* where a quantized

version (local decisions consisting of a few bits) is forwarded to the fusion center. In the digital approach, one needs to determine optimal decision rules for both the local sensors and the fusion center, which is computationally impractical even for a network with a few sensor nodes. However, the proposed detection scheme requires only the determination of a single decision rule at the fusion center, and possibly the computation of the optimal power allocation gains which can be pursued either analytically or numerically with much less complexity than that required for the digital approach. Another advantage of the analog approach is that it can readily be employed for cases where there exists correlation between the sensor observations.

7.4.1 Analog Signaling

We consider the parallel topology shown in Figure 43. The sensor node S_k employs an amplify-and-forward technique to transmit its statistics y_k; there is no quantization or coding on y_k. We consider a linear analog communication scheme such as double side-band (DSB)[3] modulation where the transmitted signal at S_k is given by [59]

$$x_k(t) = \sqrt{\rho_k} y_k \cos(2\pi f_{c_k} t),\ 0 \leq t \leq T_s$$

where ρ_k is the power gain, g_k is the channel gain and f_{c_k} is the carrier frequency, and the transmit power is $P = \rho_k \bar{y_k^2}$. The received signal (after coherent detection) can be expressed as

$$r_k = \sqrt{\rho_k}\sqrt{g_k} y_k + n_k$$

where n_k is the zero mean additive white Gaussian noise (AWGN) with variance ξ_k^2. Letting $\mathbf{y} = [y_1, \ldots, y_K]$, $\mathbf{A} = \text{diag}\{\sqrt{\rho_1}, \ldots, \sqrt{\rho_K}\}$, $\mathbf{G} = \text{diag}\{\sqrt{g_1}, \ldots, \sqrt{g_K}\}$, $\mathbf{n} = [n_1, \ldots, n_K]$, and $\mathbf{r} = [r_1, \ldots, r_K]$, we equivalently have

$$\mathbf{r} = \mathbf{AGy} + \mathbf{n}$$

Note that the channel noise covariance matrix is given by $\mathbf{C_n} = \text{diag}\{\xi_1^2, \ldots, \xi_K^2\}$. We assume that, conditioned on H_k, $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_k, \mathbf{C}_{\mathbf{y},k})$ with $\boldsymbol{\mu}_k = [\mu_{k,1}, \ldots, \mu_{k,K}]$. With the signal model above, we have $\mathbf{r} \sim \mathcal{N}(\mathbf{D}\boldsymbol{\mu}_k, \mathbf{D}\mathbf{C}_{\mathbf{y},k}\mathbf{D}^T + \mathbf{C_n})$ where $\mathbf{D} = \mathbf{AG}$.

[3] One can also employ a single-side band modulation which will be more efficient in terms of bandwidth expansion. The related analysis for this case follows similarly.

Not that according to this analog scheme, one does not need to define a local decision criterion and leaves the final decision to the fusion center that will make a decision based on the corrupted version of local statistics.

7.4.2 Neyman-Pearson Detection

The optimal detection under Neyman-Pearson criterion reduces to the likelihood ratio test given by

$$\Gamma(\mathbf{r}) = \frac{f(\mathbf{r}|H_1, \mathbf{A}, \mathbf{G})}{f(\mathbf{r}|H_0, \mathbf{A}, \mathbf{G})} \underset{H_0}{\overset{H_1}{\gtrless}} t_0. \tag{196}$$

where $f(\mathbf{v}) = (2\pi|\mathbf{\Sigma}|)^{-1/2} \exp(-\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{v}-\boldsymbol{\mu}))$ is the multivariate Gaussian density function with $\boldsymbol{\mu}$ the mean vector and $\mathbf{\Sigma}$ the covariance matrix. Substituting $f(\cdot)$ into (196) and after some manipulations, we have[4]

$$\Gamma^*(\mathbf{r}) = -\frac{1}{2}\mathbf{r}^T(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_0^{-1})\mathbf{r} + \mathbf{r}^T(\mathbf{\Sigma}_1^{-1}\mathbf{D}\boldsymbol{\mu}_1 - \mathbf{\Sigma}_0^{-1}\mathbf{D}\boldsymbol{\mu}_0) - \frac{1}{2}(\boldsymbol{\mu}_1^T\mathbf{D}^T\mathbf{\Sigma}_1^{-1}\mathbf{D}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\mathbf{D}\mathbf{\Sigma}_0^{-1}\mathbf{D}\boldsymbol{\mu}_0) \underset{H_0}{\overset{H_1}{\gtrless}} t_0^* \tag{197}$$

where $t_0^* = \log t_0 + \frac{1}{2}\log\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|}$, and $\mathbf{\Sigma}_k = \mathbf{AGC}_{\mathbf{y},k}\mathbf{G}A + \mathbf{C}_\mathbf{n}$. From (197), the false alarm and detection probabilities follows as

$$P_{F,0} = \Pr(\Gamma^*(\mathbf{r}) > t_0^*|H_0) \tag{198}$$

and

$$P_{D,0} = \Pr(\Gamma^*(\mathbf{r}) > t_0^*|H_1), \tag{199}$$

respectively. We note that $\Gamma^*(\mathbf{r})$ assumes a quadratic Gaussian form and its density can be evaluated using the probability theory [5]. However, the resulting expressions are quite lengthy and do not provide useful insight. Rather, we consider a simpler case where $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_0 = \mathbf{\Sigma}_o$, i.e., the covariance matrix of $\mathbf{y}$ are identical under both hypothesis, *e.g.*, $\mathbf{C}_{\mathbf{y},1} = \mathbf{C}_{\mathbf{y},0} = \mathbf{C}_o$. Thus, the quadratic term in (197) disappears and $\Gamma^*(\mathbf{r})$ reduces to a linear

[4]It is convenient to express likelihood ratios in logarithmic form

combination of Gaussian random variables. The mean and the variance of $\Gamma^*(\cdot)$ conditioned on H_k is then given by

$$\mu_{\Gamma^*,k} = (-1)^k \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{D}^T \boldsymbol{\Sigma}_o^{-1} \mathbf{D}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \tag{200}$$

$$\sigma^2_{\Gamma^*,k} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{D}^T \boldsymbol{\Sigma}_o^{-1} \mathbf{D}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \tag{201}$$

Finally, the false alarm and detection probabilities in (198) and (199) boils down to

$$P_{F,0} = Q\left(\frac{t_0^* - \mu_{\Gamma^*,0}}{\sigma_{\Gamma^*,0}}\right) \tag{202}$$

and

$$P_{D,0} = Q\left(\frac{t_0^* - \mu_{\Gamma^*,1}}{\sigma_{\Gamma^*,1}}\right), \tag{203}$$

respectively, where $Q(x) = 1/\sqrt{2\pi}\int_x^\infty e^{-t^2/2}dt$. According to NP-lemma, the threshold t_0^* can be determined for any prespecified size α, and with this threshold, the detection probability (a.k.a. power) is given by

$$P_{D,0} = Q\left(Q^{-1}(\alpha) - \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{D}\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}\right) \tag{204}$$

From (204), the receiver operating characteristics (ROC) curves for the detection scheme can readily be obtained for a fixed power allocation scheme $\mathbf{A}$ and channel gain matrix $\mathbf{G}$. Also note that expressions on the size and the power of the proposed detection scheme reduce to those for the centralized detection if we set $\mathbf{D} = \mathbf{I}_K$ and $\mathbf{C_n} = \mathbf{0}_K$.

7.4.3 Optimal Power Allocation

Observe that the detection rate attained by the likelihood ratio test depends on the power allocation gains $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K]$, which implies that one can obtain larger $P_{D,0}$ by suitable power allocation. In this section, we solve for the optimal power gains that attain the maximal $P_{D,0}$ for a given total power budget, P_T. Recall that the average transmit power at node k is given by $\rho_k \bar{y_k^2}$ where, for equally likely hypotheses, we have $\bar{y_k^2} = [\mathbf{C}_o]_{k,k} + (\mu_{1,k}^2 + \mu_{0,k}^2)/2$ with $[\mathbf{C}_o]_{k,k}$ denoting the k^{th} diagonal entry of $\mathbf{C}_o$. Let $P_y = \max\{\bar{y_1^2} \ldots, \bar{y_K^2}\}$. Hence, the constraint on average of the total power transmitted from all K sensors is given by

$$\sum_{k=1}^K \rho_k \leq P_T/P_y.$$

Note that $Q(x)$ is a monotonically decreasing function, therefore, to maximize $P_{D,0}$ we need to maximize $(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{D}\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)$, or minimize $-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{D}\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)$. Therefore, the optimization problem can be cast as

$$\begin{aligned} \min\ & -(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{D}\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) \\ \text{s.t.}\quad & \textstyle\sum_{k=1}^K \rho_k \le P_T/P_y. \end{aligned} \tag{205}$$

which can be simplified to

$$\begin{aligned} \min\ & -(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T(\mathbf{C}_o+\boldsymbol{\Lambda}^{-1})^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) \\ \text{s.t.}\quad & \textstyle\sum_{k=1}^K \rho_k \le P_T/P_y. \end{aligned} \tag{206}$$

where $\boldsymbol{\Lambda}=\text{diag}\{\rho_1 g_1/\xi_1^2,\ldots,\rho_K g_K/\xi_K^2\}$. This optimization is convex over $\rho_1,\ldots,\rho_K$. Although a closed form solution to the problem may not be analytically feasible for a general $\mathbf{C}_o$, one can use interior point methods or gradient-based methods to solve for the optimal power gain set [23]. We will illustrate several numerical examples in Section 7.6 for this case.

A closed form solution exists when the observations at the local sensor nodes are uncorrelated in which case the covariance matrix $\mathbf{C}_o$ is diagonal, i.e., $\mathbf{C}_o=\text{diag}\{\sigma_1^2,\ldots,\sigma_K^2\}$. In this case, the optimization in (206) reduces to

$$\begin{aligned} \min\ & -\textstyle\sum_{k=1}^K \Delta_k^2 \frac{\rho_k g_k}{\rho_k g_k \sigma_k^2+\xi_k^2} \\ \text{s.t.}\ & \textstyle\sum_{k=1}^K \rho_k \le P_T/P_y. \end{aligned} \tag{207}$$

for which the Lagrangian cost function follows as

$$J(\lambda,\boldsymbol{\rho}) = -\sum_{k=1}^K \Delta_k^2 \frac{\rho_k g_k}{\rho_k g_k \sigma_k^2+\xi_k^2} + \lambda\left(\sum_{k=1}^K \rho_k - P_T/P_y\right) \tag{208}$$

and $\Delta_k=\mu_{1,k}-\mu_{0,k}$. The KKT conditions are given by

$$\begin{aligned} -\Delta_k^2 \frac{g_k\xi_k^2}{(\rho_k g_k\sigma_k^2+\xi_k^2)^2} + \lambda &= 0 \\ \sum_{k=1}^K \rho_k &= P_T/P_y \end{aligned} \tag{209}$$

Without loss of generality, assume that $g_1\Delta_1^2/\xi_1^2 > g_2\Delta_2^2/\xi_2^2 > \ldots > g_N\Delta_N^2/\xi_N^2$, and define

$$a(L) = \frac{\sqrt{g_L}\Delta_L}{\xi_L} \frac{P_T/P_y + \sum_{k=1}^{L} \frac{\xi_k^2}{g_k\sigma_k^2}}{\sum_{k=1}^{L} \frac{\Delta_k\xi_k}{\sqrt{g_k}\sigma_k^2}} \tag{210}$$

Next determine the unique L_1 such that $a(L_1) > 1$ and $a(L_1+1) < 1$. If $a(k) \geq 1$ for $\forall k,\ k = 1, \ldots, K$, then set $L_1 = K$. By the KKT conditions in (209), we can finally write the optimal power gains as

$$\rho_j^{\text{opt}} = \begin{cases} \frac{\Delta_j\xi_j}{\sqrt{g_j}\sigma_j^2} \frac{P_T/P_y + \sum_{k=1}^{L_1} \frac{\xi_k^2}{g_k\sigma_k^2}}{\sum_{k=1}^{L_1} \frac{\Delta_k\xi_k}{g_k\sigma_k^2}} & j = 1, \ldots, L_1 \\ 0 & j = L_1+1, \ldots, K \end{cases} \tag{211}$$

From (211), the optimal power gains are determined in two steps: First, enumerate the sensor nodes in a descending order with respect to their channel signal-no-noise ratios, and then determine L_1 which, in effect, specifies a signal-to-noise ratio threshold. Second, set the power gains ρ_j as in (211). It is seen that according to the optimal power allocation, some of the sensors are not allowed to transmit if the signal-to-noise ratio for that node is below some threshold.

We note that instead of performing optimal power allocation, one can employ suboptimal methods to distribute the power, *e.g.*, using identical power gains at all sensor nodes; though one does not guarantee the best detection performance with such schemes. However, as shown by the numerical examples, even with uniform power allocation, one can attain detection performance comparable to that obtained with the optimal power gains. Furthermore, at high signal-to-noise ratios, the detection rate given by (204) can be shown to converge to

$$P_{D,0} \approx Q\left(Q^{-1}(\alpha) - \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{C}_o^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}\right) \tag{212}$$

which is exactly the detection rate attained by the centralized detector.

7.5 Sequential Distributed Detection

The distributed detection procedures discussed so far operate with a fixed number of observations, and based on that the decision rules are predetermined, *e.g.*, fixed sample size detector. Alternatively, in sequential detection, the number of observations for reaching

the decision is random and they are based on the observations themselves [91]. We illustrate the sequential method with a simple example. Consider the binary hypothesis testing problem for which an infinite sequence of iid random observations $Y_1, Y_2, \ldots$, can be generated. After taking the i^{th} observation, a sequential test is performed to find the probability $\pi = P(Y_1, \cdots, Y_i),\ i = 1, 2, \ldots$, and stop taking observations as soon as a desired clarity level is achieved. It is desirable to minimize the number of observations to reduce the cost of taking the observations and the delay for the final decision. Hence, we wish to achieve an optimal trade-off between the cost of observations and decisions quality.

The idea of sequential detection can be deployed in the area of distributed detection [88, 112,132,177,178]. In this context, the sequential test can be made either at the fusion center based on the decisions received from the local sensors [88,177], or at the local sensors that act as a team using a coupled objective function without interaction of a fusion center [40,178]. In this section, we will primarily investigate the former one, but in cases where the links between the local sensors and the fusion center are subject to multipath fading. Contrary to the previous work on distributed sequential detection that assumed error-free links, we incorporate the effect of fading and noise in the detection process and determine how the average number of required sequential tests and the detection capability is effected. We assume that (i) the observations are conditionally independent from sensor to sensor, and they are also iid at each sensor, (ii) local sensors are not allowed to communicate with each other, and (iii) there is not feedback from the fusion center except for the "stop" signal to inform the local sensors that a decision is made. The second assumption entails that the decision at each sensor will depend only on its own observations. Second and third assumptions together imply that local sensors may perform a sequential-like detection until the "stop" signal is heard. Based on the available information pattern at each sensor, the following scenarios are considered:

1. At time i, n^{th} local sensor has access to observations $\mathbf{y}_{n,i} = (y_{n,1}, y_{n,2}, \ldots, y_{n,i})$, and the fusion center has access to the received signals $(r_{1,1}, r_{1,2}, \ldots, r_{1,i}, \ldots, r_{N,1}, \ldots, r_{N,i})$ and the associated channel gains $\mathbf{h} = (h_{1,1}, \cdots, h_{N,i})$.

2. At time i, n^{th} local sensor has access only to the current observation $y_{n,i}$ and its previous decisions, $(u_{n,1}, \ldots, u_{n,i-1})$, while the fusion center has access to the same information as in *Case 1*.

3. At time i, n^{th} local sensor has access to the same information as in *Case 2*, while the fusion center has access only to the current received signal, $(r_{1,i}, \ldots, r_{N,i})$, and the associated channel gains. As a simple modification, we will also study the case when previous global decisions are also available.

While the idea of sequential processing in the distributed detection framework best suits to the parallel configuration, it can also be applied in the serial and hierarchical configurations. We leave the analysis for these configurations as future research.

We now describe a sequential detection method for channel-based distributed detection in a parallel network. We study Case 1 since this scheme intuitively outperforms the others, and the other cases are special cases of Case 1. The sequential-like decision to be performed at n^{th} sensor at time i uses the likelihood ratio test

$$\Lambda_{n,i}(\mathbf{y}_{n,i}) = \frac{p(y_{n,1}, \ldots, y_{n,i}|H_1)}{p(y_{n,1}, \ldots, y_{n,i}|H_0)} = \prod_{k=1}^{i} \frac{p(y_{n,k}|H_1)}{p(y_{n,k}|H_0)} \tag{213}$$

$$\begin{aligned} &> \tau_i \text{ decide } H_1, \text{ set } u_{ni} = 1 \\ &= \tau_i \text{ decide } H_1 \text{ with probability } \varepsilon_i \\ &< \tau_i \text{ decide } H_0 \text{ ,set } u_{ni} = 0 \end{aligned} \tag{214}$$

and the sequential test at the fusion center uses the probability

$$\pi_i = P(H_1|r_{11}, \ldots, r_{Ni}, \mathbf{h}) \tag{215}$$

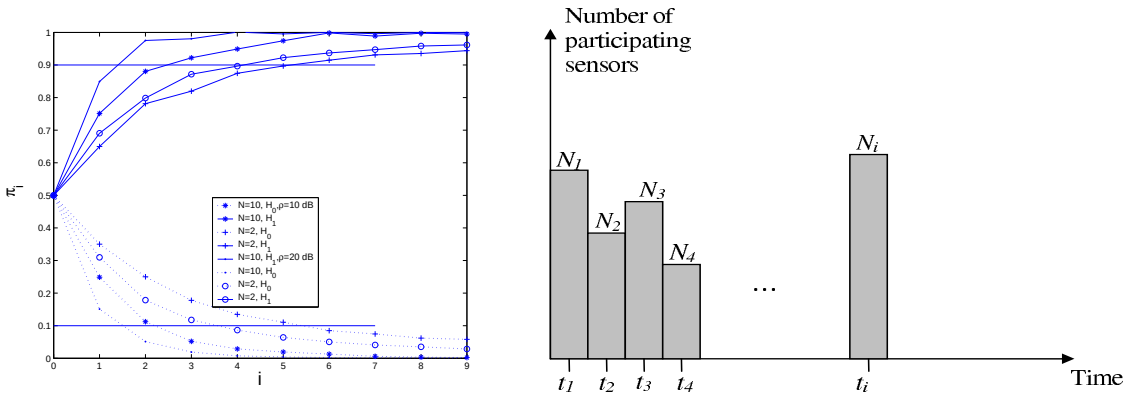
which is computed iteratively until νth step where $\nu = \min_i\{\pi_i : \pi_i \notin (\pi_L, \pi_U)\}$, and $0 \leq \pi_L < \pi_U \leq 1$ are two thresholds determined by the distributions of the observations. If $\pi_\nu \geq \pi_U$, H_1 is chosen, otherwise, H_0 is chosen. Using the Bayes' rule, the iterations can be expressed as

$$\pi_i = \frac{L_i \pi_{i-1}}{L_i \pi_{i-1} + (1 - \pi_{i-1})} \tag{216}$$

where $L_i = \prod_{n=1}^{N} L^{(n)}(r_{ni}|\pi_{i-1}, h_{ni})$ with $L^{(n)}(r_{ni}|\pi_{i-1}, h_{ni})$ being the conditional likelihood ratio between H_0 and H_1 based on r_{ni}. π_0 is set to the prior probability of H_1. From this result, we arrive at the conclusion that for the optimal sequential distributed detection, each

local sensor node performs a likelihood ratio test using (214) and the fusion center performs a sequential test using (216). The thresholds at each sensor and at the local sensor can be determined through exhaustive search, which becomes very complicated because of the consideration of fading in the decision rule. The determination of thresholds at each sensor depending on the channel realizations $\mathbf{h}$ is prohibitively complicated for the sensor nodes. We can pursue *robust design* method for these thresholds in which the channel realizations will not be used directly. For example, for Rayleigh fading, we select the average magnitude of the channel to compute all thresholds.

Figure 44(a) illustrates a sequential test for detecting DC-level in additive Gaussian noise. A parallel distributed detection network having $N = 2$ and $N = 10$ local sensors is employed. The detection and false alarm rates are assumed constant at 0.6915 and 0.3085, respectively, at each local sensor at any instant. The posterior probabilities are averaged over the channel statistics: $\pi_{i,ave} = E_{\mathbf{g}}(\pi_i)$. The dashed lines depict π_i under H_0, and the solid lines depict π_i under H_1. We observe that, at 10 dB, when the thresholds are $\pi_U = 0.9$ and $\pi_L = 0.1$, we need 6 iterations to make the decision with $N = 2$ local sensors, while we need only 3 iterations with $N = 10$ local sensors. As SNR is increased, we need less iterations to reach a decision, *e.g.*, only 2 iterations is sufficient for $N = 10$ at 20 dB. Based on the initial experiments, channel-based sequential distributed detection for wireless sensor networks is very promising in managing performance/energy tradeoff.



(a) (b)

Figure 44: (a) Posterior probabilities π_i sequentially computed using (216) when $\rho = 10, 20$ dB with $N = 2, 10$ local sensors. $P_{D,ni} = 0.6915$ and $P_{FA,ni} = 0.3085$ at n^{th} local sensor at time i; (b) Time division in sequential distributed detection.

In the above discussion, all N sensors send their observations or decisions every time unit until a satisfactory threshold is achieved. In other words, at every time unit, all N observations/decisions are transmitted to the fusion center. In wireless sensor networks, N can be in the order of hundreds of thousands. Nevertheless, not all N observations/decisions are needed to meet the decision threshold. Therefore, in the above design, there will be redundant information that unnecessarily consumes energy at the local sensors as well as at the fusion center. To overcome this problem, we propose a time-division technique where not all sensors transmit their observations/decisions every time unit (See Figure 44(b)). Instead, N_1 sensors send their decisions at time unit t_1, other N_2 sensors send their decisions at time unit t_2, other N_3 sensors send their decisions at time unit t_3, ..., other N_i sensors send their decisions at time unit t_i, where

$$\sum_{j=1}^{i} N_j = N. \tag{217}$$

The challenge is to find the optimum $[N_1, N_2, ..., N_i]$ vector that maximizes the detection probability while minimizing the consumed energy given the fading channels between the sensors and the fusion center. As an example, assume we have 1000 sensors. If we know beforehand that 6500 observations are needed before the fusion center reaches a satisfactory decision. Then, if all 1000 sensor send every time unit, then there will be 500 redundant observations, which implies 500 unnecessary transmissions and 500 unnecessary processing tasks. Instead, if we subgroup the sensors into groups of 200 sensors each, then there will be 100 redundant observations/transmissions/processing. Similarly, if we subgroup the sensors into groups with 100 sensors in each, then there will be no redundant observations/transmissions/processing. Nevertheless, the latter case requires 65 time units while the former requires 33 time units to reach the satisfactory decision. This compromise of energy-performance and delay is a key problem that will be investigated as part of the future work.

7.6 Simulation Results

We devote this section to present several numerical examples for the analysis performed in the previous sections. We illustrate the performance of different detection strategies

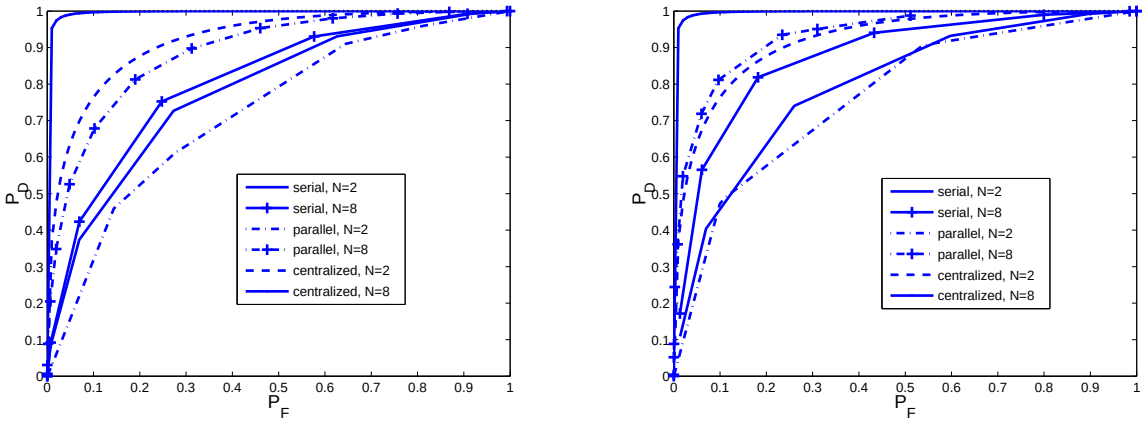
and compare their performance. In the following examples, we consider the detection of a DC-level signal in additive white Gaussian noise:

$$y_k = \begin{cases} m + \nu_k & H_1 \\ \nu_k & H_0 \end{cases} \tag{218}$$

where $\nu_k \sim \mathcal{N}(0, \sigma_k^2)$ is iid for $k = 1, \ldots, K$, and hence $\mathbf{C}_o = \text{diag}\{\sigma_1^2, \ldots, \sigma_K^2\}$. Each sensor node observes a signal that is either noise or noise plus some DC-level and transmits their summary on the observation to the fusion center.

We first illustrate the performance of the serial and parallel detection through numerical simulations (see Figure 45). We consider the case of single bit decision. Figure 45.a and Figure 45.b depict the receiver operating characteristics (ROC) for $\rho = 1$ and $\rho = 3$, respectively. The sensor nodes use BPSK modulated signals ± 1 to transmit their decisions. For the serial detection, by using the expressions (*e.g.*, (163) and (164)) developed in Section 7.2.2.1, we obtain the best detection probability $P_{D,N}$ for a given $P_{F,N}$ by an exhaustive search over $P_{F,j}$, $j = 1, \cdots, N-1$. We used the numerical integration routine QUADL in MATLAB©. In the optimization process, we observed that the best $P_{D,N}$ can always be achieved when the threshold is within the interval defined by the sufficiency condition in (174). For the parallel scheme, identical thresholds are assumed at local sensors. For the case of Rayleigh fading, we compute the average probabilities of the false alarm and detection. We observe that when $N = 2$, the serial fusion structure achieves slightly better detection performance than the parallel fusion does for both values of $\rho = 0$ and $\rho = 3$. When the number of sensors is increased to $N = 8$, it is seen that the parallel distributed detection is superior to the serial one. For all cases in this figure, we also observe that the performance degradation due to the noise channel is significant with respect to the performance of the centralized detection which assumes the availability of noise-free observations at the fusion center.

Next, we illustrate the optimal threshold design for the parallel decision fusion. We assume that the local sensors employ LR tests to perform detection and use BPSK modulated signals ± 1 to transmit their decisions. By using the expressions (*e.g.*, (183)– (186))

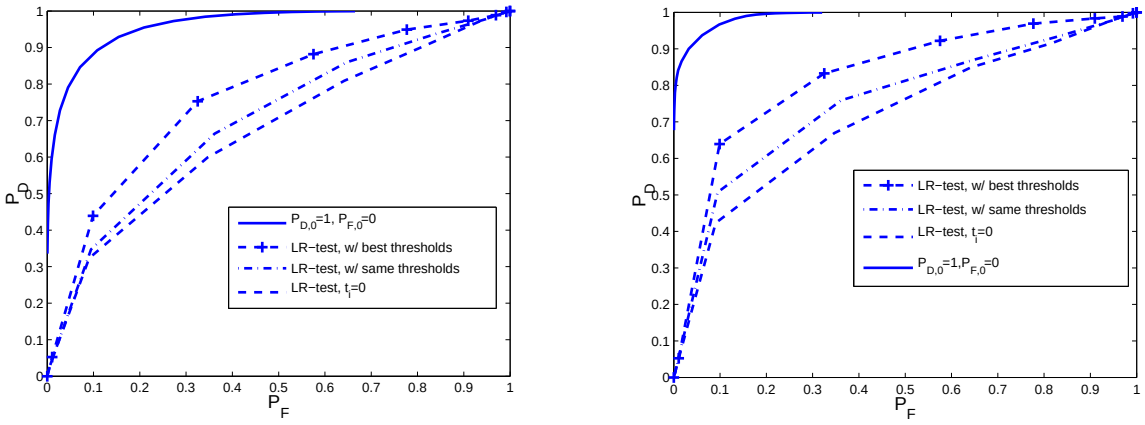


(a) (b)

Figure 45: ROC-curves for DC-level detection problem using serial and parallel networks. Simulation parameters: DC-level $m = 1$, $N = 2$ or 8 sensors, (a) $\rho = 0$ dB and (b) $\rho = 4.77$ dB, Rayleigh fading channel.

developed in Section 7.2.2, we obtain the optimal[5] threshold values $t_i,\ i = 1, \cdots, N$ by an exhaustive search. We used the numerical integration routine QUADL in MATLAB©. In Figures 46.a and 46.b, we illustrate the ROC curves for parallel distributed detection with $N = 2$ local sensors. The false alarm and detection probabilities are averaged over the channel distribution, *e.g.*, $P_{F,0} = E_{\mathbf{g}}\{P_{F,0}|\mathbf{g}\}$. We consider 4 detection schemes: (i) ideal local detectors[6] , i.e., $P_{D,j} = 1, P_{F,j} = 0$, (ii) LR tests at all sensors with different local thresholds, (iii) LR tests with the same local thresholds, and (iv) LR tests with all thresholds set to 0, i.e., $t_j^* = 0$. For the cases (ii) and (iii), we perform an exhaustive search to determine the best threshold sets. We observe that the performance with the first scheme is superior to the LR test based detection schemes. This is expected since in the first scheme, the errors associated with the distributed detection are introduced during the transmission of the local decisions. When the local detectors employ LR tests, the best detection probability is obtained when we do not put any constraint on the thresholds (Case (ii)) and perform the search over all possible threshold sets. This search is computationally very complex. If we limit the search such that each local sensor use the same threshold (Case

[5]The optimality of the thresholds is in the sense that they maximize $P_{D,0}$ for some fixed $P_{F,0}$.
[6]By ideal detector, it is implied that the detection error probabilities are 0.



(a) (b)

Figure 46: Optimal ROC-curves for DC-level detection problem with a parallel network. Simulation parameters: DC-level $m = 1$, $N = 2$ sensors, (a) $\rho = 0$ and (b) $\rho = 4.77$ dB, Rayleigh fading channel

(iii)), the computation burden is much less, but we observe some performance degradation. The performance is worst for the heuristic detection where threshold is set to 0 at each local sensor node (Case (iv)).

The performance of multiple-bit decision and its comparison with the performance of single-bit decision is depicted in Figures 47-49. In Figure 47, we consider the case of two-bit decisions transmitted by BPSK signals as described in Section 7.3.1. For a fair comparison, we scale the transmission power for the two-bit decision scheme by 2. In Figure 47, we plot the ROC curves for three different schemes when $\rho = 10$ dB: (i) $N \in \{2, 10\}$ local sensors sending single-bit decisions at 10 dB/bit, (ii) $N \in \{2, 10\}$ local sensors sending two-bit decisions at 5 dB/bit, and (iii) $N \in \{1, 5\}$ local sensors sending two-bit decisions at 10 dB/bit. All three cases use the same amount of transmit power. We observe that if we employ $N = 2$ sensors, the detection performance obtained by the two-bit sensor decisions is superior to that obtained by the single-bit sensor decisions. When N is increased to 10, we observe that the two schemes perform similarly. The results in case (iii) is very promising in that even if we employ $N = 1$ or $N = 5$ sensors with two-bit decisions at each of them, the detection probability of the distributed detection is improved significantly and superior to single-bit decision. We note that this performance improvement is achieved by using

the same total power as the other schemes but utilizing only half as many local sensors. These results hint that the two-bit sensor decisions (transmitted with binary signaling) are promising in providing better detection performance compared to the single-bit sensor decisions under the same total power constraint.

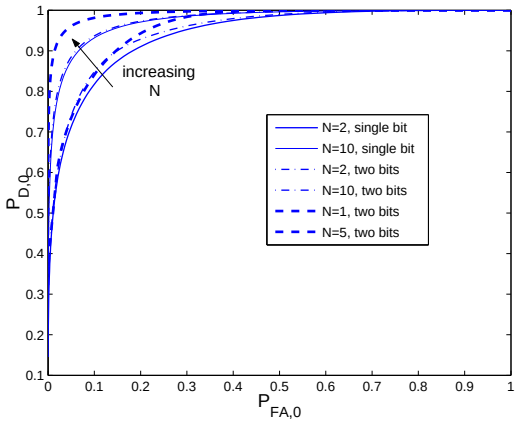
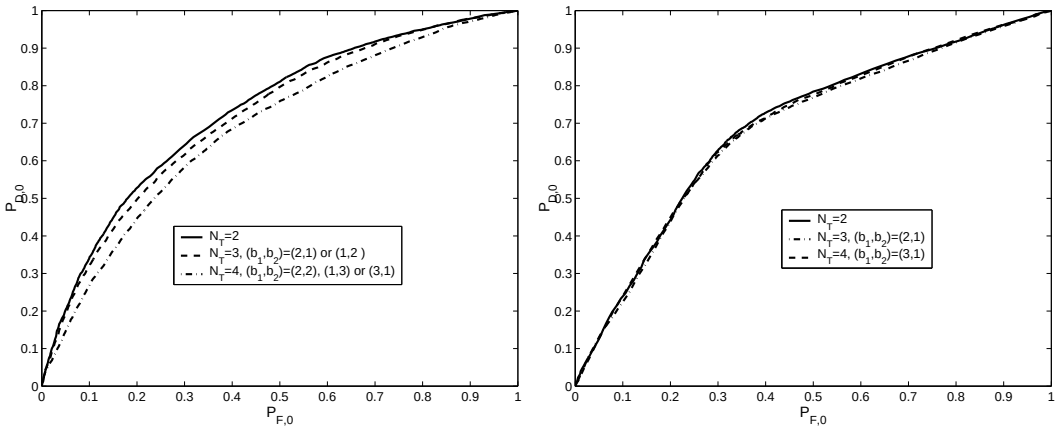


Figure 47: Comparison of ROC for distributed detection based on one-bit and two-bit sensor decisions. $\rho = 10$ dB/bit for all schemes

We next illustrate the distributed detection for different multi-bit decision schemes described in Section 7.3.2. In Figure 48, for a network of two sensor nodes, we compare the detection performance between three different schemes where a total of $N_T = 2$, 3 or 4 bits are allowed. Since the search for optimal thresholds is not practical, we use a fixed threshold scheme. Specifically, for the DC-level problem at hand, we set $\tau_l = m/2 + l\delta$, $l = -(2^{b_k-1}-1), \ldots, 0, 1, (2^{b_k-1}-1)$, where $\delta = m/2^{b_k}$ and b_k denotes the number of bits for the decision at node k. We preferred to use this threshold scheme since this one provided the best detection performance among various other threshold schemes we simulated. Figure 48.a depicts the ROC curves for the case where the links between the local nodes and the fusion center have similar fading levels whereas 48.b depicts the case when the one of the links is severely fading. For $N_T = 2$, the only possible scheme is to make single-bit decisions at both sensor nodes. For $N_T = 3$, we may have two cases: $(b_1, b_2) = (1, 2)$ or $(2, 1)$, and for $N_T = 4$, we have $(b_1, b_2) = (1, 3)$, $(2, 2)$, or $(3, 1)$. From both figures, we observe that in all cases, detection performance with single-bit decisions is superior to other schemes.

When the links have similar quality, it is seen that different bit assignment schemes provide similar quality for a given total bit budget. On the other hand, if one of the link is severely fading, the best $P_{D,0}$ for a fixed N_T is attained when more bits are assigned for the sensor node whose fading gain is smaller. For example, in Figure 48.b, for $N_T = 3$ and $N_T = 4$, the schemes with $(b_1, b_2) = (2, 1)$ and $(b_1, b_2) = (3, 1)$, respectively, detects better than those with other bit assignments with the same N_T.



(a) (b)

Figure 48: Comparison of ROC for distributed detection based on multi-bit and single-bit sensor decisions. $N = 2$ and $\rho = 1$ for all schemes. (a) $|g_1| = 0.8$, $|g_2| = 0.7$, (b) $|g_1| = 0.1$, $|g_2| = 1.01$

In Figure 49, we illustrate the performance of multi-bit decision with $N = 8$ sensor nodes. Since there are many ways to assign the number of decision bits (*e.g.*, different quantization schemes at each sensor node), we plot the ROC for three different schemes: (i) single-bit decision at all sensor nodes, (ii) single or two-bit decisions where a two-bit decision is made at a sensor node whose channel SNR is less than some pre-specified value and a single-bit decision is made otherwise, and (iii) two-bit decisions at all sensor nodes. we again observe that the best performance is obtained with a single-bit decision rule at each sensor node.

We note for the results in Figures 48 and 49 that they are obtained for a fixed threshold scheme. However, the specific result for the proposed threshold scheme is in compliance with the conclusion in [94] where from an information theoretical perspective, it is shown

that for an N-sensor network with a capacity constraint of N bits per time unit, having each sensor transmitting one bit is optimum.

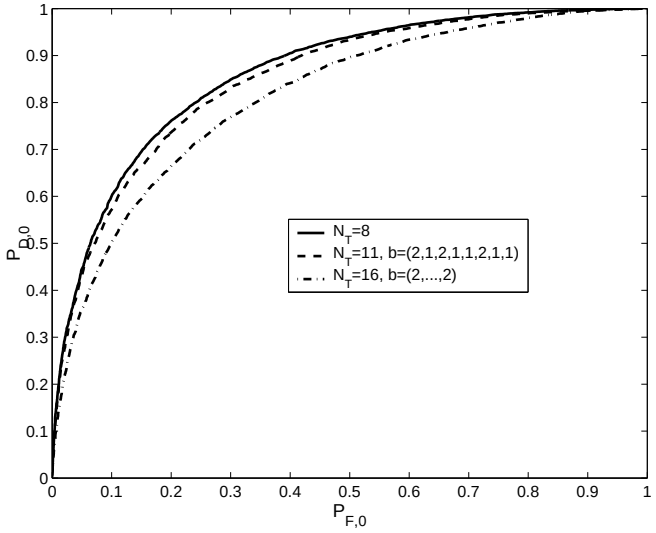


Figure 49: Comparison of ROC for distributed detection based on multi-bit and single-bit sensor decisions. $N = 8$ and $\rho = 1$ for all schemes.

We next study the performance of distributed detection with analog data gathering. We also compare its performance with the digital approach under the same total transmit power. According to NP-lemma, the optimal detection with the linear analog transmission of y_k is attained by the optimal power allocation given by (211). In the digital approach, a single-bit decision is obtained locally by a likelihood ratio test

$$\Lambda(y_k) = \log \frac{p(y_k|H_1)}{p(y_k|H_0)} \begin{cases} > t_k \ H_1, \text{send } +1 \\ \leq t_k \ H_0, \text{send } -1 \end{cases} \tag{219}$$

for which the false alarm and detection rates at the local sensor follow as

$$P_{F,k} = Q\left((t_k + m^2/\sigma_k^2)/(m/\sigma_k)\right)$$

and

$$P_{D,k} = Q\left((t_k - m^2/\sigma_k^2)/(m/\sigma_k)\right),$$

respectively. The signal received by the fusion center is given by $r_k = \sqrt{\rho_k g_k} u_k + n_k$, $k = 1, \ldots, K$ where $u_k \in \{-1, 1\}$ is the local decision at node k, and $\rho_k = \rho$ is the transmit

power gain.[7] The total transmit power is $P_T = K\rho$. The final decision is made at the fusion center by another likelihood ratio test

$$\Gamma(\mathbf{r}) = \log \frac{p(\mathbf{r}|H_1, \mathbf{g})}{p(\mathbf{r}|H_0, \mathbf{g})} \underset{H_0}{\overset{H_1}{\gtrless}} t_0. \tag{220}$$

To illustrate the comparison between the optimal analog- vs. digital-based detection, since networks of a larger number of sensor nodes do not allow for an analytical results, we first study the detection performance with a single sensor node. For $K = 1$, the global false alarm and detection rates for the digital approach described by the tests (219) and (220) are given by

$$P_{F,0} = P_{F,1} Q\left(\frac{t_0' - \sqrt{\rho_1 g_1}}{\xi_1}\right) + (1 - P_{F,1}) Q\left(\frac{t_0' + \sqrt{\rho_1 g_1}}{\xi_1}\right) \tag{221}$$

$$P_{D,0} = P_{D,1} Q\left(\frac{t_0' - \sqrt{\rho_1 g_1}}{\xi_1}\right) + (1 - P_{D,1}) Q\left(\frac{t_0' + \sqrt{\rho_1 g_1}}{\xi_1}\right) \tag{222}$$

where

$$t_0' = \frac{\xi_1^2}{\sqrt{4\rho_1 g_1}} \log \frac{(1 - P_{F,1})t_0 - (1 - P_{D,1})}{P_{D,1} - P_{F,1} t_0}.$$

For the best detection performance according to Neyman-Pearson criterion, we need to find threshold pairs (t_1, t_0) for which $P_{F,0} = \alpha$ and $P_{D,0}$ is maximized. The determination of (t_1, t_0)-pair even for this simple case is analytically infeasible. Instead, we resort to a numerical procedure where an exhaustive search for t_1 over the interval $[-W, W]$ is performed. For numerical reasons, we first obtain the uniformly-spaced L samples in this interval, *e.g.*, $t_l[n] = 2Wn/L,\ n = -L/2, \ldots, L/2$. To find the threshold pair for a specified α, we obtain t_0 for all $t_l[n]$ using (221) with the bisection method, and then calculate $P_{D,0}$ using (222). Finally, we select the threshold $t_l[n]$ that corresponds to the maximum $P_{D,0}$. It is clear that the procedure for the determination of the optimal threshold pairs is very tedious and the complexity increases very rapidly as K increases. For $K > 1$, a rather simplistic, but a suboptimal method is to employ the same thresholds at all local sensor nodes. However, this choice does not guarantee the maximal detection rates.

[7] The notation for for the channel gain is slightly changed to make the comparison between the analog and digital schemes fairly.

In Figure 50, we depict the ROC curves for $K = 1$ at the channel signal-to-noise ratios of $\rho = 1$ and $\rho = 10$ (which corresponds to $P_T = 1$ and $P_T = 10$, respectively, for the analog approach). The simulation parameters are $m = 1$, $\sigma_1^2 = 1$, $\xi_1^2 = 1$. For the digital approach, the performance of two detection rules are considered: (i) $t_1 = 0$ for all α, and (ii) optimum t_1 obtained for each α using the exhaustive search procedure described above. It is seen that for $\rho = 1$, the detection rates with the analog approach is uniformly higher than both digital approaches. For $\rho = 10$, the digital approach with optimal threshold attains better detection rates than the analog approaches does; however, the performance difference is negligibly small. We note that the cost of computing the optimal thresholds even for this simple case is very high. However, the analog approach is based on a simple amplify-and-forward technique and can achieve detection rates very close to that of the optimal scheme. We also observe that as the signal-to-noise ratio is increased, the detection performance at the fusion center for both schemes merges to the one that can be attained by the centralized detection. Another interesting observation is that at $\rho = 10$, the ROC curve for the digital approach with $t_0 = 0$ has a piece-wise linear form which bends at $P_{F,0} = 0.31$ where the $P_{D,0}$ coincides with that of the digital approach using the optimal threshold.

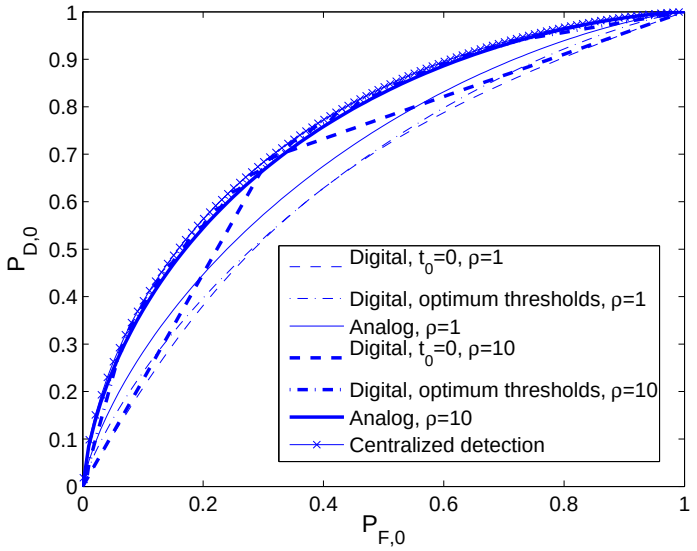


Figure 50: Comparison of ROC for distributed detection based on the digital and analog approaches. The simulation parameters: $\rho = 1$ or 10, $\sigma_1^2 = 1$, $m = 1$, $\xi_1^2 = 1$, $g_1 = 0.7$.

Next, we illustrate the results when $K = 2$ and $K = 8$ in Figures 51.a and 51.b, respectively. We assume that $\sigma_k^2 = 1$, and $\xi_k^2 = 1$, for $k = 1, \ldots, K$. The channel gains, which are the realizations of a Rayleigh random variable, are $\mathbf{g} = [0.1\ 0.9]$ and $\mathbf{g} = [1.07\ 0.50\ 0.94\ 1.27\ 1.06\ 0.71\ 0.79\ 0.67]$ for $K = 2$ and $K = 8$, respectively. For the digital approach, determining the optimal threshold vector $\mathbf{t} = [t_1, \ldots, t_K]$ is impractical because of the computational burden of the search over a set of L^K $K-$tuples; therefore, we set $t_k = 0$. From the ROC curves in Figures 51.a and 51.b, the following observations are in order: (i) For all cases under consideration, the detection performance of the analog approach with optimal power allocation is superior to the detection performance of the digital approach. Furthermore, the detection rates with the uniform power allocation is either higher than or very similar to those attained by the digital approach. (ii) As the transmit power in increased, the ROC curves for the analog approach converges to those of the centralized detection. (iii) For the analog approach at high signal-to-noise ratios, optimal and uniform power allocation schemes result in nearly same detection rates. These results indicate that the analog transmission-based detection is a promising technique for distributed detection applications in sensor networks.

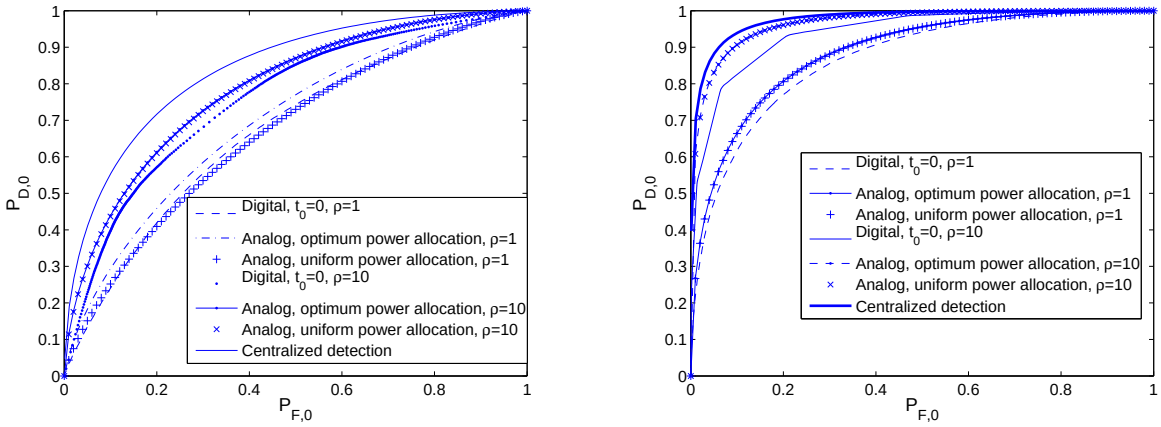


Figure 51: Comparison of ROC for distributed detection based on the digital and analog approaches. The simulation parameters: $\rho = 1$ or 10, $\sigma_k^2 = 1$, $m = 1$, $\xi_k^2 = 1$, a) (Left) $K = 2$, $\mathbf{g} = [0.1\ 0.9]$, b) (Right) $K = 8$, $\mathbf{g} = [1.07\ 0.50\ 0.94\ 1.27\ 1.06\ 0.71\ 0.79\ 0.67]$.

Figures 52.a and 52.b plot the miss probability ($P_{M,0} = 1 - P_{D,0}$) vs. signal-to-noise ratio

for $K = 1,\ 2,\ 4,\ 8$ and 16 sensor nodes. The channel fading is Rayleigh distributed and the probability of miss is obtained by averaging over the channel statistics. We observe in both figures that, at high signal-to-noise ratios, the miss probability with K sensors converges to a lower bound that coincides with the miss probability of the centralized detection, as expected. On the other hand, for $P_T/K <$ 10 dB, we see that the slopes of P_M vs. P_T/K plots increases as K increases, which indicates that significant diversity (along with the increased sample-size) can be attained by distributed detection when the channels between the sensor nodes are subject to independent fading.

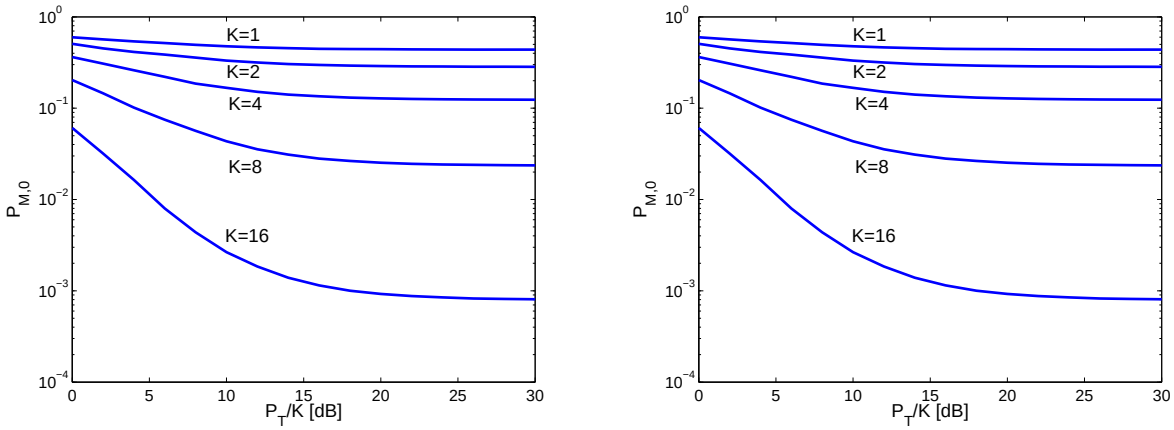


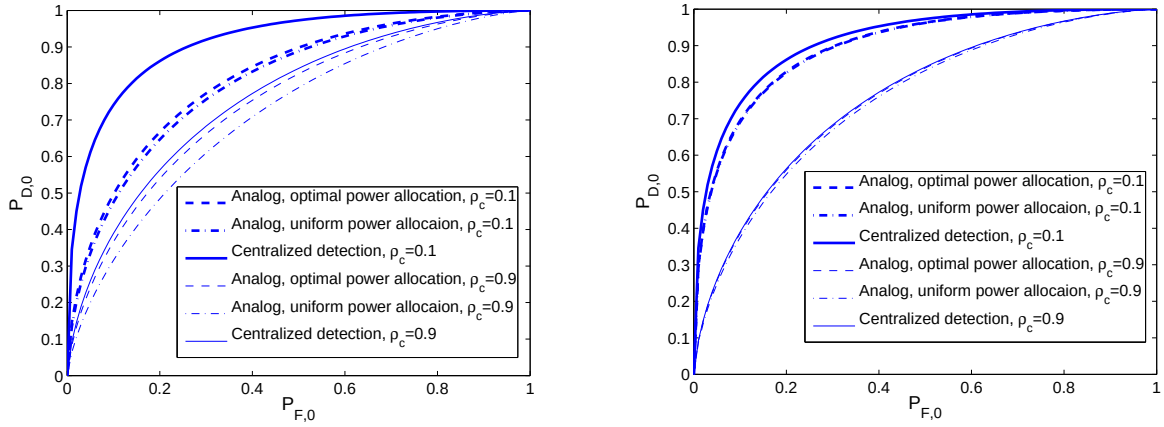
Figure 52: Miss Probability vs. P_T/K. $P_{M,0}$ is averaged over the channel statistics. The simulation parameters: $\sigma_k^2 = 1$, $m = 1$, $\xi_k^2 = 1$. a) (Left) $P_{F,0} = 0.2$, and b) (Right) $P_{F,0} = 0.3$

So far, we studied examples for which the observations at the local sensor nodes are assumed to be independent from each other, *e.g.*, $\mathbf{C}_o$ is diagonal. While the digital approach does not allow for simple solutions to the case of conditionally dependent observations, the analog approach does. In the next example, we assume that the covariance matrix has a block circulant form given by

$$[\mathbf{C}_o]_{i,j} = \rho_c^{|j-i|},\ \rho_c \leq 1 \tag{223}$$

and the mean of y_k is given by $\mu_k = e^{-(k-1)/10},\ k = 1, \ldots, K$, *e.g.*, the observation signal-to-noise ratio decreases exponentially from sensor node $k = 1$ to sensor node K. Figure 53 illustrates the ROC curves for $K = 8$. We observe that as the correlation between the

observations of the sensor nodes increases, the detection performance for both the centralized detection and the distributed detection schemes degrade significantly. For example, for the centralized detector with $P_{F,0} = 0.2$, the $P_{D,0}$ decreases to 0.563 from 0.86 while ρ_c increases from 0.1 to 0.9; meanwhile for the distributed detection scheme at $\rho = 10$, the $P_{D,0}$ decreases to 0.559 from 0.83 (see Figure 53.b). It is also seen that at higher levels of correlation, the detection rates attained by the analog-based detection scheme are very close to those of the centralized detector for both $\rho = 1$ and $\rho = 10$. This hints that the intersensor correlations can be exploited to have better detection rates by the proposed detection scheme.



(a) (b)

Figure 53: ROC when $\mathbf{C}_o$ is given by (223). The simulation parameters: $K = 8$, $\rho_c = 0.1$(thick lines) and $\rho_c = 0.9$, $\sigma_k^2 = 1$, $\xi_k^2 = 1$, a) (Left) $\rho = 1$ ($P_T = 8$) and b) (Right) $\rho = 10$ ($P_T = 80$).

7.7 Chapter Summary

We investigated the distributed detection problem in WSN under the assumption of fading channels. Various schemes are analyzed and the optimal design of distributed detectors for each scheme are studied. We derive the LR-based optimal fusion rules that incorporate fading in the distributed detection problem. Analog data gathering for the distributed detection problem is proposed. We show that with analog approach, the detection performance can be optimized by a suitable power allocation. In the digital approach, the

optimal decision rule amounts to determine the local thresholds and the global threshold to be employed for the LR test. We observe that multiple-bit sensors decisions help improve the detection performance with the same transmit power when transmitted by BPSK modulation. If multi-bit decisions are transmitted with M-ary modulation schemes, simulation results indicate that transmitting one bit with BPSK has a better detection performance. The results on the use of sequential detection over fading channels show that for large networks, sequential detection can be very efficient compared to fixed-sample size distributed detector.

CHAPTER VIII

CONCLUSIONS

MIMO communications is currently being implemented in many telecommunication systems and the next generation wireless infrastructures are expected to enjoy the benefits of MIMO communications to its ultimate extent. In this dissertation we dealt with several issues arose in the application of MIMO sysems in wireless communications. In Chapters 2–4, we consolidated the performance analysis for a MIMO antenna system using receive antenna selection. In particular, we studied three schemes (i) space-time coding over iid MIMO fading channel, (ii) space-time coding over spatially correlated MIMO fading channel , and (iii) MIMO-OFDM systems. We derived explicit closed-form upper bounds on the pairwise error probabilities for the scenarios being considered, and showed that, for each case, one can retain the asymptotical performance gain with an energy-based antenna selection. The resulting analysis is also used for designing optimum space-time codes for MIMO systems using antenna selection.

In Chapter 5, we dealt with the joint-source channel coding for a MIMO antenna system. We proposed a turbo-coded multiple description coding scheme for multiple antenna transmission. This scheme is suitable for end-to-end transmission of some real valued signal through a MIMO wireless channel. We showed by simulations that by transmitting correlated streams over an iid fading MIMO channel, we can attain less mean-square error distortion than that attained by the single-description coding. This scheme is a way to exploit the correlation between the streams that arrives at the receiver through independently fading channels. The proposed iterative decoding method achieves this by an extrinsic information exchange between the source decoder and channel decoder.

In Chapters 2-5, our assumption was that the transmitters and/or receivers were equipped with multiple antenna elements. However, because of the limited size of the mobile units, it may not be possible to deploy more than one antenna. In this case, it is still possible

to benefit from spatial diversity provided that the individual nodes cooperative with each other. In Chapter 6, we proposed a distributed turbo code for a wireless relay channel where the relay node assists the source node to attain such cooperative diversity. We assumed a full-duplex relay node which can simultaneously transmit (the previous block it decoded) and receive (the fresh block from the source). Hence, the information from the source to the destination node can be transmitted in contiguous blocks. We proposed an iterative decoding method and showed that with this scheme, it is possible to get within 1-dB of the capacity limits of the relay channel.

In the last chapter of the dissertation, we investigated distributed detection strategies for wireless sensor networks in which the links between the sensor nodes are noisy and fading. We categorized two main approaches: (i) digital approach where the local sensors makes a local decision consisting of a few bits, and (ii) analog approach where the decision statistics or observations are transmitted using a linear analog transmission scheme. For each case being considered, we derived expressions for detection and false alarm probabilities, which are then used for performance assessment and design optimization. We compared the performance of these strategies via simulations. The advantage of the analog approach is that the optimal design is much simpler than that for the digital approach. Furthermore, in many cases, the detection performance is superior with the analog approach. For the digital approach, it turns out that three is trade-off between the number of bits allocated for local decisions and the detection performance. Finally, we note that the sequential detection in wireless sensor networks is a viable method for energy-efficient detection.

APPENDIX A

UPPER BOUND IN CLOSED FORM FOR $M = 2$ AND ANY N WHEN $L = 1$

Applying the binomial expansion in (47), we obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{8N}{(\lambda_2 - \lambda_1)\rho} \sum_{n=0}^{N-1} C(N-1, n) \int_0^1 \log^n(x/e) \left(x^{n+\rho\lambda_1/8} - x^{n+\rho\lambda_2/8} \right) dx \quad (224)$$

Letting $x/e = u$ and then $log(u) = t$, we obtain

$$P(\mathbf{S} \to \hat{\mathbf{S}}) \leq \frac{8N}{(\lambda_2 - \lambda_1)\rho} \sum_{n=0}^{N-1} C(N-1, n) \left[e^{n+1+\rho\lambda_1/8} \int_{-\infty}^{-1} t^n e^{t(n+1+\rho\lambda_1/8)} - e^{n+1+\rho\lambda_2/8} \int_{-\infty}^{-1} t^n e^{t(n+1+\rho\lambda_2/8)} dt \right] \quad (225)$$

Now, consider

$$I_n = \int_{-\infty}^{-1} t^n e^{at} dt$$

Integration by parts gives us

$$I_n = (-1)^n \frac{e^{-a}}{a} - \frac{n}{a} I_{n-1}$$

Solving this recurrent equation for I_n with the initial value of $I_0 = \frac{e^{-a}}{a}$, we get

$$I_n = (-1)^n \frac{e^{-a}}{a} \left[1 + \frac{n}{a} + \frac{n(n-1)}{a^2} + \cdots + \frac{n!}{a^n} \right] \quad (226)$$

Using $a = n + 1 + \rho\lambda_1/8$ and $b = n + 1 + \rho\lambda_2/8$ in (225) along with (226), the closed form expression is obtained as

$$P(S \to \hat{S}) \leq \frac{8N}{(\lambda_2 - \lambda_1)\,\rho} \sum_{n=0}^{N-1} C(N-1, n)(-1)^n \left\{ \left(\frac{1}{a_n} + \frac{n}{a_n^2} + \frac{n(n-1)}{a_n^3} + \cdots \frac{n!}{a_n^{n+1}} \right) - \left(\frac{1}{b_n} + \frac{n}{b_n^2} + \frac{n(n-1)}{b_n^3} + \cdots \frac{n!}{b_n^{n+1}} \right) \right\} \quad (227)$$

Further simplification can be made after regrouping the terms in the summation and then using $x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \cdots + y^{n-1})$:

$$P(S \to \hat{S}) \leq N \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n \left(\frac{1}{a_n b_n} + \frac{n\,(a_n + b_n)}{a_n^2 b_n^2} + \frac{n(n-1)\left(a_n^2 + a_n b_n + b_n^2\right)}{a_n^3 b_n^3} \right.$$
$$\left. + \cdots + \frac{n!\,(a_n^n + \cdots + b_n^n)}{a_n^{n+1} b_n^{n+1}} \right)$$

APPENDIX B

PROOFS FOR LEMMA AND APPROXIMATION

B.1 Proof of Lemma

Let $I_c = \int_{\mathcal{C}} (\mathbf{z}^H \mathbf{A}\mathbf{z})^i e^{-\mathbf{z}^H \mathbf{B}\mathbf{z}} d\mathbf{z}$. First, assume that $\mathbf{A}$ is non-singular, e.g., $\lambda_i > 0,\ i = 1, \cdots, K$. Then,

$$
\begin{aligned}
I_c &= \int_{\mathcal{C}} (\mathbf{z}^H \mathbf{\Lambda}\mathbf{z})^i e^{-\mathbf{z}^H \hat{\mathbf{B}}\mathbf{z}} d\mathbf{z} && (228)\\
&= \frac{1}{|\mathbf{\Lambda}|} \int_{\mathcal{C}} (\mathbf{v}^H \mathbf{v})^i e^{-\mathbf{v}^H \bar{\mathbf{B}}\mathbf{v}} d\mathbf{v} && (229)\\
&= \frac{1}{|\mathbf{\Lambda}|} \int_{\mathcal{C}} (\mathbf{v}^H \mathbf{v})^i e^{-\mathbf{v}^H \mathbf{\Gamma}\mathbf{v}} d\mathbf{v} && (230)\\
&= \frac{\pi^K}{|\mathbf{\Lambda}|} \int_0^\infty \cdots \int_0^\infty (u_1 + \cdots + u_K)^i e^{-(\gamma_1 u_1 + \cdots + \gamma_K u_K)} du_1 \cdots d_K && (231)\\
&= \frac{\pi^K}{|\mathbf{\Lambda}|} \sum_{l_1} \cdots \sum_{l_K} \frac{i!}{l_1! \cdots l_K!} \int_0^\infty \cdots \int_0^\infty u_1^{l_1} \cdots u_K^{l_K} e^{-(\gamma_1 u_1 + \cdots + \gamma_K u_K)} du_1 \cdots d_K && (232)\\
&= \frac{\pi^K}{|\mathbf{\Lambda}|} \sum_{l_1} \cdots \sum_{l_K} \frac{i!}{\gamma_1^{l_1+1} \cdots \gamma_K^{l_K+1}} && (233)\\
&= \frac{\pi^K}{|\mathbf{\Lambda}||\bar{\mathbf{B}}|} \sum_{l_1} \cdots \sum_{l_K} \frac{i!}{\gamma_1^{l_1} \cdots \gamma_K^{l_K}} && (234)\\
&= \frac{\pi^K}{|\mathbf{B}|} f_i(\mathrm{tr}(\bar{\mathbf{B}}^{-1}), \cdots, \mathrm{tr}(\bar{\mathbf{B}}^{-i})) && (235)\\
&= \frac{\pi^K}{|\mathbf{B}|} f_i(\mathrm{tr}(\mathbf{A}\mathbf{B}^{-1}), \cdots, \mathrm{tr}((\mathbf{A}\mathbf{B}^{-1})^i)) && (236)
\end{aligned}
$$

where (228) follows by the SVD $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, and $\hat{\mathbf{B}} = \mathbf{U}^H\mathbf{B}\mathbf{U}$, (229) follows by the change of variable $\mathbf{v} = \mathbf{\Lambda}^{1/2}\mathbf{z}$, and $\bar{\mathbf{B}} = \mathbf{\Lambda}^{-1/2}\hat{\mathbf{B}}\mathbf{\Lambda}^{-1/2}$, (230) follows by the SVD $\bar{\mathbf{B}} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$, (231) follows by converting the integral to the polar coordinates, (232) follows by the binomial expansion and then changing the order of integration and summation (note that the nested summation includes the $(l_1, \cdots, l_K)$ K-tuples for which $\sum_{k=1}^K l_K = i$), (233) follows by $\int_0^\infty u^n e^{-au} = n!/a^{n+1}$, (234) follows by $\prod_{k=1}^K \gamma_k = |\bar{\mathbf{B}}|$, (235) follows by $|\bar{\mathbf{B}}| = |\mathbf{\Lambda}|^{-1}|\mathbf{B}|$, and regrouping the terms in the summation and using the fact that $\sum_{k=1}^K \gamma_k^j =$

$\text{tr}(\bar{\mathbf{B}}^{-j})$, and finally (236) follows by $\bar{\mathbf{B}}^{-1} = \mathbf{\Lambda}^{1/2}\mathbf{U}^H\mathbf{B}^{-1}\mathbf{U}\mathbf{\Lambda}^{1/2}$, and noting that $\text{tr}(\bar{\mathbf{B}}^{-j}) = \text{tr}(\mathbf{\Lambda}^{1/2}\mathbf{U}^H\mathbf{B}^{-1}\mathbf{U}\mathbf{\Lambda}^{1/2}\cdots\mathbf{\Lambda}^{1/2}\mathbf{U}^H\mathbf{B}^{-1}\mathbf{U}\mathbf{\Lambda}^{1/2}) = \text{tr}(\mathbf{A}\mathbf{B}^{-1}\cdots\mathbf{A}\mathbf{B}^{-1}) = \text{tr}((\mathbf{A}\mathbf{B}^{-1})^j)$.

If $\mathbf{A}$ is singular, we have $\mathbf{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)$, i.e., some of the eigenvalues will be 0s. Therefore, we can not use the change of variable $\mathbf{z} = \mathbf{\Lambda}^{-1/2}\mathbf{v}$ and the set of equations after (228) are not valid. Let $\mathbf{A}' = \mathbf{A} + \epsilon I_K$ for any $\epsilon > 0$, and define

$$I_c' = \int_{\mathcal{C}} (\mathbf{z}^H\mathbf{A}'\mathbf{z})^i e^{-\mathbf{z}^H\mathbf{B}\mathbf{z}} d\mathbf{z} \tag{237}$$

$$= \int_{\mathcal{C}} (\mathbf{z}^H\mathbf{\Lambda}'\mathbf{z})^i e^{-\mathbf{z}^H\hat{\mathbf{B}}\mathbf{z}} d\mathbf{z} \tag{238}$$

where $\mathbf{\Lambda}' = \mathbf{\Lambda} + \epsilon I_K$. Since $\mathbf{\Lambda}'$ is non-singular, we can now proceed to obtain

$$I_c' = \frac{\pi^K}{|\mathbf{B}|} f_i(\text{tr}(\mathbf{A}'\mathbf{B}^{-1}), \cdots, \text{tr}((\mathbf{A}'\mathbf{B}^{-1})^i)) \tag{239}$$

Taking the limit, we finally arrive at the desired result

$$\lim_{\epsilon\to 0} I_c' = \frac{\pi^K}{|\mathbf{B}|} \lim_{\epsilon\to 0} f_i(\text{tr}(\mathbf{A}'\mathbf{B}^{-1}), \cdots, \text{tr}((\mathbf{A}'\mathbf{B}^{-1})^i)) \tag{240}$$

$$= \frac{\pi^K}{|\mathbf{B}|} f_i(\text{tr}(\mathbf{A}\mathbf{B}^{-1}), \cdots, \text{tr}((\mathbf{A}\mathbf{B}^{-1})^i)) \tag{241}$$

$$= I_c \tag{242}$$

A proof for the special case of $i = 1$ can also be obtained using the results of [38] for which $f_1(\text{tr}(\mathbf{A}\mathbf{B}^{-1})) = \text{tr}(\mathbf{A}\mathbf{B}^{-1})$. ■

B.2 Approximation

Note that if λ_k, $k = 1, \ldots, n$, are the eigenvalues of the $n \times n$ matrix A, then the eigenvalues of cA and $cI_n + A$ are given by $c\lambda_k$ and $c + \lambda_k$, $k = 1, \ldots, n$, respectively. Using the equalities $|A| = \prod_{k=1}^n \lambda_k$ and $\text{tr}(A) = \sum_{k=1}^n \lambda_k$, we have

$$|I_n + \epsilon A| = \prod_{k=1}^n (1 + \epsilon\lambda_k) = 1 + \epsilon \sum_{k=1}^n \lambda_k + \mathcal{O}(\epsilon^2) \tag{243}$$

$$\approx 1 + \epsilon\text{tr}(A)$$

where in (244), we omit the higher order terms $\mathcal{O}(\epsilon^2)$. ■

REFERENCES

[1] Agrawal, D., Tarokh, V., Naguib, A., and Seshadri, N., "Space-time coded ofdm for high data-rate wireless communication over wideband channels," in *IEEE Vehicular Tech. Conf.*, pp. 2232–2236, May 1998.

[2] Alamouti, S. M., "A simple transmit diversity technique for wireless communications," *IEEE Journal of Selected Areas in Communications*, vol. 16, pp. 1451–1458, Oct. 1998.

[3] A.Nasipuri and S.Tantaratana, "Nonparametric distributed detection using wilcoxon statistics," in *26th Annu. Conf. on Inform. Sci. and Syst.*, pp. 264–269, March 1993.

[4] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, 1984.

[5] A.Papoulis, *Probability, Random Variables, and Stochasitc Process*. McGraw-Hill, 3 ed., 1991.

[6] A.R.Elias-Fuste, A.Broquetas-Ibars, J. and J.M.Yuste, "Cfar data fusion center with inhomogeneous receivers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, pp. 276–285, January 1992.

[7] At, N. and Altunbasak, Y., "Multiple description coding for wireless channels with multiple antennas," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Sept. 2001.

[8] Bahceci, I. and Duman, T. M., "Trellis coded unitary space-time modulation," *IEEE Trans. on Wireless Comm.*, vol. 3, pp. 2005–20012, Nov 2004.

[9] Bahl, L. R., Cocke, J., Jelinek, F., and Raviv, J., "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, pp. 284–287, Mar. 1974.

[10] Barkat, M. and Varshney, P. K., "Decentralized cfar signal detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, pp. 141–149, March 1989.

[11] Barkat, M. and Varshney, P. K., "Adaptive cell-averaging cfar detection in distributed sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, pp. 424–429, May 1991.

[12] Baro, S., Bauch, G., and Hansmann, A., "Improved codes for space-time trellis codes," *IEEE Communications Letters*, vol. 41, pp. 20–22, Jan. 2000.

[13] Benedetto, S., Divsalar, D., Montorsi, G., and Pollara, F., "Parallel concatenated trellis coded modulation," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 974–978, 1996.

[14] Benedetto, S., Divsalar, D., Montorsi, G., and Pollara, F., "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Transactions on Information Theory*, pp. 909–929, May 1998.

[15] Benedetto, S. and Montorsi, G., "Average performance of parallel concatenated block codes," *Electronics Letters*, pp. 156–158, Feb. 1995.

[16] Berrou, C. and Glavieux, A., "Near optimum error correcting and decoding: Turbo codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, Oct 1996.

[17] Berrou, C., Glavieux, A., and Thitimajshima, P., "Near shannon limit error-correcting coding and decoding: Turbo-codes," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 1064–1070, 1993.

[18] Blum, R. S., "Locally optimum distributed detection of dependent random signals based on ranks," *IEEE Transactions on Information Theory*, vol. 42, pp. 931–942, May 1996.

[19] Blum, R. S., Li, Y. G., Winters, J. H., and Yan, Q., "Improved space-time coding for mimo-ofdm wireless communications," *IEEE Transactions on Communications*, vol. 49, pp. 1873–1878, Nov 2001.

[20] Bolckei, H., Gesbert, D., and Paulraj, A. J., "On the capacity of ofdm-based spatial multiplexing," *IEEE Transactions on Communications*, vol. 50, pp. 225–234, Feb 2002.

[21] Bolcskei, H., Nabar, R. U., Erceg, V., Gesbert, D., and Paulraj, A. J., "Performance of spatial multiplexing in the presence of polarization diversity," *IEEE Trans. Signal Processing*, vol. 50, pp. 2553–2562, Oct 2002.

[22] Bolcskei, H. and Paulraj, A., "Performance of space-time codes in the presence of spatial fading correlation," in *Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 687–693, Nov 2000.

[23] Boyd, S. and Vandenberghe, L., *Convex Optimization*. Cambridge University Press, 2004.

[24] Chen, B., Jiang, R., Kasetkasem, T., and Varshney, P. K., "Fusion of decisions transmitted over fading channels in wireless sensor networks," in *ICAASP 2002*, pp. 1184–1188, Nov 2002.

[25] Chen, B., Jiang, R., Kasetkasem, T., and Varshney, P. K., "Channel aaware distributed detection in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 52, pp. 3454–3458, Dec 2004.

[26] Chen, B. and Willet, P. K., "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 693–699, Feb 2005.

[27] Chen, Z., Vucetic, B., Yuan, J., and Lo, K.-L., "Space-time trellis codes with two, three and four transmit antennas in quasi-static flat fading channels," in *Proc. of IEEE ICC*, vol. 3, 200x. pp. 1589-1595, April 2002.

[28] Chiani, M., Win, M. Z., and Zanella, A., "On the capacity of spatially correlated mimo rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2363–2371, Oct 2003.

[29] Chou, P. A., Mehrotra, S., Wang, A., Storer, J. A., and Cohn, M., "Multiple description decoding of overcomplete expansions using projections onto convex sets," in *Proceedings DCC'99 Data Compression Conference*, pp. 72–81, 1999.

[30] Chuah, C., Kahn, J. M., and Tse, D., "Capacity of multi-antenna array systems in indoor wireless environment," in *Globecomm'98*, vol. 4, (Sydney), pp. 1894–1899, Nov 1998.

[31] Chung, D. and Yang, W., "Multiple description image coding using signal decomposition and and reconstruction based on lapped orthogonal transforms," *IEEE Trans. Ciruits and Systems for Video Tech.*, pp. 895–908, Sept. 1999.

[32] Cover, T. and Thomas, J., *Elements of Information Theory*. New York: Wiley & Sons, 1991.

[33] Cover, T. M. and El-Gamal, A. A., "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, Sep 1979.

[34] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.

[35] Cui, S., Xiao, J.-J., Goldsmith, A. J., Luo, Z.-Q., and Poor, H. V., "Energy-efficient joint estimation in sensor networks: Analog vs. digital," in *ICASSP*, (Philadelphia, PA), pp. 745–748, IEEE, March 2005.

[36] Divsalar, D. and Pollara, F., "Multiple turbo codes for deep-space communications," TDA Progres Report 42-121, JPL, May 1995.

[37] D.Li, K.D.Wang, Y. and A.M.Sayeed, "Detection, classfication and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, pp. 17–29, March 2002.

[38] Dogandžić, A., "Chernoff bounds on pairwise error probabilities of space-time codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 1–9, May 2003.

[39] Drakopoulos, E. and Lee, C. C., "Optimal multisensor fusion of correlated local decisions," *IEEE Trans. Aerospace Elect. Syst.*, vol. 27, pp. 593–605, July 1991.

[40] D.Teneketzis and Y.C.Ho, "The decentralized wald problem," *Information Comput.*, vol. 73, pp. 23–24, 1987.

[41] Duman, T. M., "Turbo codes and turbo coded modulation systems: Analysis and performance bounds." Ph.D. Dissertation, Electrical and Computer Engineering Department, Northeastern University, Boston, MA., May 1998.

[42] Duman, T. M. and Salehi, M., "Decentralized detection over multiple access channels," *IEEE Trans. Aerospace Elect. Syst.*, vol. 34, pp. 469–476, Apr. 1998.

[43] E.Geraniotis and Y.A.Chau, "Robust data fusion for multi-sensor detection systems," *IEEE Transactions on Information Theory*, vol. 36, pp. 1265–1279, November 1990.

[44] E.K.Hussaini, A.-B. and Y.A.El-Far, "Decentralized cfar signal detection," *Signal Process*, vol. 44, pp. 299–307, July 1995.

[45] Farvardin, N. and Vaishampayan, V., "On the performance and complexity of channel optimized vector quantizers," *IEEE Transactions on Information Theory*, vol. 37, pp. 155–160, Jan 1991.

[46] Fingscheidt, T., Hindelang, T., Cox, R. V., and Seshadri, N., "Joint source-channel (de-)coding for mobile communications," *IEEE Transactions on Communications*, vol. 50, pp. 200–212, Feb 2002.

[47] Finkelstein, Y., "Antenna selection in multicarrier communication systems," *Intel Tech. Journal*, May 2003. available at http://developer.intel.com/technology/itj/2003/volume07issue02/.

[48] Foschini, Jr., G. J., "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell System Technical Journal, Autumn*, pp. 41–59, 1996.

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

[50] Gallagher, R. G., *Information Theory and Reliable Communication*. John Wiley and Sons, Inc, 1968.

[51] Galliou, S. and Belfiore, J.-C., "A new family of full rate, fully diverse space-time codes based on galois theory," in *ISIT*, (Lausanne, Switzerland), p. 419, June 2002.

[52] Gamal, H. E. and Damen, M. O., "An algebraic number theoric framework for space-time coding," in *ISIT*, (Lausanne, Switzerland), p. 132, June 2002.

[53] Garcia-Frias, J., "Joint source channel decoding of correlated sources over noisy channels," in *Proceedings of Data Compression Conference DCC'2001*, pp. 283–292, 2001.

[54] Garica-Frias, J. and Villasenor, J. D., "Joint turbo decoding and estimation of hidden markov sources," *IEEE Journal of Selected Areas in Communications*, pp. 1671–1679, Sept. 2001.

[55] Gastpar, M., Rimoldi, B., and Vetterli, M., "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, pp. 1147–1158, May 2003.

[56] Gastpar, M. and Vatterli, M., "On the capacity of wireless netwroks: The relay case," in *IEEE INFOCOMM*, pp. 1577–1586, 2002.

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

[58] Giannakis, G. B. and Zhou, S., "Optimal transmit-diversity precoders for random fading channels," in *IEEE GLOBECOM*, vol. 3, pp. 1839–1843, 2000.

[59] Goblick, T. J., "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, pp. 558–567, Oct 1965.

[60] Goff, S. L., Glavieux, A., and Berrou, C., "Turbo-codes and high spectral efficiency modulation," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 645–649, 1994.

[61] Gore, D. A., Heath, R., and Paulraj, A., "Statistical antenna selection for spatial multiplexing systems," in *IEEE International Conf. Comm. 2002*, vol. 1, pp. 450–454, 2002.

[62] Gore, D. A., Nabar, R. U., and Paulraj, A., "Selecting an optimal set of transmit antennas for a low rank matrix channel," in *ICASSP*, pp. 2785–2788, 2000.

[63] Gore, D. A. and Paulraj, A., "Space-time block coding with optimal antenna selection," in *ICASSP*, pp. 2441–2444, 2001.

[64] Gore, D. A. and Paulraj, A., "Statistical mimo antenna sub-set selection with space-time coding," in *ICC'2002*, pp. 641–645, 2002.

[65] Gore, D. A. and Paulraj, A. J., "Mimo antenna subset selection with space-time coding," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2580–2588, Oct. 2002.

[66] Gorokhov, A., "Antenna selection algorithms for mea transmissions," in *IEEE ICASSP*, vol. 3, pp. 2857–2860, 2002.

[67] Gortz, N., "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal of Selected Areas in Communications*, vol. 19, pp. 1662–1670, Sep 2001.

[68] Goyal, V. K. and Kovacevic, J., "Optimal multiple description transform coding of gaussian vectors," in *Data Compression Conference*, pp. 388–397, 1999.

[69] Goyal, V. K., Kovacevic, J., Arean, R., and Vatterli, M., "Optimal multiple description transform coding of gaussian vectors," in *In Proc ICIP98, Chicago,IL*, pp. 674–678, 1999.

[70] Goyal, V. K. and Kovačević, J., "Generalized multiple description coding with correlating transforms," *IEEE Transactions on Information Theory*, vol. 47, pp. 2199–2223, Sep 2001.

[71] Gradshteyn, I. S. and Ryzhik, I. M., *Table of Integrals, Series and Products*. New York: Academic, 1994.

[72] Gucluoglu, T., Duman, T. M., and Ghrayeb, A., "Antenna selection for space time coding over frequency-selective fading channels," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. iv, (Montreal), pp. 709–712, May 2004.

[73] Guey, J. C., Fitz, M. P., Bell, M. R., and Kuo, W. Y., "Signal design for transmitter diversity wireless communication systems over rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 4, pp. 527–537, 1999.

[74] H-Madsen, A., "On the capacity of wireless relaying," in *VTC 02*, 2002. presented in VTC02.

[75] H.-Madsen, A. and Zhang, J., "Capacity bounds and power allocation for wireless relay channel." submitted to IEEE Trans. Inform. Theory, 2003.

[76] Hagenauer, J., "Source-controlled channel decoding," *IEEE Transactions on Communications*, vol. 43, pp. 2449–2457, Sep 1995.

[77] Hagenauer, J., Robertson, P., and Papke, L., "Iterative (turbo) decoding of systematic convolutional codes with the map and sova algorithms," in *Proceedings of ITG Conference, vol. 130, Munich*, pp. 21–29, 1994.

[78] Han, J., Varshney, P. K., and Vannicola, V. C., "Some results on distributed nonparametric detection," in *Proc. 29th Conf. on Decision and Control*, pp. 2698–2703, December 1990.

[79] Hasna, M. O. and Alouni, M. S., "Application of the harmonic mean statistics to the end-to-end performance of transmission systems with relays," in *GlobeComm 2002*, vol. 2, pp. 1310–1314, IEEE, Nov 2002.

[80] Hasna, M. O. and Alouni, M. S., "A performance study of dual-hop transmissions with fixed gain relays," in *ICASSP*, vol. 4, pp. 2461–2465, IEEE, Apr 2003.

[81] Hasna, M. O. and Alouni, M., "Outage probability of multihop transmission over nakagami fading channels," *IEEE Communications Letters*, vol. 7, pp. 216–218, May 2003.

[82] Heath, R. and Paulraj, A., "Antenna selection for spatial multiplexing systems based on minimum error rate," in *Proc. IEEE Int. Contr. Conf.*, pp. 2276–2280, 2001.

[83] Helstrom, C. W., "Gradient algorithms for quantization levels in distributed detection systems," *IEEE Trans. Aerospace Elect. Syst.*, vol. 31, pp. 390–398, Jan. 1995.

[84] Hochwald, B. M. and Marzetta, T. L., "Unitary space time modulation for multiple-antenna communications in rayleigh fading," *IEEE Transactions on Information Theory*, vol. 46, pp. 2041–2052, Mar 2000.

[85] Hochwald, B. M., Marzetta, T. L., and Richardson, T. J., "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, Sep 2000.

[86] Hochwald, B. M. and Sweldens, W., "Differential unitary space-time modulation," *IEEE Transactions on Communications*, vol. 48, pp. 2041–2052, Dec 2000.

[87] Hong, Z., Liu, K., Heath, R. W., and Sayeed, A. M., "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE Journal of Selected Areas in Communications*, vol. 21, pp. 856–866, June 2003.

[88] H.R.HASHEMI and I.B.RHODES, "Decentralized sequential detection," *IEEE Transactions on Information Theory*, vol. 35, pp. 509–520, May 1989.

[89] HUGHES, B. L., "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, pp. 2567–2578, Nov 2000.

[90] HUNTER, T. E. and NOSRATINIA, A., "Cooperation diversity through coding," in *ISIT*, p. 220, IEEE, June 2002.

[91] H.V.POOR, *An introduction to Signal Detection and Estimation, New York, NY.* Springer-Verlag, 2 ed., 1994.

[92] I.F.AKYILDIZ, W.SU, Y. and E.CAYIRCI, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102–114, August 2002.

[93] IVRLAC, M. T., UTSCHICK, W., and NOSSEK, J. A., "Fading correlations in wireless mimo communication systems," *IEEE Journal of Selected Areas in Communications*, vol. 21, pp. 819–828, June 2003.

[94] J.CHAMBERLAND and V.VEERAVALLI, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, pp. 407–416, February 2003.

[95] JR., A. R. H. and GAMAL, H. E., "On the theory of space-time codes for psk modulation," *IEEE Transactions on Information Theory*, pp. 524–542, Mar. 2000.

[96] KAMACI, N. and ALTUNBASAK, Y., "Multiple description coding with multiple transmit and receive antennas for wireless channels: the case of digital modulation," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, pp. 3272–3276, Sept. 2001.

[97] KASSAM, S. A. and POOR, H. V., "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, pp. 443–481, March 1985.

[98] KONG, N. and MILSTEIN, L. B., "Combined average snr of a generalized diversity selection combining scheme," in *IEEE Int. Conf. on Commun.*, pp. 1556–1560, June 1998.

[99] LANEMAN, J. N., *Cooperative Diversity in Wireless Networks: Algorithms and Architectures.* PhD thesis, Massachusetts Institute of Technology, MA, Aug 2002.

[100] LANEMAN, J. N. and WORNELL, G. W., "Energy-efficient antenna sharing and relaying for wireless networks," in *Wireless Comm. and Networking Conf.*, vol. 1, pp. 7–12, IEEE, Sept 2000.

[101] LANEMAN, J. N. and WORNELL, G. W., "Distributed space-time protocols for exploiting cooperative diversity in wireless networks," in *Globecom*, vol. 1, pp. 77–81, IEEE, Nov 2002.

[102] LANEMAN, J. N., WORNELL, G. W., and TSE, D. N., "An efficient protocol for realizing cooperative diversity in wireless networks," in *ISIT*, p. 294, IEEE, June 2001.

[103] LAUER, G. S. and SANDELL, N. R., "Distributed detection of known signal in correlated noise," tech. rep., ALPHATECH, Burlington, MA, Mar. 1982.

[104] Li, Y. G., Winters, J. H., and Sollenberger, N. R., "Mimo-ofdm for wireless communications: signal detection with enhanced channel estimation," *IEEE Transactions on Communications*, vol. 50, pp. 1471–1477, Sept 2002.

[105] Lin, Y., Chen, B., and Varshney, P. K., "Decision fusion rules in multi-hop wireless sensor networks," in *Proceedings of the Seventh International Conference on Information Fusion*, pp. 828–835, Jun 2004.

[106] Liu, Z. and Giannakis, G. B., "Space-time block coded multiple access through frequency selective channels," *IEEE Transactions on Communications*, vol. 49, pp. 1033–1044, June 2001.

[107] Lu, B., Wang, X., and Narayanan, K. R., "Ldpc-based space-time coded ofdm systems over correlated fading channels:performance analysis and receiver design," *IEEE Transactions on Communications*, vol. 50, pp. 74–88, Jan 2002.

[108] Ma, X. and Giannakis, G. B., "Layered space-time complex field coding: Full diversity with full-rate, and tradeoffs," in *Sensor Array and Multichannel Signal Processing Workshop*, pp. 442–446, 2002.

[109] Marzetta, T. L. and Hochwald, B. M., "Capacity of a mobile multiple–antenna communication link in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, pp. 139–157, Jan 1999.

[110] M.D.Srinath, P. and R.Viswanathan, *Introduction to statistical signal processing with applications, Englewood Cliffs, NJ.* Prentice-Hall, 1995.

[111] M.Longo, T. and R.M.Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Transactions on Information Theory*, vol. 36, pp. 241–255, March 1990.

[112] M.M.Al-Ibrahim and P.K.Varshney, "Nonparametric sequential detection based on multisensor data," in *Proc. 23rd Annu. Conf. on Information. Sci. and Syst.*, pp. 157–162, March 1989.

[113] Molisch, A. F., Win, M. Z., and Winters, J. H., "Capacity of mimo systems with antenna selection," in *Proceedings of Int. Conf. Comm.*, pp. 570–574, 2001.

[114] Molisch, A. F., Win, M. Z., and Winters, J. H., "Space-time-frequency (stf) coding for mimo-ofdm systems," *IEEE Communications Letters*, vol. 6, pp. 370–372, Sept 2002.

[115] Muller-Weinfurtner, S. H., "Coding approaches for multiple antenna transmission in fast fading and ofdm," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2442–2450, Oct 2002.

[116] Nabar, R., Bolcskei, H., and Paulraj, A. J., "Transmit optimization for spatial multiplexing in the presence of spatial fading correlation," in *IEEE GLOBECOM*, vol. 1, pp. 131–135, 2001.

[117] Nabar, R., Gore, D., and Paulraj, A., "Selection and use of optimal transmit antennas in wireless systems," in *Proc. of Int. Conf. Telecomm. (ICT)'99*, 2000.

[118] Nabar, R. U., Bölcskei, H., and Kneubühler, F. W., "Fading relay channels: Performance limits and space-time signal design," *submitted to IEEE Journal on Selected Areas in Communications*, 2003.

[119] Niu, R., Chen, B., and Varshney, P. K., "Decision fusion rules in wireless sensor networks using fading channel statistics," in *Proc. of the 2003 Conference on Information Sciences and Systems*, March 2003.

[120] Orchard, M. T., Wang, Y., Vaishampayan, V., and Reibman, A. R., "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms," in *Proc. IEEE Int. Conf. Image Process (ICIP97), Santa Barbara,CA*, pp. 608–611, Oct. 1997.

[121] Papastavrou, J. D. and Athans, M., "Distributed detection by a large team pf sensors in tandem," *IEEE Trans. Aerospace Elect. Syst.*, vol. 28, pp. 988–994, July 1992.

[122] P.K.Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1996.

[123] Poor, H. V., *"Robustness in signal detection," in: Communications and networks, ed. I.F.Blake and H.V.Poor, New York*. Springer-Verlag, 1986.

[124] Proakis, J. G., *Digital Communications*. McGraw-Hill, 4th ed. ed., 2001.

[125] Proakis, J. G., *Digital Communications*. New York, N.Y.: McGraw-Hill, Inc., 1995.

[126] Reznik, A., Kulkarni, S. R., and Verdu, S., "Capacity and optimal resource allocation in the degraded gaussian relay channel with multiple relays," in *Proc. 40th Allerton Conf. on Commun., Control and Computing*, Oct 2002.

[127] Robertson, P., "Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, pp. 1298–1303, 1994.

[128] Robertson, P. and Woerz, T., "Novel bandwidth efficient coding scheme employing turbo codes," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 962–967, 1996.

[129] R.S.Blum, S. and H.V.Poor, "Distributed detection with multiple sensors: Part ii - advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–79, January 1997.

[130] R.Viswanathan and P.K.Varshney, "Distributed detection with multiple sensors-part i: Fundamentals," *Proceedings of IEEE*, vol. 85, pp. 54–63, Jaunary 1997.

[131] R.Viswanathan, S. and R.Tumuluri, "Optimal serial distributed decision fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, pp. 366–376, July 1988.

[132] R.W.Crow and S.C.Schwartz, "Quickest detection for sequential decentralized decision systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 267–283, January 1996.

[133] S. A. Kassam, H. V. P. and Thomas, *"Nonparametric signal detection," in: Advances in statistical signal processing-Vol. 2: signal detection, Greenwich, CT.* JAI Press, 1993.

[134] Sampath, H. and Paulraj, A. J., "Linear precoding for space-time coded systems with known fading correlations," in *Proc. 35th Asilomar Conf. Signals, Systems, Computers*, vol. 1, pp. 246–251, 2001.

[135] Sandhu, S., Nabar, R., Gore, D. A., and Paulraj, A., "Near optimal antenna selection of transmit antennas for a mimo channel based on shannon capacity," *Proc. 34th Asilomar Conference*, pp. 567–571, Nov. 1999.

[136] Sayood, K. and Borkenhagen, J. C., "Use of residual redundancy in the design of joint source-channel coders," *IEEE Transactions on Communications*, vol. 39, pp. 838–846, Jun 1991.

[137] Schelegel, C. and Costello, Jr., D. J., "Bandwidth efficient coding for fading channels: Code construction and performance analysis," *IEEE Journal of Selected Areas in Communications*, pp. 1356–1368, Dec. 1989.

[138] Sendonaris, A., Erkip, E., and Aazhang, B., "Increasing uplink capacity via user cooperation diversity," in *ISIT*, p. 196, Aug 1998.

[139] Sendonaris, A., Erkip, E., and Aazhang, B., "User cooperation diversity-part i: System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, Nov 2003.

[140] Sendonaris, A., Erkip, E., and Aazhang, B., "User cooperation diversity-part ii: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, pp. 1939–1948, Nov 2003.

[141] Seshadri, N. and Winters, J. H., "Two signaling schemes for improving the error performance of frequency-division-duplex (fdd) transmission systems using transmitter antenna diversity," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, (Secaucus, NJ), pp. 18–20, May 1993.

[142] Shannon, C. E., "A mathematical theory of communications," *Bell System Technical Journal*, pp. 1–10, Jan. 1948.

[143] Shao, X., Yuan, J., and Rapajic, P., "Antenna selection for mimo-ofdm spatial multiplexing," in *IEEE ISIT*, p. 90, IEEE, July 2003.

[144] Shiu, D. and Kahn, J., "Layered space-time codes for wireless communications using multiple transmit antennas,," in *IEEE International Conf. Telecomm. (ICC'99)*, (Vancouver, BC), June 1999.

[145] Shiu, D.-S., Foschini, G. J., Gans, M. J., and Kahn, J. M., "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, pp. 502–513, Mar 2000.

[146] Sirinavasan, M., "Iterative decoding of multiple descriptions," in *Proceedings of Data Compression Conference DCC'99*, pp. 463–472, Sept. 1999.

[147] Smith, P. J., Roy, S., and Shafi, M., "Capacity of mimo systems with semicorrelated flat fading," *IEEE Transactions on Information Theory*, vol. 49, pp. 2781–2788, Oct 2003.

[148] S.S.Pradhan, J. and K.Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, March 2002.

[149] Stefanov, A. and Duman, T. M., "Turbo coded modulation for systems with transmit and receive antenna diversity," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, vol. 5, pp. 2336–2340, 1999.

[150] Stefanov, A. and Duman, T. M., "Turbo coded modulation for wireless communications with antenna diversity," in *Proccedings of VTC-Fall*, vol. 3, pp. 1565–1569, 1999.

[151] Stefanov, A. and Duman, T. M., "Turbo coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches and practical considerations," *IEEE Journal of Selected Areas in Communications*, vol. 19, pp. 958–968, May 2001.

[152] Stefanov, A. and Erkip, E., "Cooperative coding for wireless networks," *IEEE Transactions on Communications*, vol. 52, pp. 1470–1476, Sep 2004.

[153] Stuber, G. L., *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.

[154] Su, H. and Evaggelos, G., "Space-time turbo codes with full antenna diversity," *IEEE Transactions on Communications*, vol. 49, pp. 47–57, Jan 2001.

[155] Swaszek, P., "On the performance of serial networks in distributed detection," *IEEE Trans. Aerospace Elect. Syst.*, vol. 29, pp. 254–260, Jan 1993.

[156] Tang, Z. B., Pattipati, K., and Kleinman, D. L., "An algorithm for determining the detection thresholds in a distributed detection problem," *IEEE Trans. Syst. Man Cybern.*, vol. 21, pp. 231–237, Jan./Feb. 1991.

[157] Tang, Z. B., Pattipati, K., and Kleinman, D. L., "Optmization of detection networks: Part i–tandem structures," *IEEE Trans. Syst. Man Cybern.*, vol. 21, pp. 1044–1059, Sept 1991.

[158] Tang, Z. B., Pattipati, K., and Kleinman, D. L., "Optimization of distributed detection networks: Part ii–generalized tree structures," *IEEE Trans. Syst. Man Cybern.*, vol. 23, pp. 211–221, Jan./Feb. 1993.

[159] Tarokh, V. and Jafarkhani, H., "A differential detection scheme for transmit diversity," *jsac*, vol. 18, pp. 1169–1174, July 2000.

[160] Tarokh, V., Jafarkhani, H., and Calderbank, A. R., "Space–time block codes for high data rate wireless communications: Performance results," *IEEE Journal of Selected Areas in Communications*, vol. 17, pp. 451–460, Mar. 1999.

[161] Tarokh, V., Jafarkhani, H., and Calderbank, A. R., "Space–time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[162] Tarokh, V., Naguib, A., Seshadri, N., and Calderbank, A. R., "Combined array proscessing and space–time coding," *IEEE Transactions on Information Theory*, vol. 45, pp. 1121–1128, May 1999.

[163] Tarokh, V., Seshadri, N., and Calderbank, A. R., "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, Mar 1998.

[164] Telatar, E., "Capacity of multi–antenna gaussian channels," *AT&T-Bell Labs Internal Tech. Memo.*, June 1995.

[165] Thomopoulos, S. C. and Zhang, L., "Distributed decision fusion with networking delays and channel errors," *Inform. Sci.*, vol. 66, pp. 91–118, Dec. 1992.

[166] Torabi, M. and Soleymani, M. R., "Adaptive ofdm with space-time coding and antenna selection for broadband wireless communications," in *IEEE ISIT*, p. 35, IEEE, July 2003.

[167] Tsitsiklis, J. N., "On threshold rules in decentralized detection," in *Proc. 25th IEEE Conf. on Decision and Contr.*, (Athens, Greece), pp. 232–236, 1986.

[168] Uner, M. K. and Varshney, P. K., "Distributed cfar detection in homogenous and nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 84–97, Jan. 1996.

[169] Uysal, M. and Georghiades, C., "Effect of spatial fading correlation on performance of space-time codes," *Electronics Letters*, vol. 37, pp. 181–183, Feb 2001.

[170] Vaishampayan, V., "Multiple description scalar quantization," *IEEE Transactions on Communications*, vol. 39, pp. 821–834, May 1993.

[171] Valenti, M. C. and Correal, N., "Exploiting macrodiversity in dense multihop netwroks and relay channels," in *Wireless Comm. and Networking*, vol. 3, pp. 1877–1882, IEEE, Mar 2003.

[172] van der Meulen, E. C., "Three-terminal communication channels," *Adv. Appl. Probl.*, vol. 3, pp. 120–154, 1971.

[173] VanTrees, H. L., *Detection, Estimation and Modulation Theory*, vol. 1. New York: Wiley, 1968.

[174] Veeravalli, V. V., "Comments on decentralized sequential detection," *IEEE Transaction on Information Theory*, vol. 38, pp. 1428–1429, July 1992.

[175] Veeravalli, V. V., "Decentralized quickest change detection," in *Proc. 1995 IEEE Int. Symp. Inform. Theory*, p. 294, 1995.

[176] Viswanathan, R. and Ansari, A., "Distributed detection of a signal in generalized gaussian noise," *IEEE Transactions Acoust., Speech and Signal Process*, vol. 37, pp. 775–778, May 1989.

[177] V.V.Veervalli, T. and H.V.Poor, "Decentralized sequential detection with a fusion center performing the squential test," *IEEE Transactions on Information Theory*, vol. 39, pp. 443–442, March 1993.

[178] V.V.Veervalli, T. and H.V.Poor, "Decentralized sequential detection with sensor performing the sequential test," *Math. Contr., Signals Syst.*, vol. 7, no. 2, pp. 292–305, 1994.

[179] V.V.Veervalli, T. and H.V.Poor, "Minimax robust decentralized detection," *IEEE Transactions on Information Theory*, vol. 40, pp. 35–40, Jan. 1994.

[180] Wang, Y., Orchard, M. T., and Reibman, A. R., "Multiple description image coding for noisy channels by pairing transform coefficients," in *In Proc. 1997 IEEE First Workshop on Multimedia Signal Processing, Princeton, NJ*, pp. 419–424, Jun 1997.

[181] Win, M. Z. and Winters, J. H., "Analysis of hybrid selection-maximal ratio combining in rayleigh fading," *IEEE Transactions on Communications*, vol. 47, pp. 1773–1776, Dec. 1999.

[182] Wittneben, A., "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 1630–1634, 1993.

[183] Yang, S. M. and Vaishampayan, V., "Low delay communication for rayleigh fading channels: An application of the multiple description quantizer," *IEEE Transactions on Communications*, vol. 43, pp. 2771–2783, Nov 1995.

[184] Zhao, B. and Valenti, M. C., "Distributed turbo coded diversity for relay channel," *Electronics Letters*, vol. 39, pp. 786–787, May 2003.

VITA

İsrafil Bahçeci was born in Mersin, Turkey, in 1977. He received his B.S. Degree from Bilkent University, Ankara, Turkey in 1999, and his M.S. Degree from Arizona State University, Tempe, Arizona, U.S.A. in 2001, all in Electrical Engineering. He will receive his PhD Degree in Electrical Engineering from Georgia Institute of Technology in December 2005. His research interests include multiple-input multiple-output wireless communications, distributed source/channel coding for wireless systems, signal processing for wireless sensor networks, distributed detection and estimation, and analog error correction codes. He is currently working as a Postdoctoral Fellow at the University of Waterloo, Waterloo, Ontario, Canada.