CONCATENATED CODES FOR THE MULTIPLE-INPUT MULTIPLE-OUTPUT QUASI-STATIC FADING CHANNEL

A Dissertation

by

VIVEK GULATI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Electrical Engineering

CONCATENATED CODES FOR THE MULTIPLE-INPUT MULTIPLE-OUTPUT QUASI-STATIC FADING CHANNEL

A Dissertation

by

VIVEK GULATI

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Krishna R. Narayanan (Chair of Committee)	
Costas N. Georghiades (Member)	A. L. N. Reddy (Member)
Riccardo Bettati (Member)	Chanan Singh (Head of Department)

December 2004

Major Subject: Electrical Engineering

ABSTRACT

Concatenated Codes for the Multiple-Input Multiple-Output Quasi-static Fading Channel. (December 2004)

Vivek Gulati, B. Tech.(Hons.), Indian Institute of Technology, Kharagpur, India

Chair of Advisory Committee: Dr. Krishna R. Narayanan

The use of multiple antennas at the transmitter and/or the receiver promises greatly increased capacity. This can be useful to meet the ever growing demand of wireless connectivity, provided we can find techniques to efficiently exploit the advantages of the Multiple-Input Multiple-Output (MIMO) system.

This work explores the MIMO system in a flat quasi-static fading scenario. Such a channel occurs, for example, in packet data systems, where the channel fade is constant for the duration of a codeword and changes independently from one transmission to another. We first show why it is hard to compute the true constrained modulation outage capacity. As an alternative, we present achievable lower bounds to this capacity based on existing space-time codes. The bounds we compute are the fundamental limits to the performance of these space-time codes under maximum-likelihood decoding, optimal outer codes and asymptotically long lengths. These bounds also indicate that MIMO systems have different behavior under Gaussian signaling (unconstrained input) and under the finite alphabet setting. Our results naturally suggest the use of concatenated codes to approach near-capacity performance. However, we show that a system utilizing an iterative decoder has a fundamental limit – it cannot be universal and therefore it cannot perform arbitrarily close to its outage limit.

Next, we propose two different transceiver structures that have good performance. The first structure is based on a novel BCJR-decision feedback decoder which results in performance within a dB of the outage limit. The second structure is based

on recursive realizations of space-time trellis codes and uses iterative decoding at the receiver. This recursive structure has impressive performance even when the channel has time diversity. Thus, it forms the basis of a very flexible and robust MIMO transceiver structure.

To His Holiness Sri Sri Ravishankar, and to all teachers: past, present and future.

ACKNOWLEDGMENTS

In an effort of this magnitude and taking as long as it has, though the journey is personal, one does not travel alone. Along the path one meets many who show the way – each in their own unique way, each one worthy of being acknowledged. Yet, when it comes time to put the acknowledgment in words, one only realizes the severe limitation of words.

No matter what I write in the following lines, it won't be able to fully express my gratitude to my advisor, Dr Krishna Narayanan. He is a rare blend of sharp intellect and a very personable nature. He has inspired me, nurtured me, been very patient with me and, at the same time, given me enough space to grow. It is truly a unique honor to have him as my advisor.

The head of the Telecommunications and Signal Processing Group, Dr Georghiades, leads the group by example. His insights into technical matters and the ease with which he handles administrative matters are truly admirable. He has been a constant inspiration for me.

The other committee members Dr Reddy and Dr Bettati, have always been supportive of my work. The classes that I took with them were very enjoyable. They have always been there for me whenever I needed them, for whatever I needed them. The same can also be said about the staff at A&M, be it the academic departments – the Wireless Communications Lab., Electrical Engineering; or other departments – International Student Services, University Apartments, Memorial Students Center.

My numerous friends – from the research group, from my undergraduate years, from among the Pachucos, from KEOS, from Art of Living and many that are not from any of these groups – have been my essential support system. During the times of trials and tribulations, they have cheered me, egged me on, given me strength and

yet allowed me the space to be myself and express my individuality. It is amazing how each one of them has come to my life only to enrich me!

It is hard for me to say that I acknowledge my family – my mom, my dad, my brother, Patti (my mom away from home) – they are, after all, my very own. It seems strange to say to one's own heart, "I acknowledge you – that I live because you beat?"

Yet, the only person I wish to truly acknowledge – from each particle of my being – is my spiritual master, my Guru, Sri Sri Ravishankar. With his unconditional love, infinite wisdom and child-like innocence, he has truly transformed me. Again and again, he has shown me that I am much more than what I think I am. He has kindled in me the essence of life – enthusiasm, joy, love, inner peace and the spirit of service.

It is to Him that I owe all that I am and all that I ever will be.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A signal propagating through a wireless channel experiences a multiplicative noise, or *fade*. The statistical properties of the fade depend on the frequency and bandwidth of the signal and may vary with time. In an equivalent baseband model of the channel, a fade is often represented by a circularly symmetric complex random variable with zero mean and unit variance. A common and well understood method of exploiting the randomness of the channel is to employ multiple antennas at the receiver (also called *receive diversity*). This enables the receiver to average out the bad fades, so to speak. The exponent of the signal-to-noise ratio at which the probability of a bit error at the receiver decreases is called the *diversity order* of the system. With receive diversity, it is possible to obtain a diversity order equal to the number of receive antennas [3].

In many practical scenarios, however, employing multiple antennas at the receiver is not feasible. A common example is the down-link channel in a cellular system. This has sparked interest in the possible use of multiple-antennas at the transmitter, or, transmit-diversity systems. The challenge of transmit-diversity systems is that the signals transmitted from different antennas interfere with each other at the receiver. Some of the fundamental questions one may ask are as follows. What is the capacity of a transmit diversity system? Is it more useful to avoid interference (e.g. by using antenna selection)? How does one design codes to achieve close to capacity performance? What is the diversity order that a transmit-diversity system may achieve? How are the diversity order and the capacity related? How does one design codes to achieve a given diversity order? Is this design the same as that

---

The journal model used is *IEEE Transactions on Information Theory*.

for achieving close-to-capacity performance? Is there a trade-off in complexity and achievable diversity-order? Is there a trade-off in the achievable rate and the diversity order? Do the answers to these questions change depending on how the channel changes with time or whether or not the channel is known at the receiver? How does the imperfectness of the channel estimates affect the capacity and the performance of the system?

All these questions carry over to a more general system in which multiple antennas are used at both the transmitter and the receiver. Such a system is called a multiple-input multiple-output (MIMO) system. Researchers have focused on mainly on the case where the channel is known perfectly at the receiver (also called perfect channel state information at receiver, or, perfect CSI-R). It is also common to assume one of the three cases (a) the channel remains constant for the period of one transmission (quasi-static) (b) the fades change independently from one symbol to the next (fast-fading or infinitely interleaved) (c) the channel changes in a block fashion, taking on an independent value in each block and there being multiple blocks in a transmission (block-fading). Of these, the quasi-static channel and the block fading channel (with a fixed number of blocks in a transmission) has a zero Shannon capacity and so the concept of outage capacity [4] is used. In other words, no matter what rate of transmission we choose, the channel realization may not support the rate and so we must consider the probability with which this *outage* event occurs. The fast-fading case is relatively uninteresting from the code design perspective due to the availability of large temporal diversity.

In [5, 6] the outage information rates for i.i.d. Gaussian signaling in quasi-static fading (QSF) with CSI-R have been computed. These results are often called outage capacity of the MIMO-QSFC, though there is no proof yet of the optimality of iid Gaussian signaling. In [7], Marzetta and Hochwald compute the capacity in the

absence of CSI-R and show that there is no capacity increase if the number of transmit antennas exceeds the coherence interval of the channel. They further provide insights into the structure of capacity achieving space-time codes in this scenario. The case of perfect channel information at the transmitter (perfect CSI-T) and perfect CSI-R has been treated in [8] for block-fading channels. The key result therein is that the capacity increases with SNR as the minimum of transmit and receive antennas, independent of the number of fading blocks.

This thesis deals almost exclusively with the case of quasi-static fading with no CSI-T and perfect CSI-R. Note that the capacity in this case is a random variable depending on the channel realization. In [9] it has been shown that this random capacity has a Gaussian distribution about its mean, especially as the number of transmit and receive antennas becomes large. The diversity order and the given channel realization are related through the Demmel condition number in [10]. Reference [11] relates the three variables – rate, SNR and diversity order – of the MIMO system. Zheng and Tse explore the trade-off between rate and diversity at high SNR [12]. The case of finite modulation is treated in [13–15] for quasi-static fading and in [16] for fast-fading.

The code design for the MIMO-QSFC is studied in great details in [2, 17]. It is shown that the diversity order achieved by a channel code is related to the minimum rank of the codeword difference matrices. Some hand crafted example codes have also been provided therein. Improvements in the design criterion and improved codes found using computer searches have been presented in [1, 18, 19]. The cases of large number of receive antennas and low SNRs have been treated in [20–23]. In [24], the diversity criterion is given in terms of the binary rank and techniques to achieve full diversity using convolutional codes are presented.

A different approach to achieving full spatial diversity is taken in [25] wherein a simple block code that achieves full diversity with two transmit antennas for any

constellation is proposed. This idea is extended to multiple transmit antennas using Hurwitz-Radon theory in [26]. It is shown that orthogonal designs (that achieve full spatial diversity) with linear processing maximum-likelihood decoding exist only for select cases and may incur a rate-loss. That these codes incur a capacity loss as well has been shown in [27, 28]. Many researchers have also explored the option of doing away with one or more of the three conditions – linear processing, orthogonality for all constellations and using identical constellations at all the transmit antennas – to obtain full spatial diversity for arbitrary number of transmit and receive antennas. See, for example, [29–31]. Trading off diversity of block codes for obtaining higher rate is considered in [32, 33].

Foschini *et al* have proposed the use of spatial multiplexing at the transmitter to obtain very high data rates with a MIMO system [34, 35]. The underlying principle is to arrange the output of a channel code into a space-time matrix (often called space-time *modulation*) and transmit this matrix. There have been various proposals on how this formatting should be done, for example, spatial multiplexing [34], threaded space-time codes [36] and wrapped space-time codes [37] etc. The receiver typically performs some signal processing (some form of interference cancellation) followed by decoding. Full diversity is obtained either through constraints on the channel code [24, 38] or on the formatting [37].

The space-time modulator may be replaced by a space-time block or trellis code to obtain a concatenated system. In [39], the Alamouti scheme [25] is concatenated with a turbo-TCM outer code. The performance of such a system is analyzed in [40]. The use of differential encoders on each transmit antenna is proposed in [41]. The use of recursive realizations of space-time trellis codes as inner codes in a serial concatenation is proposed in [42] and analyzed in [42, 43]. The use of recursive realizations of the codes from [24] is proposed in [44].

The fundamental questions we address in this thesis are the following. Given a fixed number of transmit (N_t) and receive (N_r) antennas, an overall rate R, a constellation $\mathcal{A}_x$ (which may differ for each transmit antenna) and an outage probability P_{out}, how does one choose the space-time code or modulator? What is the limiting performance of the system once this choice has been made? What outer code should be used so that this limiting performance may be approached? Are iterative coding techniques well suited to this problem? Is there a non-iterative system that we can apply?

A. Limiting Performance of Space-time Encoders or Modulators

The Shannon capacity of the quasi-static fading channel is zero and so we must consider the *outage probability* – that given a transmission rate, the channel cannot support this rate [4]. In [5, 6] the outage information rates for iid Gaussian signaling have been computed. These are usually taken to be upper bounds on the performance of any space-time system with finite modulations [2]. Since practical systems use finite modulations, there is a need to determine the outage capacity with a modulation constraint. This in itself is a hard problem since the capacity computation involves a maximization of mutual information over the distribution of input sequences $\mathbf{x}$, conditioned on the channel realization $\mathbf{H}$ [13]. Instead, we can compute the mutual information rates for specific space-time encoders/modulators. These serve as *achievable lower bounds* on the outage capacity of the MIMO-QSF channel. Further, our results of these computations indicate significant differences between the infinite alphabet (Gaussian signaling) and the finite alphabet case. For example, whereas in the former case iid signaling is optimal, this no longer holds in the latter case.

The added advantage of this computation is that information rate is a more fun-

damental parameter (than, say the frame error rate or the bit error rate) to quantify and compare the performance of space-time systems. This is so because the frame error rate performance of many space-time codes depends on the length of the code-word and an appropriate choice of the length may make the same code look good or bad.

B. Code Design

Once the limiting performance with a given space-time encoder or modulator is known, the next question that arises is how to achieve this limiting performance. We first investigate a serial concatenation structure that naturally arises out of the computations of the limiting performance. We find that any (serial) concatenation structure that utilizes iterative decoding is intrinsically lossy. If we want to have a *universal* design – a scheme that works well for a variety of channel realizations $\mathbf{H}$ – then iterative decoding schemes do not fit the bill. We must look for alternative designs.

One such non-iterative design is proposed. This design is based on the fact that perfect decision feedback is optimal in terms of mutual information. This structure easily obtains close to limit performance.

We also propose and analyze a robust and flexible design based on recursive realizations of space-time trellis codes. This design works well not only for the quasi-static channel but also for independent fading and block fading channels. This idea is further extensible to the case of parallel concatenation and automatic repeat request systems.

C. Thesis Organization

Chapter II deals with the constrained modulation outage capacity of a MIMO-QSF channel. Section II.B presents the problem formally and explains why capacity computation under constrained modulation is a much harder problem than the unconstrained case. In Section II.C, achievable lower bounds to this capacity are computed by considering spatial multiplexing, space-time block codes and space-time trellis codes.

In Chapter III, the problem of designing an "optimal" outer code for a serial concatenation scheme with a given inner space-time code is considered. Low density parity check (LDPC) codes are considered as potential candidates for the outer code. The results of this chapter immediately point to a fundamental limitation of iterative decoding. This limitation is explained and characterized in Chapter IV. It is shown, using EXIT charts as a tool, that iterative decoding based systems cannot be universal in the sense of approaching outage limits.

Chapter V presents a novel transceiver architecture which approaches the fundamental limits of space-time trellis codes computed in Chapter II. Two variations are presented, both of which have near capacity performance.

The idea of recursive space-time trellis codes is introduced in Chapter VI. In Section VI.B, it is shown that most space-time trellis codes have a recursive realization. A serial concatenation scheme based on these recursive realizations is analyzed in Section VI.C. Extensions of this idea to ARQ systems and channels with time variations are considered in Chapter VII.

The material in Chapter II has been presented in part in [13] and is under review in [45]. The contents of Chapters III,V are in preparation for publication in [46]. The material in Chapter VI has been published in [47].

CHAPTER II

CONSTRAINED OUTAGE INFORMATION RATES

In this chapter, we look at constrained modulation outage capacity. We show that an iid input distribution is not always optimal (unlike the unconstrained case). The actual outage capacity is hard to compute. We find achievable lower bounds by computing outage information rates assuming specific input distributions – (i) iid (spatial multiplexing) (ii) block dependence (space-time block codes) and (iii) Markov dependence (space-time trellis codes). We observe interesting differences between the constrained and the unconstrained case.

Mutual information rate between the transmit sequence and the received sequence is a more fundamental quantity than frame-error rate especially for systems whose performance depends on length (e.g. trellis codes). The former is also independent of the decoding algorithm and the outer code. Thus, our results are the fundamental limit on the performance of specific space-time codes. Mutual information rate is also a more robust measure to compare different space-time codes. We will use these limits in the subsequent chapters on code design to see how well a system performs in comparison to its ultimate limit.

A. System Model

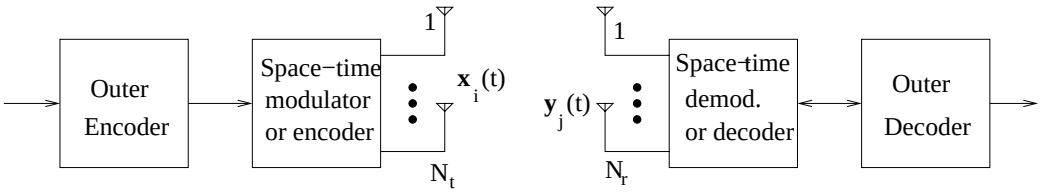

Fig. 1. System model with N_t transmit and N_r receive antennas. The various systems considered in this chapter are special cases of this scheme.

We consider a system equipped with N_t antennas at the transmitter and N_r antennas at the receiver (see Fig. 1) operating over a flat Rayleigh fading channel. Data is transmitted in frames of length N_f symbols. It is assumed that the transmitter does not know the channel and the receiver has perfect knowledge of the channel gains (that is, no CSI-T and perfect CSI-R). All channels gains are assumed to be statistically independent. While antenna separation and the propagation geometry heavily influence this assumption, it is often justified in practice. We denote the transmitted signals by $X_i(t),\ i = 1, \ldots, N_t$ and hence the complex baseband receive signals may be written as:

$$\mathbf{Y}[k] = \sqrt{\frac{\rho}{N_t}}\mathbf{H}\mathbf{X}[k] + \mathbf{N}[k]. \tag{2.1}$$

The entries of the $N_r \times N_t$ matrix $\mathbf{H}$, H_{ji}, represent the complex channel gains between receive (j)-transmit (i) antenna pairs and $N_j(t)$ is an additive white Gaussian noise (AWGN). The multiplicative term $\sqrt{\frac{\rho}{N_t}}$ is a normalization introduced in order to make the SNR, ρ, at each receive antenna independent of the number of transmit antennas. All simulation results refer to this quantity as the SNR. We assume that the noise is independent at each receive antenna. The channel gains are assumed to remain constant for the duration of transmission of a codeword and then change independently to a different realization for the next codeword.

B. Constrained Modulation Information Rates

1. Problem Formulation

Consider the problem of transmitting an $N_t \times N_f$ matrix $\mathbf{X}$ of symbols over a MIMO-QSFC. For a fixed rate of transmission R, an outage is defined as the event that the mutual information between the received matrix $\mathbf{Y}$ and $\mathbf{X}$ conditioned on the channel

realization $\mathbf{H}$ is less than R. In order to find the outage information rate, one needs to find the *a priori* distribution $p(\mathbf{X})$ that minimizes the outage probability. That is,

$$P_{out} = \min_{p(\mathbf{X})} \lim_{N_f \to \infty} \Pr\left(\frac{I(\mathbf{X};\mathbf{Y}|\mathbf{H})}{N_f} \leq R\right). \tag{2.2}$$

An alternate formulation for outage capacity may be written in terms of maximizing the rate at a given outage probability:

$$C_{P_{out}}(\rho) = \max_{p(\mathbf{X})} R \ : \Pr\left(\frac{I(\mathbf{X};\mathbf{Y}|\mathbf{H})}{N_f} \leq R\right) = P_{out},$$

where the limit has been absorbed into the probability. It should be noted that in the case of ergodic channels, the problem is that of maximizing the expected mutual information $\mathcal{E}[I(\mathbf{X};\mathbf{Y}|\mathbf{H})]$ and has been addressed in [6, 16, 48].

2. Unconstrained Modulation

Let us first consider the case when the symbols in $\mathbf{X}$ are not restricted to be from any finite sized constellation. Let the $N_t \times N_f$ transmit vector $\vec{x}$ be given by $\vec{x} = [x_{11}\ x_{12}\ \ldots\ x_{1N_t}\ \ldots\ x_{N_fN_t}]^t = [\bar{x}_1^t\ \bar{x}_2^t\ \ldots\ \bar{x}_{N_f}^t]^t$. The correlation matrix of $\vec{x}$ is given by:

$$\mathbf{Q} = \mathcal{E}[\vec{x}\vec{x}^H] = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} & \ldots & \mathbf{Q}_{1N_f} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \ldots & \mathbf{Q}_{2N_f} \\ \vdots & & & \vdots \\ \mathbf{Q}_{N_f1} & \mathbf{Q}_{N_f2} & \ldots & \mathbf{Q}_{N_fN_f} \end{bmatrix},$$

where $\mathbf{Q}_{ij} = \mathcal{E}[\bar{x}_i\bar{x}_j^H]$ is the correlation matrix between the vectors $\bar{x}_i$ and $\bar{x}_j$ transmit at times i and j. Clearly, $\mathbf{Q}_{ij} = \mathbf{Q}_{ji}^H$.

In Appendix A we outline the steps involved in proving that it suffices to choose $p(\mathbf{X})$ such that the vector of symbols $\bar{x}_k$ transmitted at time k is independent of $\bar{x}_j, j \neq k$ (i.e., temporal independence). Moreover, within each time instant, it suffices

to choose the symbols to be independent of each other (i.e., spatial independence). Hence, a diagonal correlation matrix Q (i.i.d elements in the vector $\bar{x}$) suffices to optimize the outage probability. This proof of sufficiency relies on two facts about $I(\mathbf{X};\mathbf{Y}|\mathbf{H})$:

- It has a closed form:

$$I(\mathbf{X};\mathbf{Y}|\mathbf{H}) = N_f \times \log\det(I_{N_r} + \mathbf{H}\mathbf{Q}\mathbf{H}^H), \tag{2.3}$$

  which allows for the use of Fischer's inequality [49].

- Its distribution is invariant under a unitary transformation of the input:

$$p\left(I(\mathbf{X};\mathbf{Y}|\mathbf{H})\right) = p\left(I(\mathbf{V}\mathbf{X};\mathbf{Y}|\mathbf{H})\right); \quad \mathbf{V}\text{unitary} \tag{2.4}$$

The importance of these two observations will become obvious when we attempt to prove a similar result in the constrained modulation case.

As a last remark, recall that in [6] there is a conjecture that the Q should be of the form $Q = (P/k)\text{diag}\{1,1,\ldots,1,0,0,\ldots,0\}$, where P is the power constraint and k out of N_t transmit antennas are used.

3. Constrained Modulation

In practice, the elements of $\mathbf{X}$ have to be chosen from a constellation, usually finite sized, such that it possesses the required peak to average power ratio. In this case, we attempt to shed insight into what type of $p(\mathbf{X})$ are good in terms of minimizing the outage probability defined in (2.2). We also compute $I(\mathbf{X};\mathbf{Y}|\mathbf{H})$ for some specific $p(\mathbf{X})$ when the symbols in $\mathbf{X}$ are restricted to be M-PSK symbols.

Unlike the unconstrained case (2.3), the conditional mutual information $I(\mathbf{X};\mathbf{Y}|\mathbf{H})$ does not have a closed form when $\mathbf{X}$ is restricted to a finite alphabet. Hence, it is not

always possible to use Fischer's inequality to show that it suffices to consider transmission of independent vectors in time (see Appendix A). Secondly, it is also not possible to use unitary transformations such as (2.4) since a transformation applied to the transmitted signals will change the input constellation. Therefore, in general, we cannot say anything about the optimality (or otherwise) of iid signaling from a given constellation to achieve outage capacity for the constrained modulation case. In fact, our results will show that iid signaling is *not* optimal.

One has to solve (2.2) in its most general form. The optimization must be carried out over the distribution of the $N_t \times N_f$ *matrix* $\mathbf{X}$, and over the ensemble of channel matrices $\mathbf{H}$. Further, the simplifying assumptions made in the unconstrained case do not carry over to the constrained case. The sheer dimensionality of this problem makes it quite hard.

Instead, we consider specific class of distributions of the transmit matrix $\mathbf{X}$ and compute the achievable information rate for a given outage probability and SNR. These provide a lower bound on the true outage capacity. Although the computations are mainly numerical, the results provide insight into what type of space-time constellations (or, space-time codes) provide good performance in what range of the SNR. For a wide range of rates, the computed lower bounds are close to the unconstrained case (which may be treated as an upper bound) and, hence, show near optimal space-time constellations for these range of rates.

C. Computing Constrained Modulation Information Rates

In this section, we consider three different distributions of the transmit matrix $\mathbf{X}$. First, we assume that the elements of $\mathbf{X}$ are all independent and identically distributed, taking values from a finite (say M-PSK) constellation and each value occur-

ring with equal probability. This corresponds to spatial multiplexing. Next, space-time block codes are considered. These block codes lend a certain correlation to the transmit matrix. For example, the Alamouti scheme [25] makes two adjacent transmissions dependent on each other. Finally, we turn our attention to space-time trellis codes and their punctured versions, which lend a Markov structure to the transmit matrix $\mathbf{X}$.

In each of the these cases, the achievable mutual information rate is computed numerically. The information rate conditioned on knowing the channel matrix $\mathbf{H}$ is computed first. A quantile plot of this conditional information rate is then generated for a fixed channel SNR by simulating a large number (10,000) of channels. The outage information rate is then read off this quantile plot.

1. Spatial Multiplexing: i.i.d. Transmission

Consider the case when the transmit matrix $\mathbf{X}$ has entries that are i.i.d. and drawn equally-likely from a finite alphabet $\mathcal{A}_x$. In other words, we have:

$$\Pr(\mathbf{X}) = \prod_{i=1}^{N_t} \prod_{j=1}^{N_f} \Pr(x_{ij})$$

Such a case arises, for example, when random-like codes (Turbo codes, LDPC codes etc) are transmit using spatial multiplexing. Given this correlation structure, we can consider $\mathbf{X}$ and $\mathbf{Y}$ to be, respectively, $N_t \times 1$ transmit and $N_r \times 1$ receive vectors, instead of $N_t \times N_f$ and $N_r \times N_f$ matrices. The mutual information between these vectors, conditioned on knowing the channel $\mathbf{H}$, is given by:

$$I(\vec{x}; \vec{y}|\mathbf{H}) = h(\vec{x}|\mathbf{H}) - h(\vec{x}|\vec{y}, \mathbf{H}) = h(\vec{x}) - \mathcal{E}_n[\log \Pr(\vec{x}|\vec{y}, \mathbf{H})],$$

where $h()$ represents the differential entropy function and the subscript n indicates the expectation over the noise realization. With this formulation, the *aposteriori* probability (APP) of the input given the output, $\Pr(\vec{x}|\vec{y},\mathbf{H})$, can be computed for each channel and noise realization. The average entropy associated with the APP can be computed and averaged over the noise realization and this quantity determines the resultant information rate. It should be noted that this APP computation requires an evaluation of the distance spectrum of the constrained modulation set, conditioned on the channel. When large sets are considered, for example when $N_t \geq 3$, this computation becomes numerically intensive.

Returning to the computation itself, the quantity $h(\vec{x})$ is simply $N_t \times \log(|\mathcal{A}_x|)$. The APP is obtained easily since the received vector has a multi-variate Gaussian distribution. It is easy to see that [16]:

$$\begin{aligned} I^{iid}_{N_t,N_r}(\mathbf{H}) &= N_t \log(|\mathcal{A}_x|) - \\ &\quad \frac{1}{|\mathcal{A}_x|^{N_t}}\, \mathcal{E}\left[\sum_{\vec{x}} \log \sum_{\vec{x}'} \exp\left[-\frac{1}{2}\left(\|\vec{n}\|^2 - \left\|\vec{n} - \sqrt{\frac{\rho}{2}}\mathbf{h}(\vec{x}-\vec{x}')\right\|^2\right)\right]\right], \end{aligned} \tag{2.5}$$

where the dependence of $I(\vec{x};\vec{y}|\mathbf{H})$ on N_t, N_r and $\mathbf{H}$ has been made explicit and the superscript *iid* indicates the distribution of $\mathbf{X}$. The expectation in (2.5) is over the noise vector $\vec{n}$. Monte Carlo simulations are used to evaluate the mutual information.

The first interesting result from these computations is illustrated in Fig. 2 where the densities of the achievable mutual information are plotted for BPSK and QPSK modulations and $N_r = 2$. Note that the difference in the 1% outage information rates for the two constellations is much smaller than the difference in the average information rates. Thus, when outage is of interest, it is perhaps more beneficial (in terms of receiver complexity) to use a smaller constellation per transmit antenna.

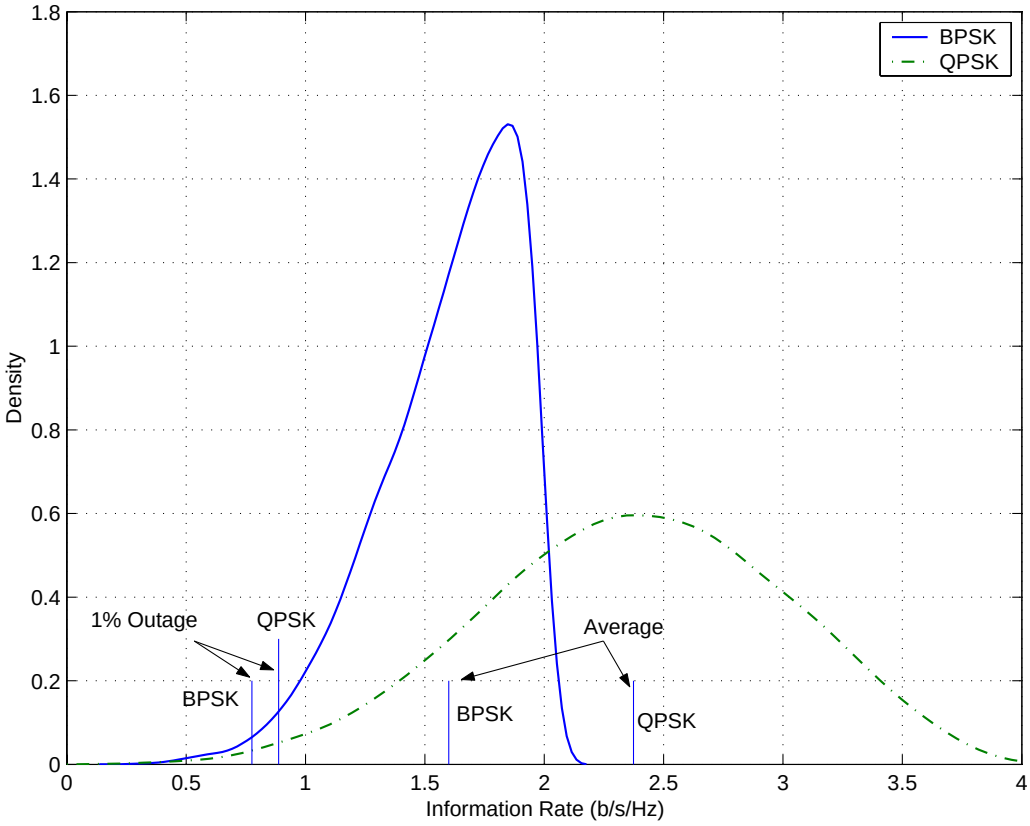

Fig. 2. The densities of the instantaneous information rate for spatial multiplexing using $N_t = N_r = 2$, Rayleigh fading, $SNR = 4$ dB.

When average mutual information is of interest, such as in the ergodic case, it makes more sense to employ as large a signal constellation as possible. More results and discussion about the achievable information rates for spatial multiplexing schemes compared to other space-time codes are presented in the following sections.

2. Space-time Block Codes

Recall that we are considering the effect upon mutual information of the distribution of the transmit matrix $\mathbf{X}$. When space-time block codes (STBCs) are used, the correlation among the entries of the transmit matrix $\mathbf{X}$ is specified by the code. For example, the Alamouti code [25], given by the code matrix:

$$\mathbf{X}(s_0, s_1) = \begin{bmatrix} s_0 & s_1 \\ -s_1^* & s_0^* \end{bmatrix},$$

makes the symbols transmitted from $N_t = 2$ antennas in $N_f = 2$ successive symbol intervals dependent on each other. In particular, we can write the 4×1 transmit vector as $\vec{x} = [s_0 \ s_1 \ -s_1^* \ s_0^*]$, where $s_0, s_1 \in \mathcal{A}_x$. This corresponds to the input distribution given by $p(\vec{x}_i) = \frac{1}{|\mathcal{A}_x|^2}$ if $\vec{x}_i = [s_0 \ s_1 \ -s_1^* \ s_0^*]$, for any $s_0, s_1 \in \mathcal{A}_x$ and $p(\vec{x}_i) = 0$ otherwise. Further $\vec{x}_i$ is independent of $\vec{x}_j$ for $i \neq j$. For other space-time block codes also a similar input distribution can be found. An alternate way to look at it is in terms of the correlation introduced by the block code. For the Alamouti scheme it is given by the Kronecker product of the matrix $\mathbf{Q}$ and the $(N_f/2)N_t$ sized identity matrix, where:

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & -a & 0 \\ 0 & -a & 1 & 0 \\ a & 0 & 0 & 1 \end{bmatrix},$$

and $a = \mathcal{E}(s^2); s \in \mathcal{A}_x$.

If $s_0(k)$ and $s_1(k)$ are transmitted in the time intervals k and $k+1$ via the code matrix above, then we have:

$$y_i(k) = \sqrt{\frac{\rho}{N_t}}\tilde{h}s_i(k) + \tilde{n}_i(k), \quad i = 0,$$

The quantity $\tilde{h} = \sum_{i=1}^{N_r}(|h_{i1}|^2 + |h_{i2}|^2)$ is the sum of the squares of the channel gains and $\tilde{n}_i(k) \sim \mathcal{N}(0, \tilde{h}); i = 0, 1$, are samples of white Gaussian noise. The information rate of this *instantaneously* AWGN channel depends only on the distribution of the effective SNR (or, equivalently, of $\tilde{h}$). For all linear processing orthogonal space-time block codes, the effective SNR has a closed form distribution (for example the Alamouti code has $\rho_{\text{eff}} = \lambda\rho/2; \lambda \sim \chi^2_{2N_r}(0.5)$). Since good closed form approximations to the constrained input capacity of the AWGN channel are available in literature [50], these may be used to find an analytical expression of the mutual information rate of such space-time block codes. For example, for the binary input AWGN channel with SNR ρ_{eff}, a good approximation of the capacity is given by [50]

$$C(\rho_{\text{eff}}) = 1 - e^{-1.24\rho_{\text{eff}}}$$

The outage capacity can then be easily computed without Monte Carlo methods by integrating the above expression over the pdf of ρ_{eff} which is $\chi^2_{2N_r}(0.5)$. In case a closed form distribution is not available, Monte Carlo methods [51] may be used.

This method is fairly general and can be used to compute the information rate of, for example, the entire class of linear processing space-time block codes [26]. In the absence of a linear processing ML decoder (for example, for the code described in [30], or the code space spanned by super-orthogonal matrices [52, 53]), the information rate can be computed by an exhaustive search over the conditional distance spectrum of the received signal. This computation is similar to (2.5) but numerically

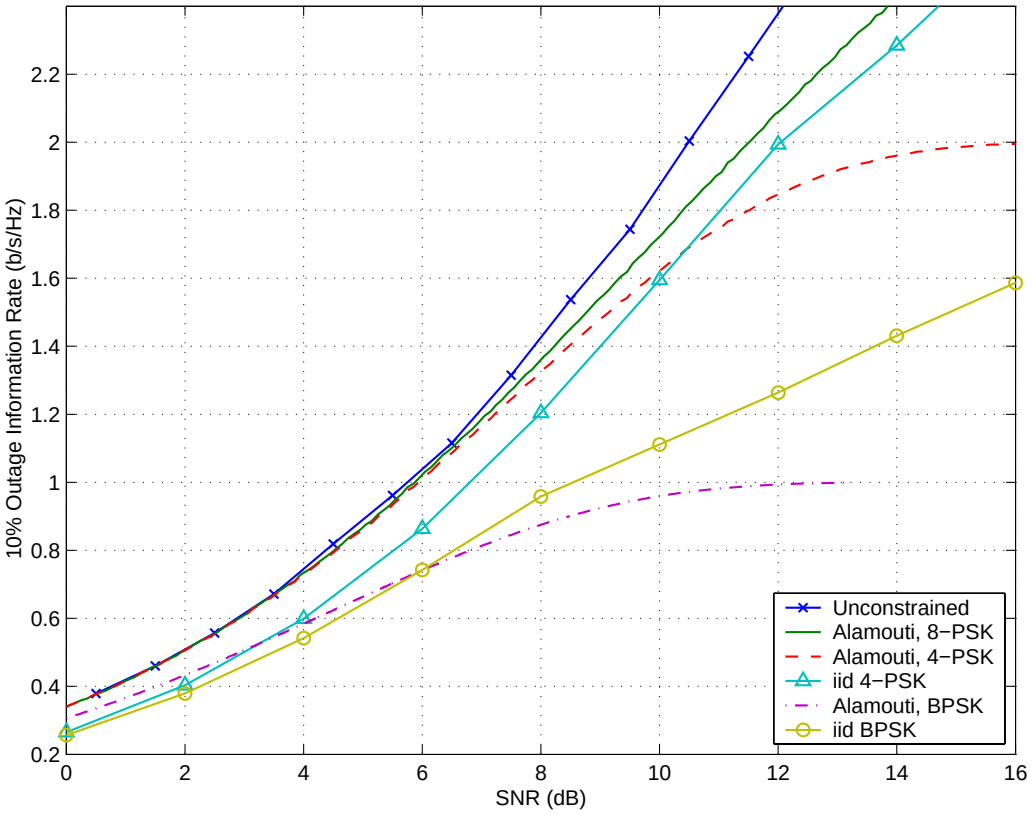

Fig. 3. Constrained modulation information rate: The space-time code is fixed to be the Alamouti scheme. The input to the Alamouti scheme is equi-probable iid M-PSK and the number of receive antennas is $N_r = 1$.

more intensive for the general block code due to the fact that multiple symbols are transmitted over multiple time slots.

It is instructive to look at the results of these computations. Fig. 3 shows the 10% outage information rates for the Alamouti scheme when different constellation sizes are used with $N_r = 1$. Note that for the entire range of rates between 0.2 b/s/Hz and 1 b/s/Hz, the QPSK and 8-PSK curves are near-optimal (almost on top of the unconstrained curve) whereas the BPSK curve is much worse. This motivates the use of expanded signal sets. That is, in order to achieve a target information rate, it suffices to expand the signal constellation used on each antenna and to use Alam-

outi's scheme. This can be thought of as the analogy of bandwidth efficient coding but applied to multiple transmit antennas. Note that the receiver complexity does not increase with increase in the constellation size and, hence, provides a practical solution.

Further, note that the iid QPSK curve is worse than the Alamouti QPSK curve up to 1.7 b/s/Hz. This shows that at low spectral efficiencies, introducing specific correlation structure (both in time and across transmit antennas) to the transmit matrix $\mathbf{X}$ outperforms a diagonal correlation structure (spatial multiplexing of iid symbols) significantly. This shows that at low rates the overall space-time code must be a concatenation of an outer code and a inner space-time code rather than being a code followed by spatial multiplexing such as in [38]. On the other hand for the range 1.7 b/s/Hz and higher, the iid QPSK outperforms the Alamouti scheme. This shows neither the Alamouti scheme nor spatial multiplexing is optimal for the entire range of rates. Contrast this with the unconstrained case where both iid and Alamouti scheme result in identical mutual information rates [28, 54] and both are optimal for the entire range of rates. The optimality of the Alamouti's scheme for the constrained case must be interpreted as follows - if the rate of transmission is significantly smaller than the maximum achievable rate, Alamouti's scheme is near optimal.

With more than one receive antennas, $N_r \geq 2$, the Alamouti scheme is no longer optimal even in the unconstrained case [28, 48]. Therefore, it is not surprising that in the constrained modulation case, there is a gap between the information rate achievable by the Alamouti scheme and the unconstrained modulation upper bound. This gap is shown in Fig. 4 for $N_r = 2$.

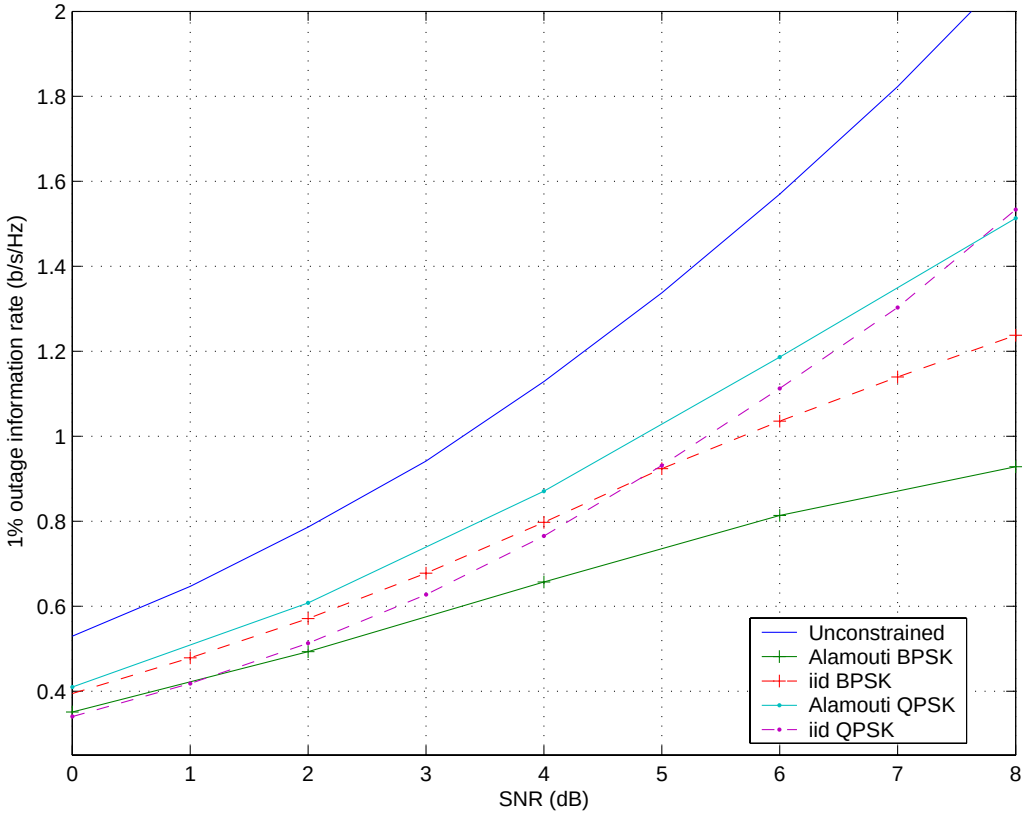

Fig. 4. 1% outage information rates for the Alamouti scheme, $N_t = N_r = 2$.

3. Space-time Trellis Codes

In this section we consider the case when the transmit matrix $\mathbf{X}$ is the output of a Markov chain. This is the case for all trellis codes. The Markov structure of the code forces the signals transmitted during a given time instant to be dependent on what was transmit during previous time instants. Consider, for example, the 4-state 4-PSK space-time trellis code from [2] (Fig. 5). Let the set of QPSK symbols be indexed by $\{0, 1, 2, 3\}$. Let x_{k1}, x_{k2} be the indices of the symbols transmit from the two antennas at time k. This 4-state code corresponds to a first order Markov chain with transition probabilities of the indices transmitted during the $k+1$-th symbol interval given by:

$$\begin{aligned} \Pr(x_{(k+1)1} = i | x_{k2} = j) &= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \\ \Pr(x_{(k+1)2} = i | x_{k2} = j) &= 0.25. \end{aligned}$$

Different space-time trellis codes correspond to Markov chains with different transition matrices. The method of computing the mutual information for a given transition matrix conditioned on the channel is explained in details in Appendix B.

In Fig. 6, we compare the outage information rates for $N_t = 2$, $N_r = 1$ with QPSK for space-time trellis codes of different constraint lengths. It can be seen that space-time trellis codes provide higher information rates than the Alamouti scheme for the medium range of rates and increasing the constraint length (higher correlation in time) provides better performance. In Fig. 7, we compare the information rates for some space-time trellis codes for QPSK modulation with $N_t = 2$ and $N_r = 2$. The space-time trellis codes compared are the AT&T codes [2], the codes by Chen, Yuan and Vucetic (CYV) [1] and super-orthogonal codes [53]. It can be seen that the CYV code provides better outage capacity performance than the AT&T code and the performance is quite close to the unconstrained outage capacity for rates below 1.5

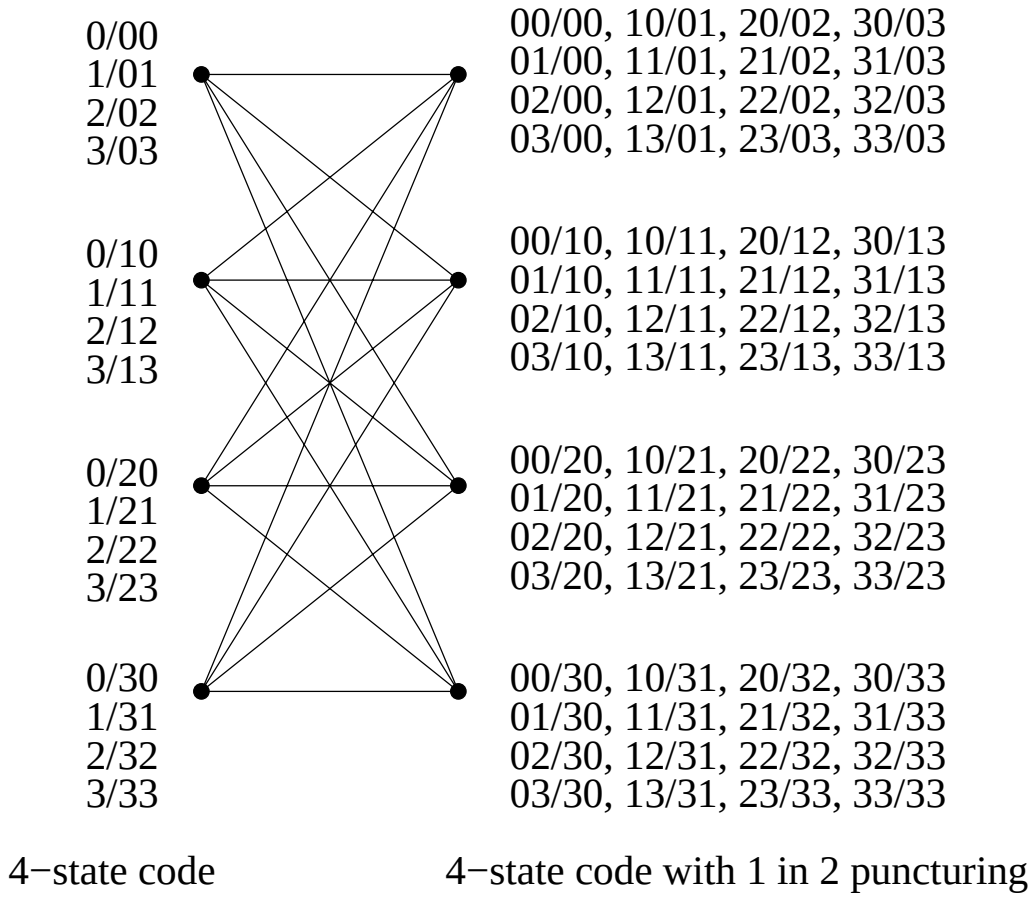

Fig. 5. The 4-state AT&T code and its punctured version ($p = 2$).

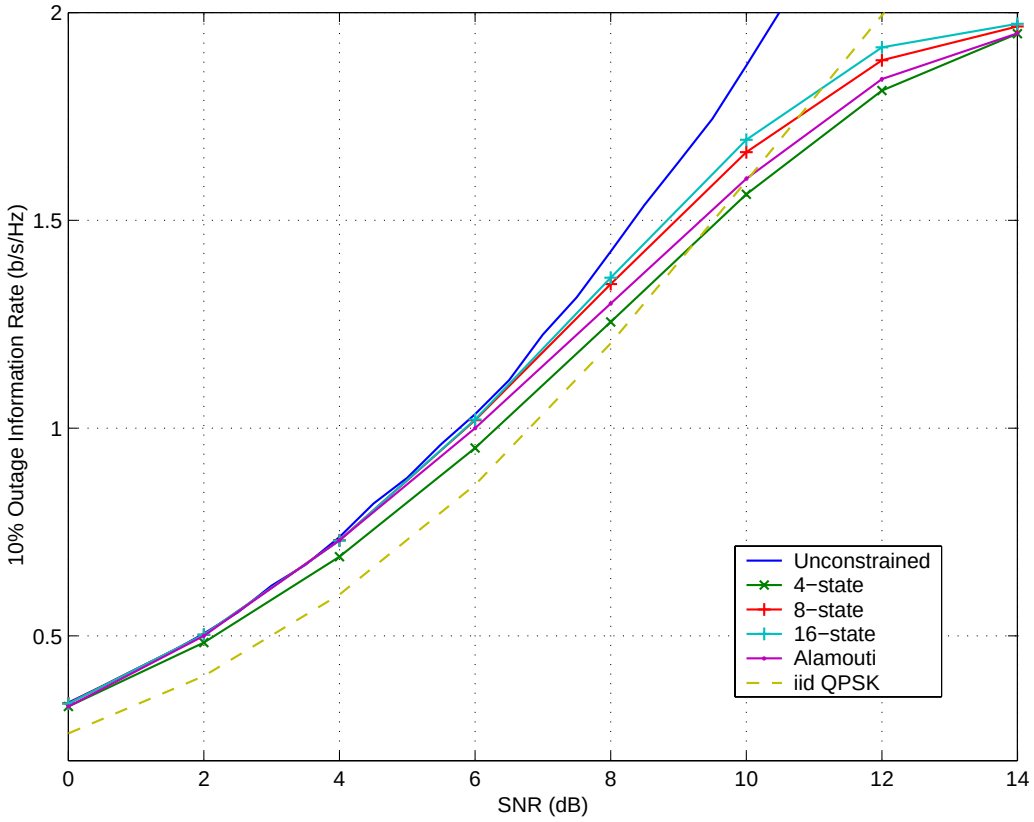

Fig. 6. 1% outage information rates of QPSK based space-time trellis codes for $N_t = 2$ and $N_r = 1$.

b/s/Hz.

4. Punctured Space-Time Trellis Codes

Here we compute information rates for some space-time trellis codes whose outputs are punctured in order to increase the data rate. Punctured space-time trellis codes also introduce a Markov structure to the transmitted signals. However, it makes the probability transition matrix time dependent. In Appendix C, we describe the modifications necessary to compute the mutual information rate for a punctured space-time trellis code.

In Fig. 8, we compare the 1% outage information rates of the 4-state CYV code

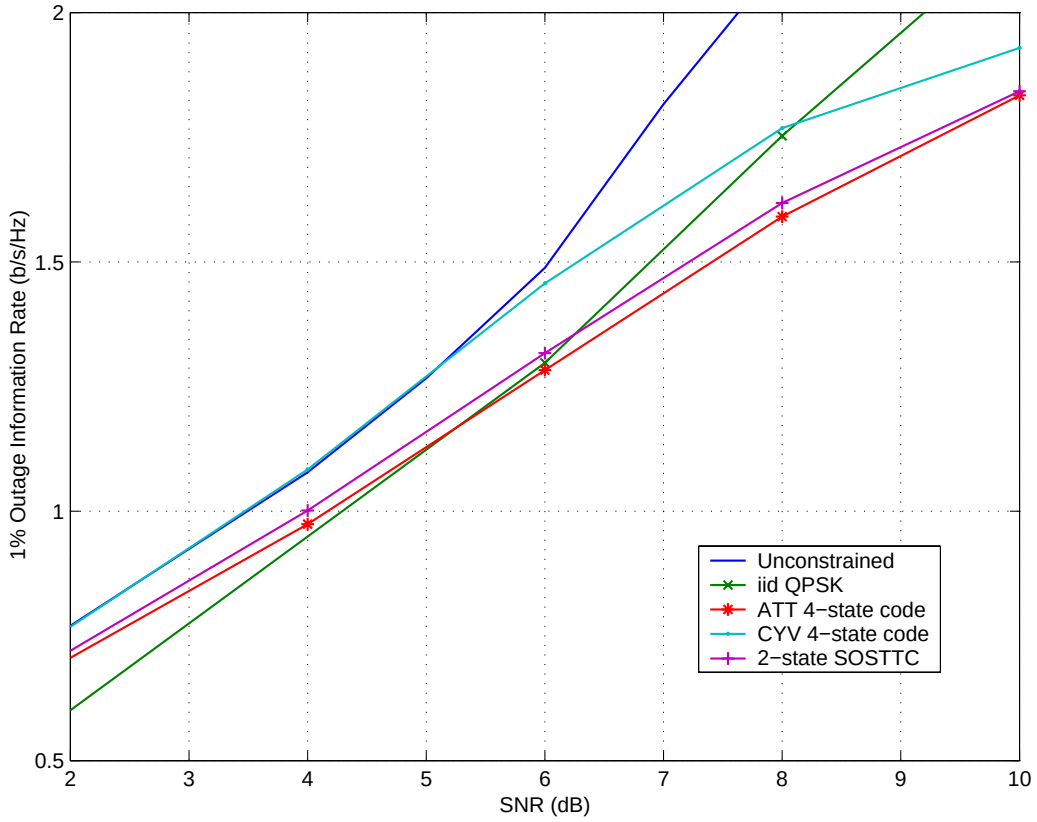

Fig. 7. A comparison of 1% outage information rates of different QPSK based space-time trellis codes for $N_t = 2$ and $N_r = 2$.

with puncturing $p = 3$ (one of three trellis stages are punctured) and $N_r = 2$. We can see that for the range of rates between 1.5 b/s/Hz and 2.4 b/s/Hz, the punctured space-time code performs better than the mother code and the spatial multiplexing. It performs fairly close to the unconstrained outage capacity. Note that the punctured space-time code does not give full diversity but still provides near optimal performance for a range of rates. This shows that for concatenated schemes with inner space-time codes, it is not essential that full diversity be provided by the space-time code. In fact, for a range of rates, it is better to use space-time codes that do not provide full diversity. The decoding complexity for punctured space-time codes is similar to that of the mother code and, hence, puncturing is a practical approach to extending the

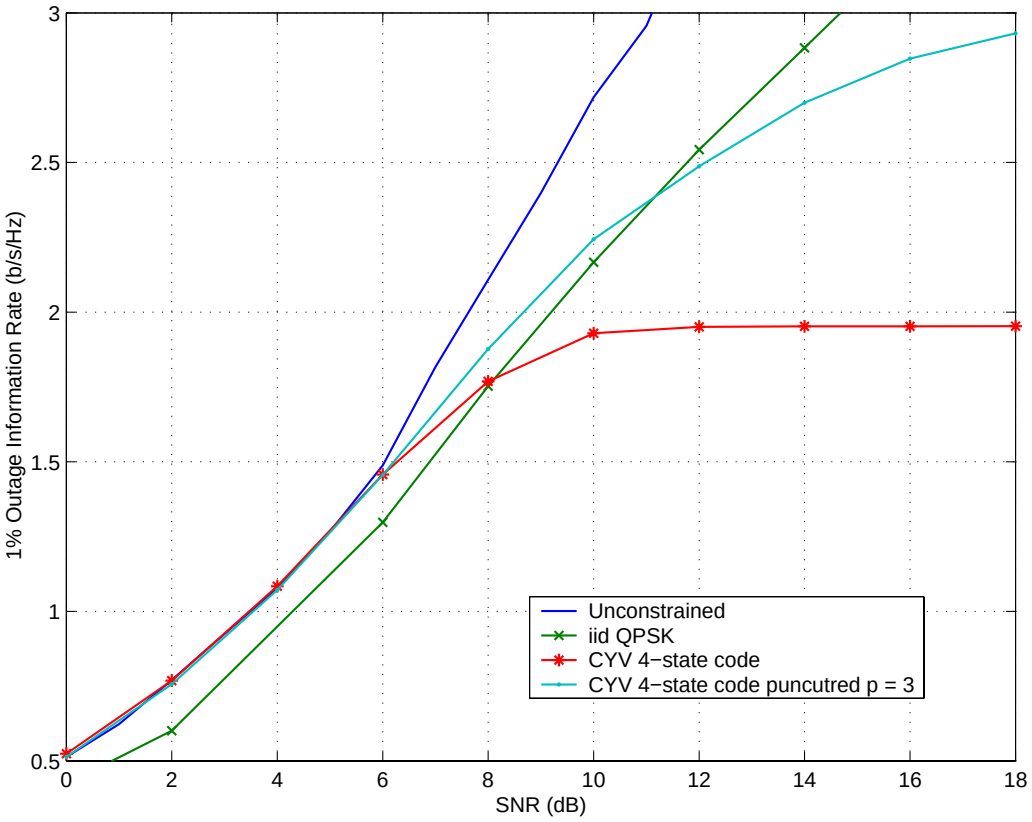

Fig. 8. 1% outage information rates for the CYV space-time trellis code and its punctured version $N_r = 2$.

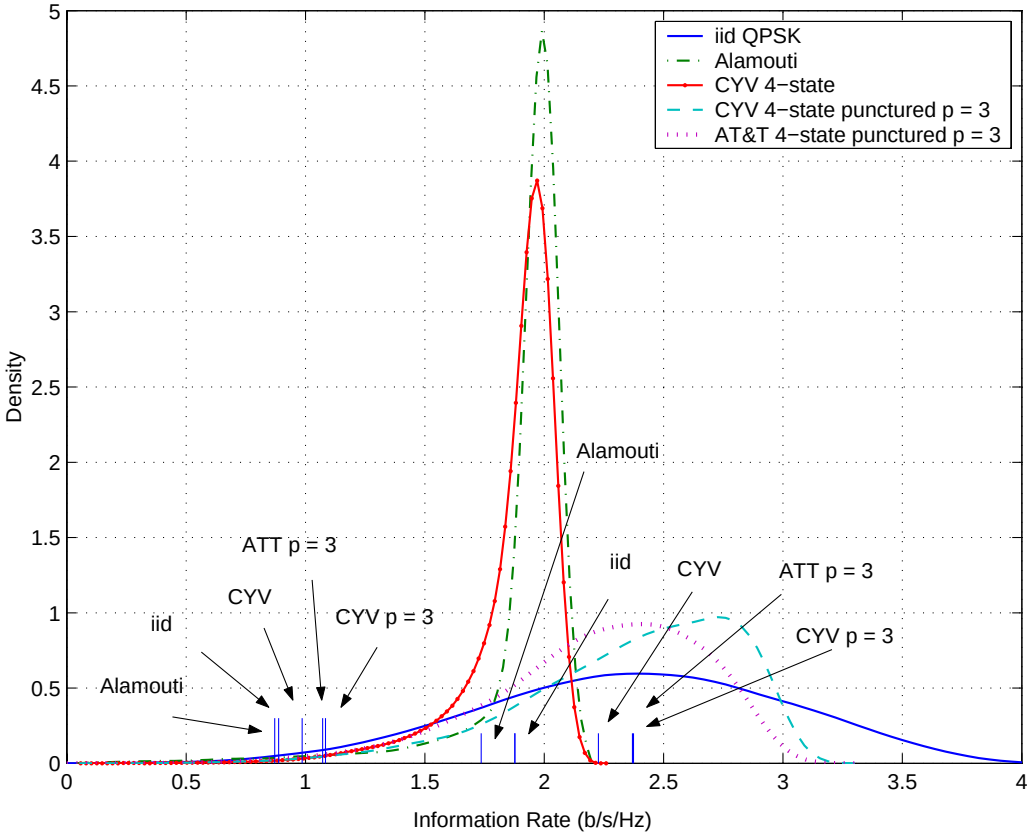

Fig. 9. The histogram of the instantaneous information rate for some space-time trellis codes and their punctured versions, Rayleigh fading, $SNR = 4$dB and $N_r = 2$.

range of rates over which space-time trellis codes perform well.

Finally, consider Fig. 9, where the densities of mutual informations achieved by different $N_t = 2, N_r = 2$ space-time codes at SNR = 4 dB are shown. Notice that if the target outage probability is small, the difference among the various space-time codes is not significant. On the other hand, when average mutual information is of interest, the choice of the space-time code being used becomes much more critical.

D. Conclusions

In this chapter we have considered different space-time coding and modulation schemes from a capacity perspective. We have shown that in the constrained modulation setting, temporal correlation among the transmit signals can achieve higher mutual information than the iid case. This is unlike the unconstrained case. The computation of the true constrained outage capacity of a multiple-input multiple-output system used over a quasi-static Rayleigh fading channel is a hard problem. This is primarily because such a computation involves determining the distribution of a matrix of transmit symbols which minimizes the outage probability. Instead, we compute the mutual information achieved by specific space-time systems.

We find that depending upon the target rate, different space-time systems are optimal in the mutual information rate sense. At low rates, space-time block codes and space-time trellis codes achieve mutual information that is close to the unconstrained case. Hence, it is wiser to use a system which has the least complexity. For a range of low-to-medium rates, imposing a Markov-like temporal correlation provides a better outage than using either block codes or spatial multiplexing. At high rates, both block codes and trellis codes offer limited choices and spatial multiplexing becomes an attractive option. A punctured space-time trellis codes does not promise full spatial diversity but it still offers near-optimal information rates for a range of SNRs. Further, just like the AWGN case, these results also indicate the advantage of using expanded constellations. In particular, for the two transmit antenna case, our results provide a strong motivation for using expanded constellations in conjunction with the Alamouti scheme.

From a practical stand-point, we have shown that the best choice (in terms of mutual information) of the space-time code to be employed depends on a number of

factors – whether outage or average mutual information is of interest, the target rate, the complexity that the receiver can afford etc.

CHAPTER III

CODE DESIGN FOR THE MULTIPLE-INPUT MULTIPLE-OUTPUT FLAT QUASI-STATIC FADING CHANNEL

In the previous chapter, we computed the constrained modulation information rates for various space-time codes. The next obvious question to ask is how to obtain system performance close to these constrained modulation limits. One frame-work that is obvious from the outage computations is that of a concatenated code (Fig. 10). We just slap an appropriate outer code on top of the inner space-time code. In this chapter, we discuss this strategy for the three space-time codes considered in Chapter II. We use low-density parity-check (LDPC) codes as outer codes. For spatial multiplexing, we find that there is a slight advantage in using codes that are designed for a channel that is a mixture of an erasure channel and an AWGN channel. When an orthogonal space-time block code, such as the Alamouti scheme, is used, an LDPC code designed for the AWGN channel is optimal and can have performance arbitrarily close to the constrained limit of the block code. In the case of space-time trellis codes, the performance of this concatenation scheme is far away from outage capacity. The reason for this gap from the limit will be the topic of the next chapter.

We begin this chapter with a brief description of the system. Thereafter, we consider spatial multiplexing, space-time block codes and finally space-time trellis codes. We conclude with a brief summary of the results from this chapter.

A. System Model and Motivation

Once again, assume that there are N_t transmit and N_r receive antennas. The channel is flat Rayleigh fading, with the channel coefficients unknown at the transmitter (no CSI-T) and perfectly known at the receiver (perfect CSI-R). The data is transmitted

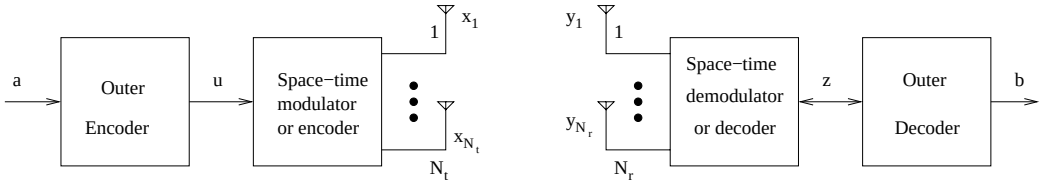

Fig. 10. A serial concatenation model with N_t transmit antennas and N_r receive antennas.

in frames of length N_f. The $N_r \times N_t$ fade coefficients are assumed statistically independent, an assumption that is often justified in practice. When linear modulation is used, the received signal at time k may be written in a matrix notation as:

$$\mathbf{Y}_k = \sqrt{\frac{\rho}{N_t}}\mathbf{H}\mathbf{X}_k + \mathbf{N}_k, \quad k = 0, 1, \ldots, N_f - 1, \tag{3.1}$$

where $\mathbf{H}$ is the $N_r \times N_t$ matrix of channel gains with complex, zero-mean, unit variance, i.i.d. Gaussian entries. The vector $\mathbf{N}_k$ is the additive noise with zero-mean, unit variance, iid complex Gaussian entries. The ensures that ρ is the received SNR at each receive antenna, independent of the number of transmit antennas.

The concatenated system is shown in Fig. 10. The primary challenge in designing an outer code that enables the serial concatenation to have a performance close to the theoretical limit is the following. A capacity achieving design would lead to a concatenated code that decodes successfully each time the channel realization can support the rate of the overall code. Conversely, the only time the decoder makes an incorrect decision is when there is an outage (i.e. channel cannot support the desired rate). Practical code design techniques, such as those using density evolution [55] or EXIT charts [56], will work well if either the design were for one channel realization or if we could average over the channel realization (i.e. ergodic case). In order to get around this problem, we make an assumption that the channels are ordered according to achievable mutual information. In other words, if a code decodes successfully for

a given channel realization, we assume that it will decode successfully for all channel realizations with a higher mutual information as well. So, our strategy is the following. We pick a code, find a channel realization whose mutual information is higher than the desired rate but the code decodes incorrectly. For this channel realization, we design a new code that would decode correctly. Finally we iterate this process.

We can design a low-density parity check (LDPC) outer code using density evolution [55] but this process is computationally involved. A much simpler, albeit less accurate, design process utilizes extrinsic information transfer (EXIT) charts in conjunction with a non-linear optimization process such as differential evolution [57, 58].

B. Spatial Multiplexing

When spatial multiplexing is done, the system looks very similar to a bit-interleaved coded modulation (BICM) system. It is shown in Fig. 11. At the encoder, the outputs of the binary rate R_o LDPC code are mapped on to some constellation (say, an $M-$PSK alphabet) and multiplexed on to the N_t transmit antennas. The overall rate is, therefore, $R = \frac{R_o N_t}{\log M}$. At the decoder, the space-time demodulator and the LDPC decoder iteratively exchange soft information[1].

The first obvious thing to do is to see how well an LDPC code optimized for the Gaussian channel works. On a closer look we find that the equivalent channel as seen by the LDPC decoder resembles a mixture of Gaussian channels. This is not surprising at all since the MIMO channel may be decomposed into independent parallel spatial channels via a singular value decomposition (SVD) of the channel matrix. Each of these spatial channels is an AWGN channel and has an effective channel SNR, given by the respective singular value.

[1]For details of these operations, please refer to chapter IV

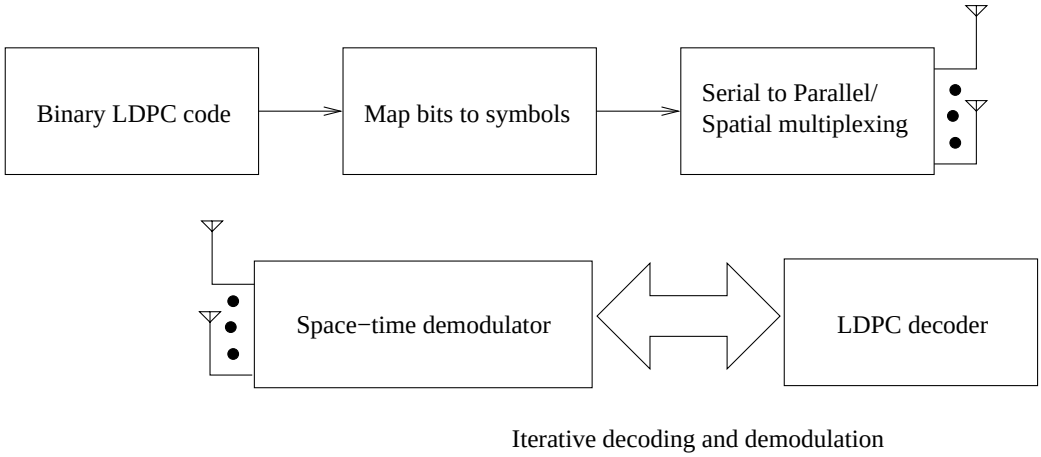

Fig. 11. Spatial multiplexing of a binary LDPC code with N_t transmit and N_r receive antennas.

In the extreme case, if the effective channel SNR for a particular sub-channel is close to zero, this sub-channel resembles an erasure channel. If all the sub-channels are "erased", we have an outage. In the $N_t = 2$ case, we can use this insight to design the outer code. We can use as the outer code an LDPC code that has been optimized for a mixture of an AWGN channel and an erasure channel. Such an optimization has been done in literature in the context of magnetic recording channels [59].

In Fig. 12, we present the results for a $N_t = N_r = 2$ system with QPSK modulation and an outer rate 1/2 LDPC code. The overall rate of this scheme is $R = 2$ b/s/Hz. The decoder performs 5 (outer) iterations between the space-time demodulator and the LDPC decoder. The LDPC decoder performs 25 iterations for each such outer iteration.

As expected, the mixture code is slightly better than the AWGN code. The performance is about 1 dB from its fundamental limit. The reason for this gap from constrained outage will be explored in the following chapter.

In Fig. 13, we use a rate 1/4 LDPC code optimized for the AWGN channel (in place of the rate 1/2 code in Fig. 12), so that the overall rate is 1 b/s/Hz. Again, the

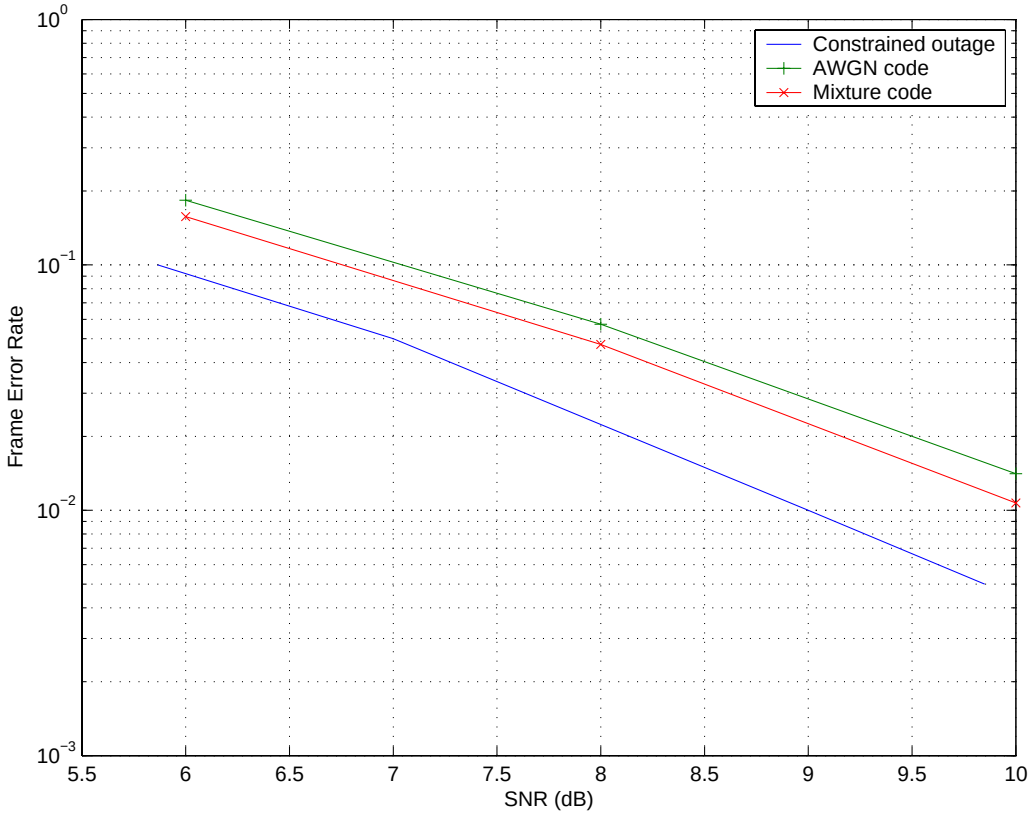

Fig. 12. Spatial multiplexing: $N_t = N_r = 2$, rate 1/2 outer LDPC code, QPSK modulation. Overall rate 2 b/s/Hz.

gap from the fundamental limit is 1–1.25 dB.

C. Space-time Block Codes

Orthogonal space-time block codes have the property that they convert the MIMO fading channel to an instantaneous AWGN channel [25, 28, 40]. The effective SNR at the output of the space-time (block) decoder equals the MIMO channel SNR scaled by the sum of the channel gains squared. It follows that codes that achieve capacity on the AWGN channel are optimal in the MIMO case as well. Furthermore, the channel as seen by the outer-code is ordered with respect to mutual information. The import of this fact will be become apparent when we discuss trellis codes in the next section.

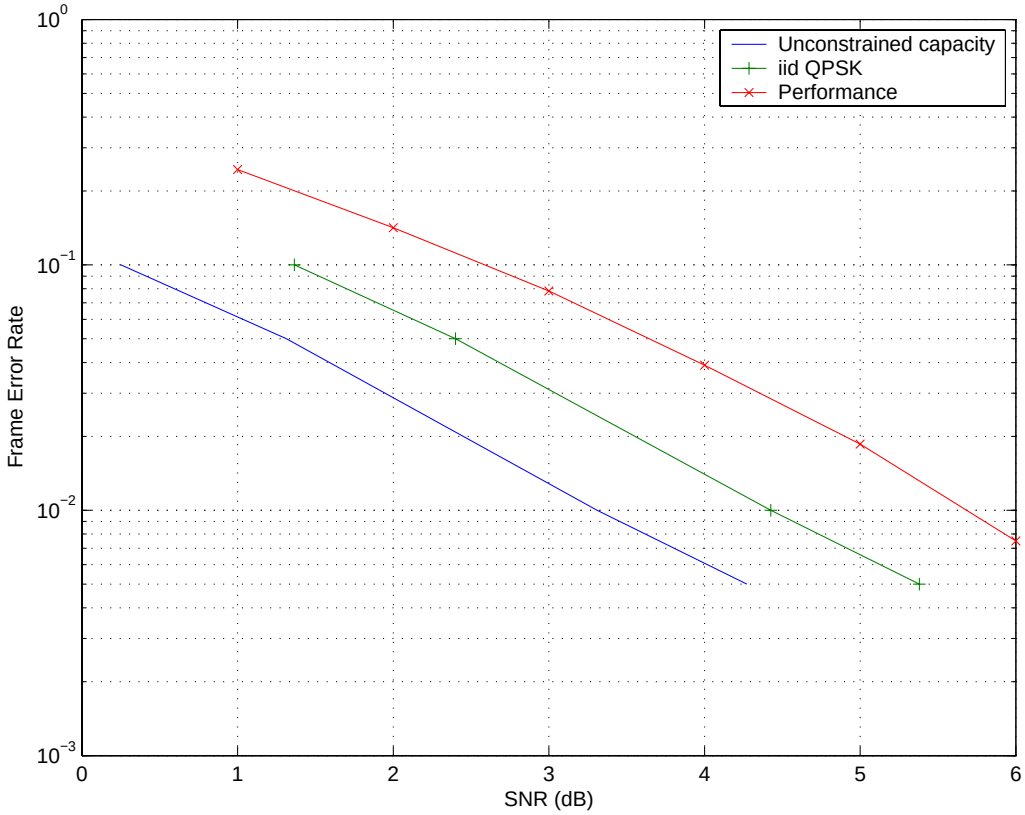

Fig. 13. Spatial multiplexing: $N_t = N_r = 2$, rate 1/4 outer LDPC code, QPSK modulation. Overall rate 1 b/s/Hz.

LDPC codes, if designed carefully, can achieve close to capacity performance on the binary input AWGN channel [60]. If a higher modulation format is employed, binary-AWGN codes can still be used with Gray mapping. If a non-Gray mapping is used, similar design procedures lead to different designs [61]. We can borrow the AWGN designs for use over the MIMO channel in conjunction with orthogonal space-time block codes.

An advantage of using LDPC codes is the ease with which the performance of this concatenation scheme can be predicted. Since LDPC codes exhibit a threshold phenomenon, we can assume that if the effective SNR ρ_{eff} is below the threshold of the code ρ_{code}, the decoder will always make an error. Otherwise, the decoder can

be assumed to decode to the correct codeword. Since the closed form distribution of the effective SNR $f_{\rho_{\text{eff}}}(x)$ is known, the codeword error rate of the concatenation is simply the integral of this distribution from the threshold of the code to infinity:

$$P_w(\rho) = \int_{\rho_{\text{code}}}^{\infty} f_{\rho_{\text{eff}}}(x)dx. \tag{3.2}$$

As a particular case, consider the Alamouti scheme where the received signal may be written as [25]:

$$\hat{s} = \sqrt{\frac{\rho}{N_t}} \left(\sum_{i=1}^{N_r} \sum_{j=1}^{2} |h_{ij}|^2 \right) s + n. \tag{3.3}$$

The noise is complex Gaussian with zero-mean and variance $\lambda = \sum_{i=1}^{N_r} \sum_{j=1}^{2} |h_{ij}|^2$. So, the effective SNR of the instantaneous AWGN channel is $\rho_{eff} = \frac{\rho\lambda}{N_t}$. Assume that a binary rate R outer code is used with an M-ary modulation. Let γ be the SNR above which we can achieve reliable communication on the M-ary input AWGN channel at a rate R b/s/Hz (i.e., capacity). The probability of frame error is given by:

$$\begin{aligned} P_f &= \Pr(\frac{E_b}{N_0} < \gamma) \\ &= \Pr(\frac{1}{R\log_2(M)} \frac{E_s}{N_0} < \gamma) \\ &= \Pr(\lambda < \frac{RN_t \log_2(M)}{\rho} \gamma) \\ &= \Pr(\lambda < \alpha) \end{aligned}$$

where $\lambda = \sum_{i=1}^{N_r} \sum_{j=1}^{2} |h_{ij}|^2$ is χ^2 distributed with $n = 2N_tN_r$ degrees of freedom and variance $\sigma^2 = 1/2$. Letting $m = n/2 = 2N_r$ (since $N_t = 2$ for Alamouti scheme), we get:

$$P_f = 1 - \exp(-\alpha) \sum_{i=0}^{m-1} \frac{\alpha^i}{i!}. \tag{3.4}$$

In Fig. 14, we show the agreement between the prediction and the performance for a concatenation of a LDPC code with the Alamouti scheme. The modulation is QPSK and both $N_r = 1$ and $N_r = 2$ are considered.

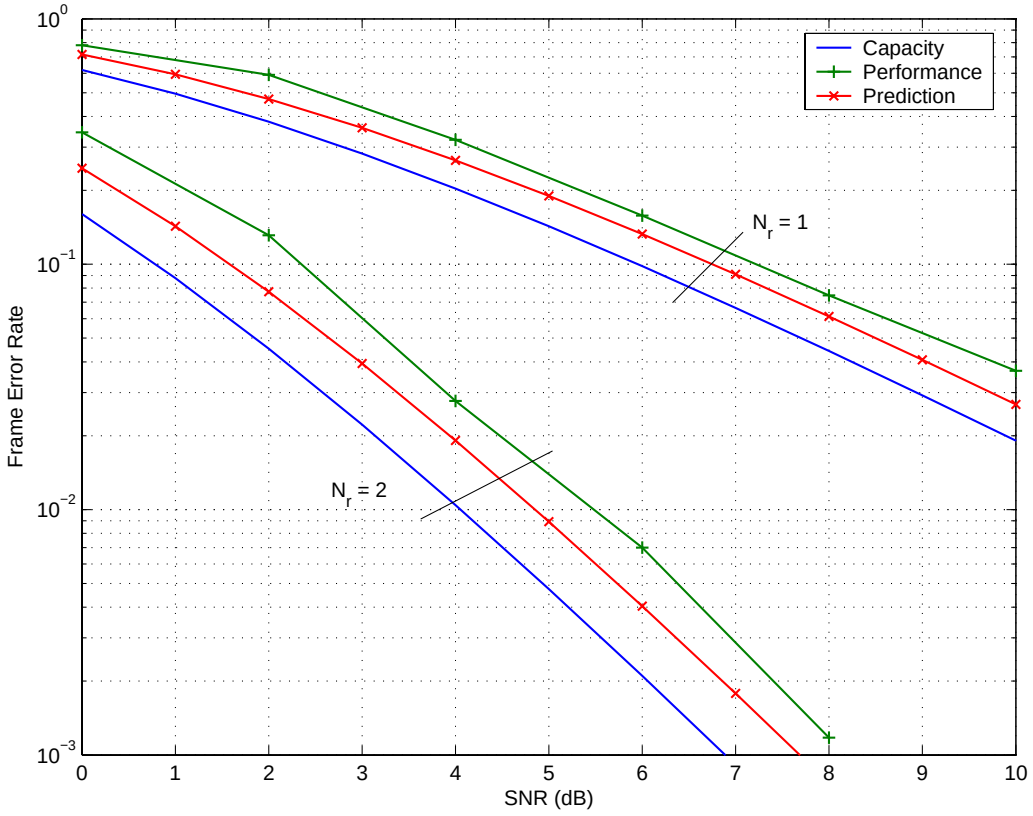

Fig. 14. Concatenation of an irregular rate-1/2 LDPC code with Alamouti scheme. The code length is 8192 bits. The code threshold is about a dB away from the BIAWGN capacity and the actual code performance is less than 0.5 dB away from the threshold. As the code length is increased, this gap between threshold and performance will close.

The mutual information rates for orthogonal space-time block codes also suggest the use of expanded signal sets. At low rates, the Alamouti scheme offers information rates that are close to the unconstrained case (which is an upper bound) when one receive antenna is used (see Fig. 3 in Chapter II). In order to achieve a target infor-

mation rate, a computationally simple way is to use expanded signal constellations on each of the transmit antennas and to use Alamouti's scheme. This can be thought of as the analogy of bandwidth efficient coding but applied to multiple transmit antennas. As an example of this, consider that we wish to transmit at 2 b/s/Hz. We save in terms of power efficiency if we use 8-PSK (with a rate-2/3 outer code) instead of (uncoded) QPSK. This is shown in Fig. 15. The important point to note is that no matter what modulation format is used, it is easy to design an outer code which will perform close to the outage information rate of the space-time block code.

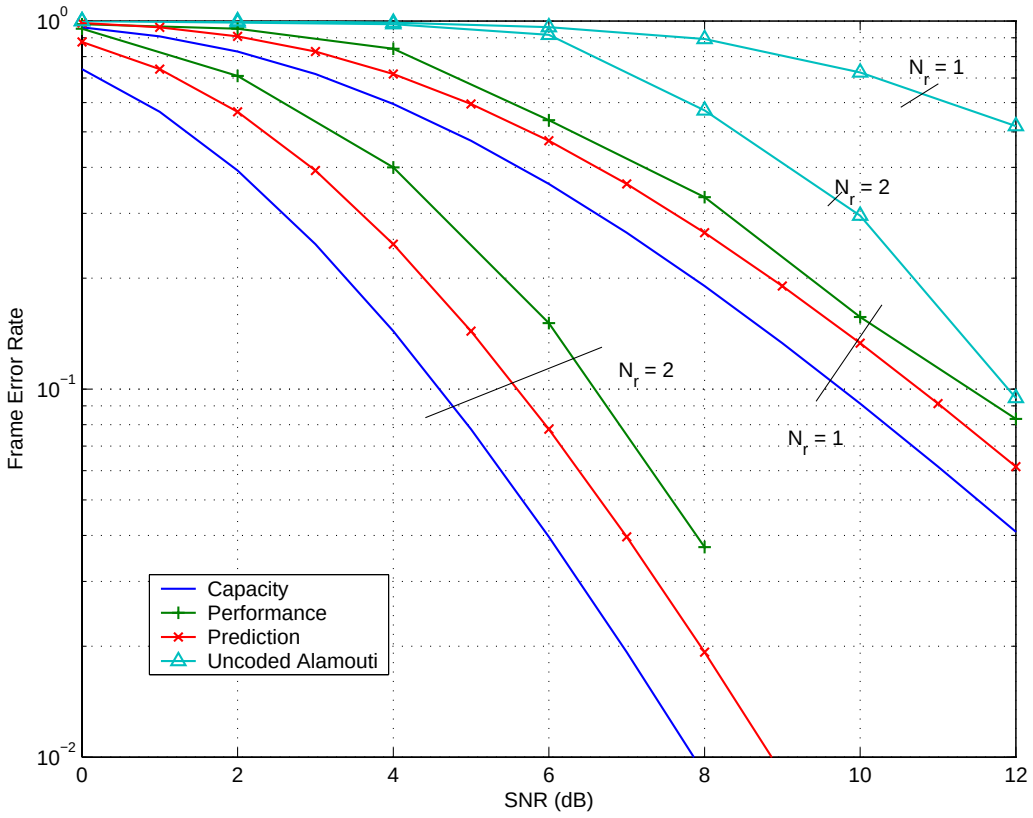

Fig. 15. Bandwidth efficient coding using Alamouti scheme: $N_t = 2$, overall rate 2 b/s/Hz. An outer code of rate 2/3, length 6144 bits is used with 8-PSK and it is compared against uncoded Alamouti with QPSK.

When more than one receive antennas are used, the Alamouti scheme (and other block codes) is no longer optimal in the capacity sense [28]. Even so it is easy to find outer codes that perform close to this (non-optimal) limit. Thus, the concatenation of LDPC codes with space-time block codes in general (and the Alamouti scheme in particular) is a practical and computationally simple method to achieve performance close to the outage limits of the specific block codes. It is simple to design good outer LDPC codes and their performance can be predicted easily and accurately. When expanded signal constellations are used, the decoder does not have to iterate between the space-time decoder and the LDPC decoder. If Gray mapping is used, then we also do not need to iterate between the symbol de-mapper and the LDPC decoder.

D. Space-time Trellis Codes

We have seen before that space-time trellis codes offer better information rates than block codes but it is more difficult to achieve this information rate. In the search for outer codes that will perform close to outage limits, we seek to design low-density parity-check codes which are matched to a particular inner code and a particular channel realization. This is to be contrasted with the ergodic channel case [62] where one can design the code based on an average over different channel realizations. We characterize the iterative decoding using extrinsic information transfer (EXIT) charts. These charts plot the input-output characteristics of the various components of the receiver (e.g. inner and outer codes) in terms of the mutual information [56].

For a particular channel realization:

$$\mathbf{H} = \begin{bmatrix} 0.904371 + 0.709265i & -0.586554 - 0.30.958i \\ -0.109425 + 0.152505i & 0.0631015 - 0.472781i \end{bmatrix}$$

with mutual information 1.367 b/s/Hz, the EXIT charts are computed at an SNR of

2 dB. The averaged trajectory of extrinsic mutual information is also plotted. The EXIT charts are seen to be optimistic by about 0.25 dB. This is shown in Fig. 16.

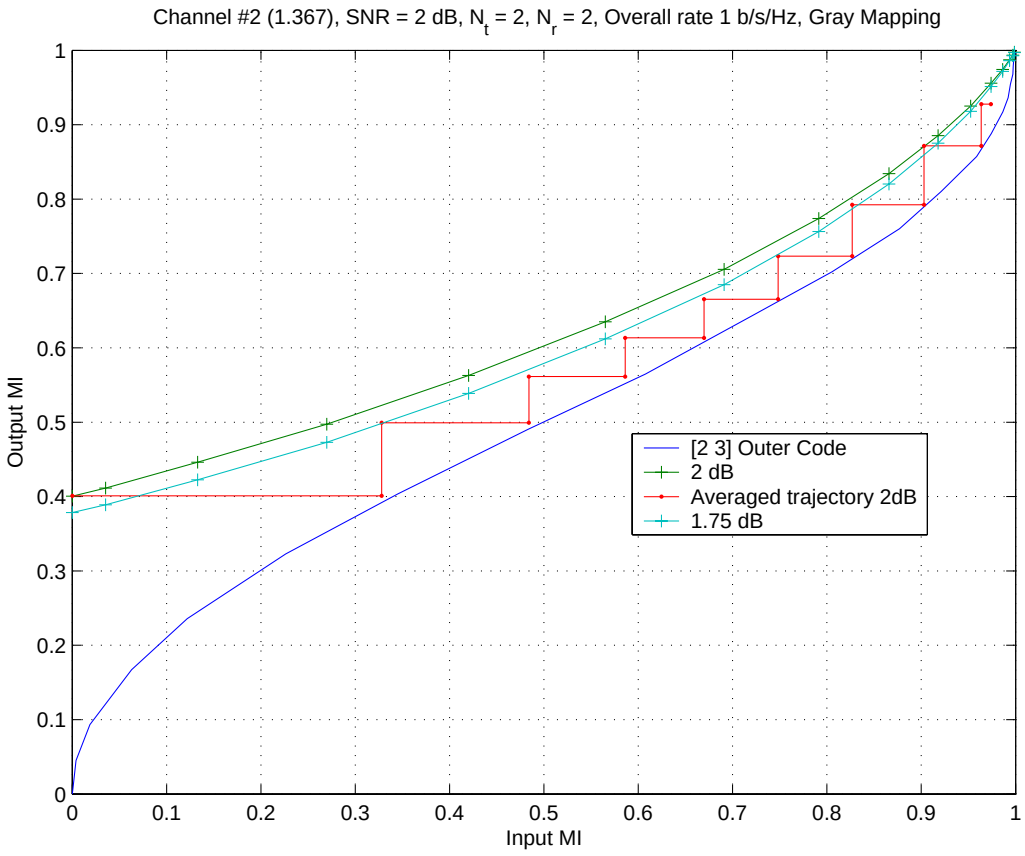

Fig. 16. EXIT charts for a serial concatenation scheme with [2 3] outer convolutional code and an inner recursive space-time trellis code. As shown, these charts are slightly optimistic.

We observe that the achievable mutual information does not provide a strict ordering of the channels. In other words, it is possible for the decoder to converge when the channel realization has a low mutual information and have a decoding failure when the channel has a higher mutual information. This is surprising since one would expect the achievable mutual information to be a robust and ordered

measure of performance. The robustness of this (or any other) measure is a property of the channel model and this is seen in the ergodic channel case (where the mutual information *is* ordered). Other measures like condition number, Demmel condition number, least singular value etc also show similar "non-ordering" behavior.

This "non-ordering" arises from the fact that the space-time trellis code itself is an information rate transformer. This transformation, however, is non-linear in the following sense. Consider two channel realizations $\mathbf{H}_1, \mathbf{H}_2$, with mutual information rates R_1 and R_2; $R_1 > R_2$. After this transformation is applied $R_1' = f(R_1, \mathbf{H}_1); R_2' = f(R_2, \mathbf{H}_2)$, there is no guarantee that $R_1' > R_2'$. This is precisely the reason for the non-ordering behavior. In other words, if, under maximum likelihood decoding, the concatenation of an LDPC code with the space-time trellis code decodes successfully on channel 1 and not on channel 2, it is necessary that $R_1' > R_2'$ but it is not necessary that $R_1 > R_2$. The problem is now compounded by the fact that the decoding is sub-optimal. In this case even the ordering $R_1' > R_2'$ under the transformed mutual information rate may also not be preserved. This is shown with the help of an example in Fig. 17. We have considered a serial concatenation scheme with a rate-1/2 [2 3] outer convolutional code and recursive inner space-time CYV code. At a given SNR (say 2 dB), we plot the pdf of mutual information between u and z (refer Fig. 10), $I_{inner}(\rho, \mathbf{H}) = I(u; z|\mathbf{H})$. However, we make a distinction between the channel realizations $\mathbf{H}$ where the decoding is successful and where it is not. As seen in Fig. 17, there is an overlap between the two pdfs. In other words, there are instances where data sent over a channel realization with lower $I_{inner}(\rho, \mathbf{H})$ is decoded correctly while data sent over a channel with a higher $I_{inner}(\rho, \mathbf{H})$ is not decoded correctly. A similar behavior is observed with other measures such as highest singular value, condition number, Demmel condition number etc.

These observations reinforce our claim in Chapter II that the conditional mutual

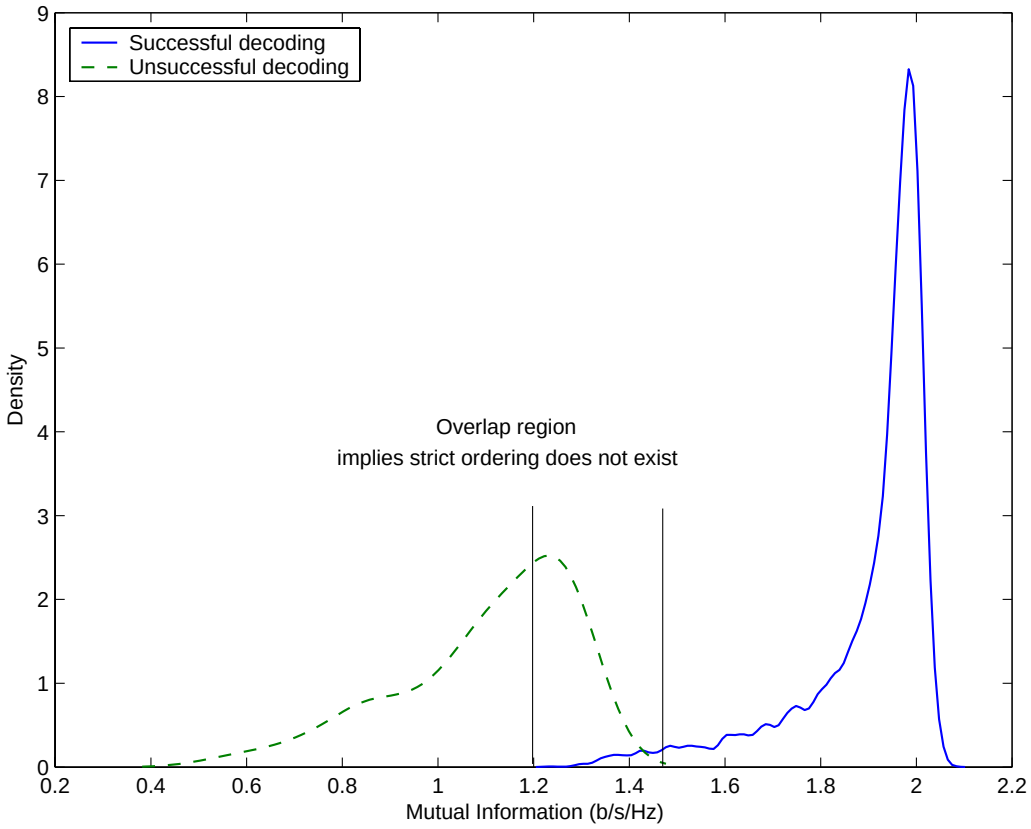

Fig. 17. Mutual information does not impose strict ordering.

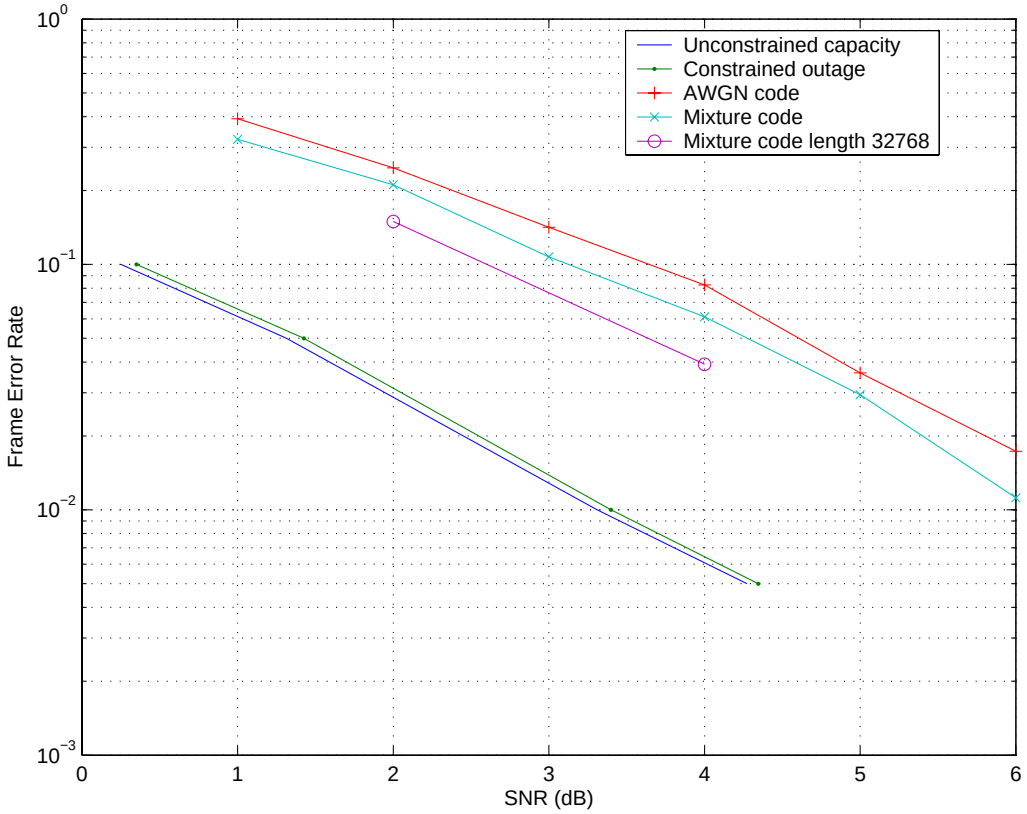

Fig. 18. Different rate 1/2 outer LDPC codes concatenated to a 4-state, 4-PSK space-time trellis code [1]. Overall rate is 1 b/s/Hz.

information (of the effective channel) is a more fundamental quantity than either the frame error rate or the Gaussian mutual information rate of the channel.

Finally, we compare the concatenation of different outer LDPC codes with the same inner space-time trellis code (from [1] and [2]) to their ultimate limits. The LDPC codes chosen are an AWGN optimized code [63] and a code optimized for the a Gaussian channel with erasures (marked Mixture code) [59]. The length of the LDPC code chosen is 8192. The results are presented in Fig. 18-19. We see that there is a significant gap of about 2.5 dB between the performance and the outage limit with the CYV code and of about 2 dB with the delay diversity code. There is only

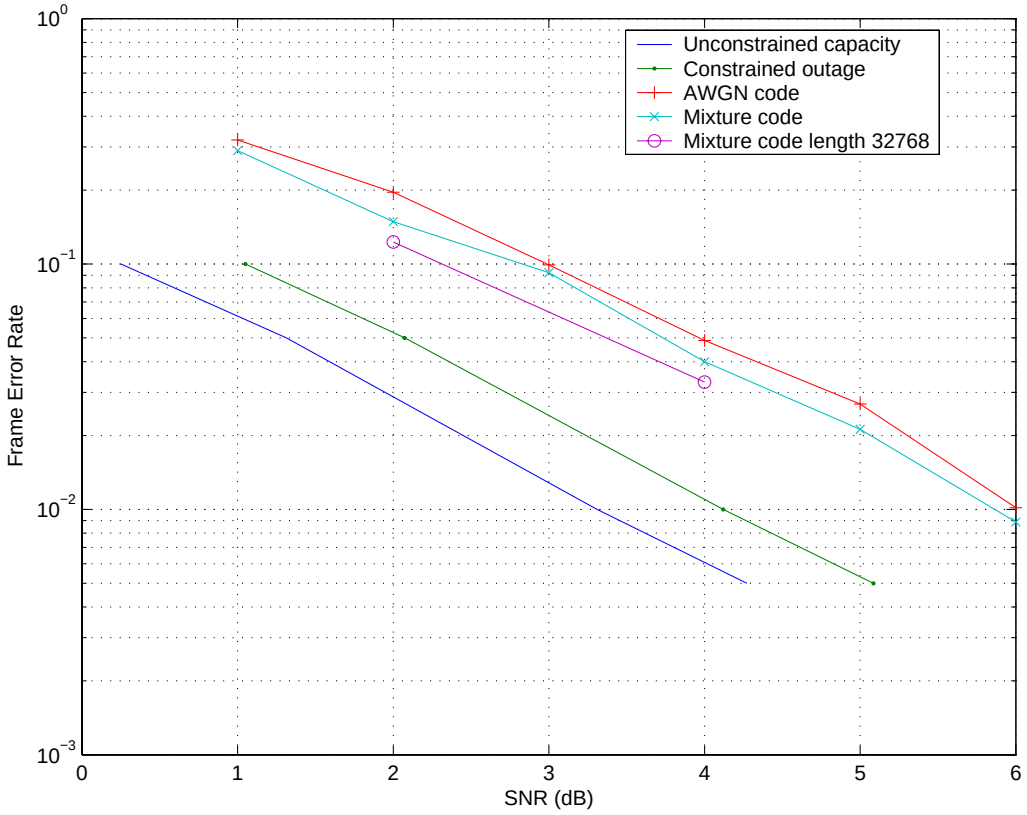

Fig. 19. Different rate 1/2 outer LDPC codes concatenated to a 4-state, 4-PSK delay diversity trellis code [2]. Overall rate is 1 b/s/Hz.

a slight improvement when the LDPC code is changed (from one optimized for the AWGN channel [63] to the one optimized for an AWGN channel with erasures [59]). Also, increasing the length (to 32768) gives very little improvement (about 0.2 dB). The reason for this significant gap is the fundamental limitation of iterative decoding, which is explored in more details in the next chapter.

E. Conclusions

In this chapter we have considered the serial concatenation structure, consisting of an outer LDPC code and an inner space-time code. When the inner space-time code is a simple spatial multiplexer, we find that LDPC codes optimized for a Gaussian channel with erasures performs slightly better than a code optimized for the pure Gaussian channel. While this can be explained by the fact that the equivalent channel (as seen by the outer decoder) is a mixture channel, this method does not give much insight into a general design with more than 2 transmit antennas. We do not know apriori whether such designs will exist or not. These issues point toward a fundamental limit of iterative decoding which will be discussed in the next chapter.

Space-time block codes, inspite of their capacity limitations, are much easier to handle since they reduce the MIMO channel into an equivalent AWGN channel. We are able to predict the performance easily. Furthermore, bandwidth efficient coding is achieved easily (and this is similar to the concept of TCM in AWGN channels). Specifically for the case of one receive antenna and at low rates, this scheme is near-optimal among all space-time systems. This is so because first the outage capacity is close to the unconstrained case and second we are able to achieve this capacity. For multiple receive antennas, the latter part of the statement still holds.

In the case of space-time trellis codes, the code design problem is compounded

by the fact that under sub-optimal (iterative) decoding, the conditioned mutual information is not ordered. This again points to a fundamental limitation of iterative systems which is explored in the next chapter. Furthermore, we must look for alternative transceiver structures to bridge the gap between performance and outage limit. Such an alternative design is presented in Chapter V.

CHAPTER IV

LIMITS ON ITERATIVE DECODING

In Chapter II, we computed the outage limits under a constrained modulation setting. Theory tells us that if we have a "good" code design and we do maximum likelihood decoding, we should be able to approach the outage limits computed before. In practice, however, maximum likelihood decoding is not always computationally feasible. Often times an iterative set up is employed. In the background material for this chapter, we first explain what is iterative decoding. A commonly employed technique to study any iterative set-up is the extrinsic information transfer (EXIT) chart. We explain briefly the concept of an EXIT chart and the assumptions commonly made in using this technique. Next we employ this technique to study iterative decoding of a concatenated space-time system. We find that iterative decoding has a fundamental limitation – it cannot be *universal*. In other words, on quasi-static channels, a single outer code cannot give near capacity performance with iterative decoding.

A. Background

1. Iterative Decoding and Demodulation

We can consider the channel code and the space-time modulator as a serial concatenation (Fig. 20). The binary data $u_k, k = 1, 2, \ldots, N_u$ is encoded by the channel code to produce the binary sequence $c_k, k = 1, 2, \ldots, N_c$, where $N_c = N_u/R_o$ and R_o is the rate of the channel code. This coded sequence is mapped to an $M-$ary sequence of symbols $x_k, k = 1, 2, \ldots, N_m$, where $N_m = \frac{N_c}{m}$, $m = \log M$ and $x_k \in \mathcal{A}_x$. This sequence is then serial to parallel converted to obtain transmit vectors $\mathbf{x}_k, k = 1, 2, \ldots, N_f$, where $N_f = (N_m/N_t) = \frac{N_c}{mN_t}$, N_t is the number of transmit antennas and $\mathbf{x}_k \in \mathcal{A}_x^{N_t}$. So, the

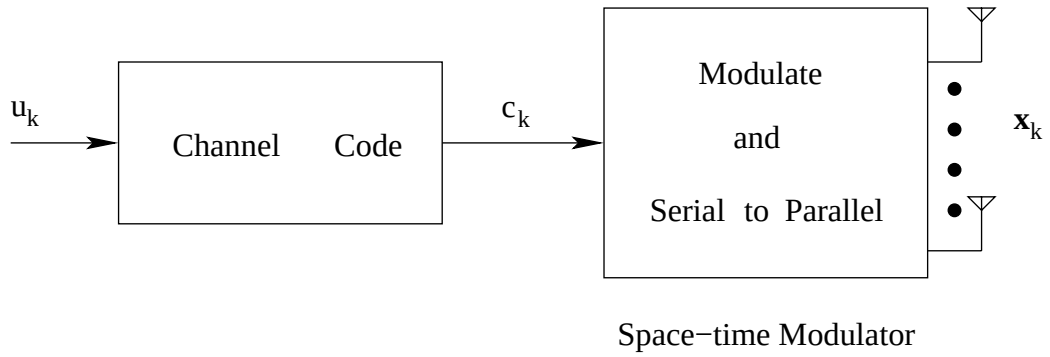

Fig. 20. Space-time modulation.

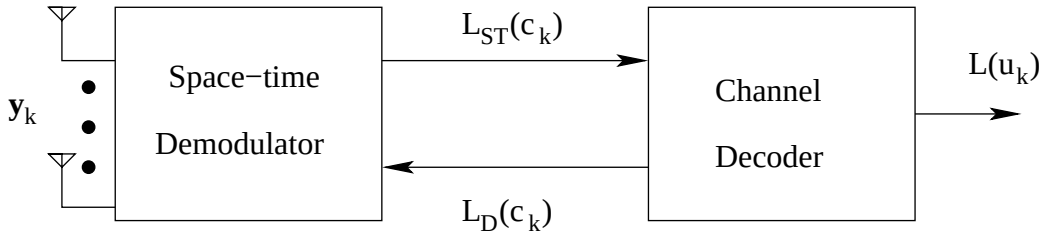

Fig. 21. Iterative decoding of space-time modulation.

overall rate of the space-time system is $R = R_o \times N_t \times m$ bits per channel use, where $m = \log M$.

The receive vector at time k is given by $\mathbf{y}_k = \sqrt{\frac{\rho}{N_t}}\mathbf{H}\mathbf{x}_k + \mathbf{n}_k; k = 1, 2, \ldots, N_f$, where $\mathbf{H}$ is the $N_r \times N_t$ matrix of channel coefficients, ρ is the SNR and $\mathbf{n}_k$ is the $N_r \times 1$ noise vector. As before, all components of $\mathbf{H}$ and $\mathbf{n}$ are independent and have a $\mathcal{CN}(0, 1)$ distribution.

The receiver operates in an iterative fashion. The space-time demodulator looks at the receive vector $\mathbf{y}$ and the apriori information about the bits c_k (which is obtained from the decoder, and is zero initially), and generates an extrinsic information $L_{ST}(c_k)$ about the bits c_k. The channel decoder treats this extrinsic information as apriori

information and generates additional information $L_D(c_k)$ about the coded bits c_k. This $L_D(c_k)$ is treated as apriori by the demodulator in the next iteration. This simple procedure for information exchange is shown in Fig. 21. The mathematical details of the demodulator operation are described below.

Consider that the coded bit-vector $\mathbf{c} = [c_1 \ \ldots \ c_{mN_t}]; c_i \in \{0,1\}$ is divided into N_t groups of m bits each and then modulated to form the symbol vector $\mathbf{x} = [x_1 \ x_2 \ \ldots \ x_{N_t}]; x_i \in \mathcal{A}_x$. There are $\mathcal{A}_x^{N_t}$ possible such vectors and if $\mathbf{y}$ and $\mathbf{H}$ are known, the probability of $\mathbf{x} = \vec{x}$ is given by:

$$\log \Pr(\mathbf{x} = \vec{x}|\mathbf{y}, \mathbf{H}) = -\frac{1}{2}||\mathbf{y} - \sqrt{\frac{\rho}{N_t}}\mathbf{H}\vec{x}||^2 + \text{constant}. \tag{4.1}$$

At the demodulator, let the apriori log-likelihood ratio be denoted by $L_{ap}(c_k)$, so that:

$$L_{ap}(c_k) = \log \frac{\Pr_{ap}(c_k = 0)}{\Pr_{ap}(c_k = 1)},$$

or, conversely,

$$\log \Pr{}_{ap}(c_k = 0) = \frac{\exp\left(L_{ap}(c_k)\right)}{1 + \exp\left(L_{ap}(c_k)\right)}; \quad \log \Pr{}_{ap}(c_k = 1) = \frac{1}{1 + \exp\left(L_{ap}(c_k)\right)}.$$

The extrinsic information generated by the demodulator is then given by:

$$\begin{aligned} L_{ST}(c_k) \ = \ & \sum_{\vec{x}:c_k=0} \left[\log \Pr(\mathbf{x} = \vec{x}|\mathbf{y}, \mathbf{H}) + \prod_{i=1;i\neq k}^{mN_t} \log \Pr{}_{ap}(c_i = \vec{c}_i) \right] \\ & - \sum_{\vec{x}:c_k=1} \left[\log \Pr(\mathbf{x} = \vec{x}|\mathbf{y}, \mathbf{H}) + \prod_{i=1;i\neq k}^{mN_t} \log \Pr{}_{ap}(c_i = \vec{c}_i) \right], \end{aligned} \tag{4.2}$$

where $\vec{c}_i$ is the $i-$th component of $\vec{c}$.

2. EXIT Charts

In order to study the behavior of this iterative demodulator-decoder, we use the technique of transfer function charts. The two components (the space-time demodulator

and the outer-code-decoder, in our case) are treated as amplifiers/transformers. The quantity that we track can be the signal-to-noise ratio (SNR) or some monotonic function of the LLRs [64]. A robust measure commonly employed in such studies is the mutual information between the binary data c_k and the log-likelihood ratios $L_{ap}(c_k)$ and $L_{ST}(c_k)$, i.e.,

$$I_{A,ST} = I\left(c_1^{mN_t}; L_{ap}(c_k)\right); \quad I_{E,ST} = I\left(c_1^{mN_t}; L_{ST}(c_k)\right). \tag{4.3}$$

The plot of I_A against I_E is called an EXtrinsic Information Transfer (EXIT) chart [56]. In order to compute the EXIT chart, one must track the probability densities of $L_{ap}(c_k), L_{ST}(c_k)$ etc.

We can now easily track the progress of the iterative set-up as follows. We first plot the EXIT chart of the inner module (the space-time decoder in this case). Next, we super-impose the EXIT chart of the outer module with its axis reversed. A staircase plot tracks the actual evolution of mutual information as iterations proceed. If these two plots cross each other, then there is a fixed-point in the iterations and the decoder will not converge. If these plots do not converge, then the decoder will converge to a codeword. This is illustrated in Fig. 22.

a. Gaussian Assumption

Usually, the apriori LLRs $L_{ap}(c_k)$ are assumed to have a Gaussian distribution with variance that is twice the mean [56]:

$$L_{ap}(c_k) \sim \mathcal{N}(c_k m, 2m).$$

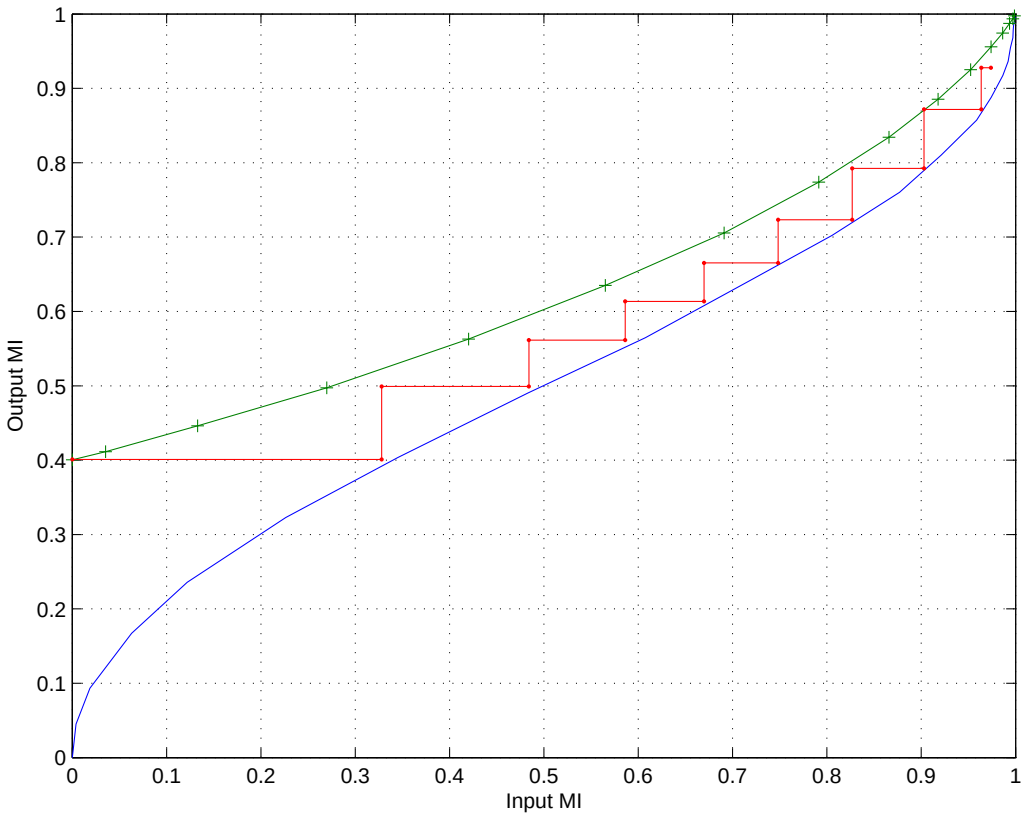

Fig. 22. Progress of iterations in an iterative decoder can be tracked via an EXIT chart.

In many cases, especially when the channel is AWGN, this assumption is fairly accurate[1]. Even when the LLRs are not Gaussian, the transfer functions drawn with the Gaussian assumption are found to be fairly accurate [56, 65]. The real advantage of using this assumption is that a single variable (the mean of the Gaussian) describes the density completely. Instead of computing and tracking the entire densities of

[1]The special relationship between the mean and the variance originated from the symmetry condition in [63]. If the message densities satisfy this condition, then low-density parity check codes can be designed to have good performance on the given channel.

messages, now one needs only to track one variable.

In the ergodic MIMO case, EXIT charts have been used to design good codes [62].

b. Area Property

In [65], the authors prove that if the decoder's apriori information comes from a binary erasure channel (BEC) then the area under the EXIT function is one minus a conditional entropy. The practical implication of this property is that the code design for the BEC reduces to a curve-fitting problem for several classes of codes (e.g. LDPC codes, turbo-codes, RA codes etc).

This area property holds approximately for a large class of channels, though there is no proof in the literature yet. The curve-fitting approach to code-design has been used by many researchers in a variety of problems [62, 66].

B. Iterative Decoding and Demodulation of MIMO Systems in QSFC

When the channel has a quasi-static behavior, as has been assumed throughout this work, each realization of the channel $\mathbf{H}$ results in a different EXIT chart for the space-time demodulator. Analysis and code-design becomes extremely time consuming and tedious if one has to track actual densities of apriori and extrinsic messages for each channel realization. We circumvent this problem by invoking the Gaussian assumption. As we show by examples, the received density is more closely approximated by a mixture of two Gaussian densities. However, for the purposes of this chapter, we still make the Gaussian assumption while computing the EXIT charts. Even though the EXIT charts themselves become inaccurate (in terms of predicting code convergence thresholds etc), the behavior (i.e. the shapes of the EXIT charts) of the demodulator and the decoder is still well characterized.

1. Message Densities at the Output of Space-time Demodulator

Consider a system with $N_t = 2$ transmit antenna that uses BPSK ($M = 2$) modulation. The maximum achievable rate is 2 b/s/Hz with uncoded transmission. We assume that the rate of the outer code is $R_o = 1/2$, so that the overall rate is $R = 1\ b/s/Hz$.

It is easy to see that the receive signal $y = \sqrt{\rho/2}(h_1x_1 + h_2x_2) + n$ yields the log likelihood ratio of the bit b_1 (corresponding to the symbol x_1) as:

$$L(x_1) = \frac{1}{2}L_1(x_1) + \frac{1}{2}L_{-1}(x_1), \tag{4.4}$$

where

$$\begin{aligned} L_1(x_1) &= \log \frac{\Pr(x_1 = 1|x_2 = 1)}{\Pr(x_1 = -1|x_2 = 1)} \\ &= \log \exp\left(-\frac{1}{2}||y - \sqrt{\rho/2}(h_2 + h_1)||^2 + \frac{1}{2}||y - \sqrt{\rho/2}(h_2 - h_1)||^2\right) \\ &= 2\left[\mathcal{R}(h_1)\ \mathcal{R}(y - \sqrt{\rho/2}h_2) + \mathcal{I}(h_1)\ \mathcal{I}(y - \sqrt{\rho/2}h_2)\right]. \end{aligned}$$

Here $\mathcal{R}()$ and $\mathcal{I}()$ represent the real and imaginary parts of the argument, respectively. Similarly,

$$\begin{aligned} L_{-1}(x_1) &= \log \frac{\Pr(x_1 = 1|x_2 = -1)}{\Pr(x_1 = -1|x_2 = -1)} \\ &= \log \exp\left(-\frac{1}{2}||y - \sqrt{\rho/2}(-h_2 + h_1)||^2 + \frac{1}{2}||y - \sqrt{\rho/2}(-h_2 - h_1)||^2\right) \\ &= 2\left[\mathcal{R}(h_1)\ \mathcal{R}(y + \sqrt{\rho/2}h_2) + \mathcal{I}(h_1)\ \mathcal{I}(y + \sqrt{\rho/2}h_2)\right]. \end{aligned}$$

Thus (4.4) becomes:

$$L_{ST}(b_1) = \sqrt{2\rho}\left[\mathcal{R}(h_1)\mathcal{R}(y) + \mathcal{I}(h_1)\mathcal{I}(y)\right]. \tag{4.5}$$

The real and the imaginary parts of y have a Gaussian distribution. Thus, atleast

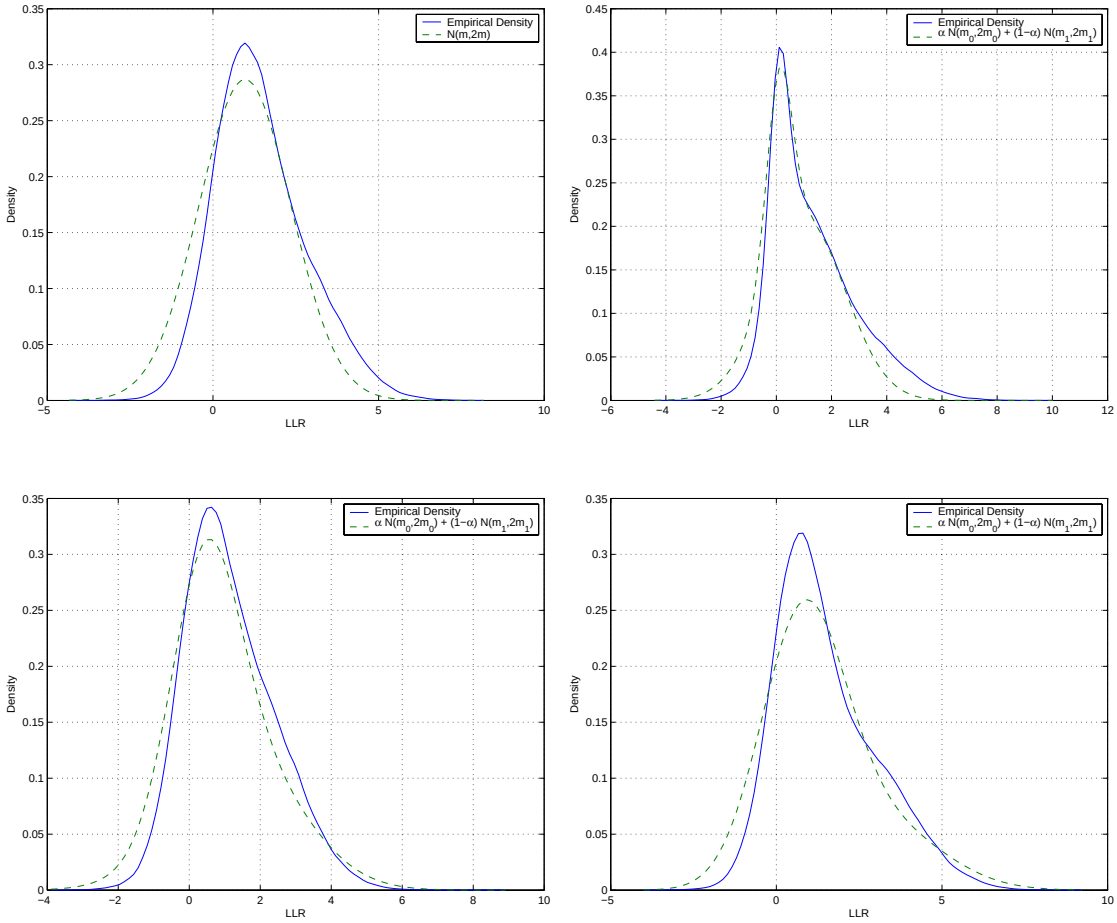

Fig. 23. Probability densities of received LLRs for different channel realizations. BPSK modulation, $N_t = N_r = 2$, $\rho = 6$dB. All these channels have iid BPSK information rate $\approx 1\ b/s/Hz$.

in the $N_r = 1$ case with BPSK we can show that the output of the space-time demodulator is a sum of two Gaussians distributed random variables.

In Figs. 23-24, we show empirically that the density of the received likelihoods at the output of the space-time demodulator can be well approximated as a mixture of two Gaussian distributed random variables, each with a variance that is twice its mean. Further, this property holds even for higher number of receive antennas and for higher modulation order.

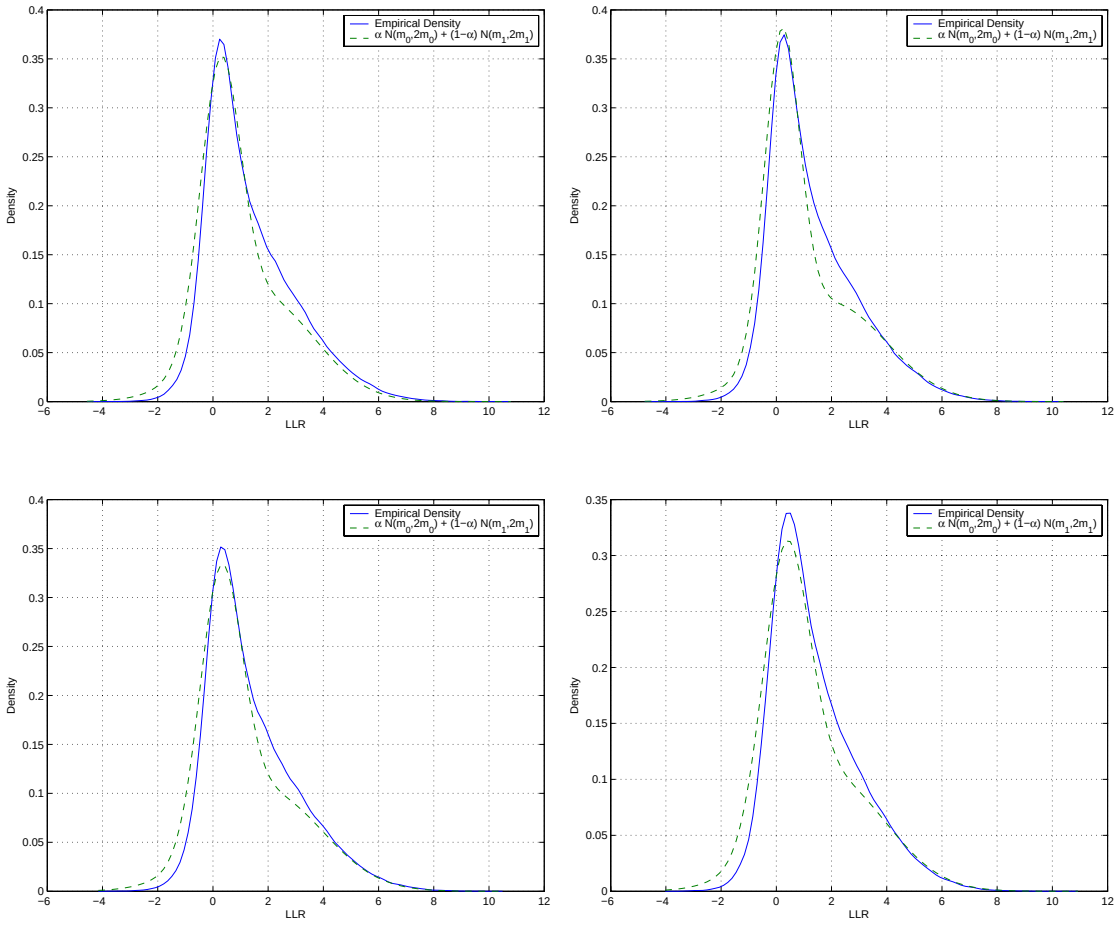

Fig. 24. Probability densities of received LLRs for different channel realizations. QPSK modulation, $N_t = N_r = 2$, $\rho = 9$ dB. All these channels have iid QPSK information rate $\approx 2\ b/s/Hz$.

2. EXIT Charts

In this sub-section, we study the behavior of the space-time demodulator in greater detail. We use the technique of EXIT charts mentioned before. We make the following assumptions:

- The bits $\mathbf{c} = [c_1 \ \ldots \ c_{mN_t}]; c_i \in \{0, 1\}$ that are generate the transmit vector $\mathbf{x} = [x_1 \ x_2 \ \ldots \ x_{N_t}]; x_i \in \mathcal{A}_x$ are independent and equally likely.

- The apriori messages received by the space-time demodulator are Gaussian with a variance that is twice the mean:

$$L_D(c_k) \sim \mathcal{N}(m, 2m). \tag{4.6}$$

- The apriori messages to the outer decoder are Gaussian distributed with a variance that is twice the mean:

$$L_{ST}(c_k) \sim \mathcal{N}(m, 2m). \tag{4.7}$$

The first of these conditions is easily satisfied if the outer code is a random-like code (e.g. LDPC code, turbo code etc) or by assuming that the output of the outer code passes through a long (ideally infinite length) interleaver before the grouping occurs. While (4.7) is different from what we observed in the previous sub-section, it does not change the behavior of the transfer charts. The accuracy of the EXIT charts is sacrificed only slightly. Once (4.7) is assumed, (4.6) holds for most codes [56, 67].

Once again, we assume a $N_t = N_r = 2$ system with BPSK modulation. Recall that the EXIT chart depends on the channel realization $\mathbf{H}$ and the channel SNR ρ. We choose a set of channels wherein the i.i.d. BPSK information rate is ≈ 1 b/s/Hz. We simulate a large number of bits and compute the log-likelihoods $L_{ST}(c_k)$.

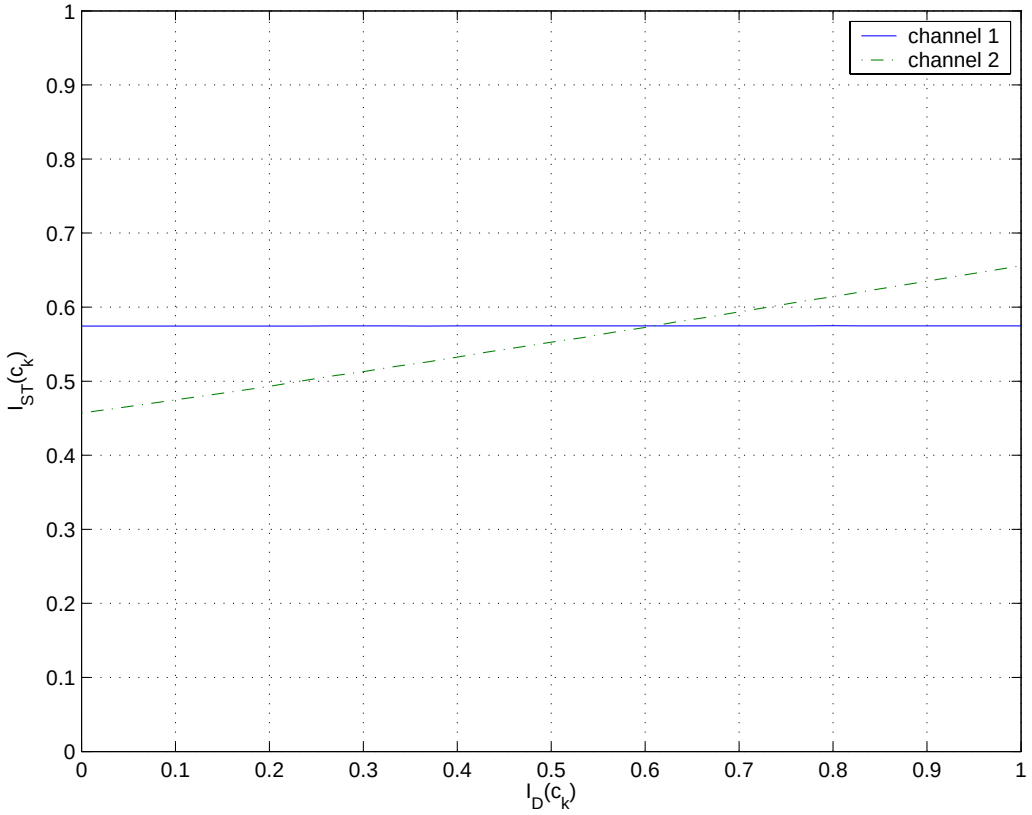

Fig. 25. EXIT charts for $N_t = N_r = 2$ system with BPSK modulation, $\rho = 6$ dB. The two channels H_1 and H_2 have a constrained i.i.d. BPSK mutual information ≈ 1 b/s/Hz.

Finally the extrinsic mutual information $I\left(c_k; L_{ST}(c_k)\right)$ is plotted as a function of $I\left(c_k; L_D(c_k)\right)$.

As an example, we consider different channel realizations with the same constrained modulation information rate. We expect the same area under the curves and we want to see how the shape of the curve changes. In Fig. 25, we show the EXIT charts for two channels

$$H_1 = \begin{bmatrix} 0.3449 + j0.1697 & 0.3987 - j0.4771 \\ -0.1311 + j0.1540 & 0.1008 - j0.3759 \end{bmatrix},$$

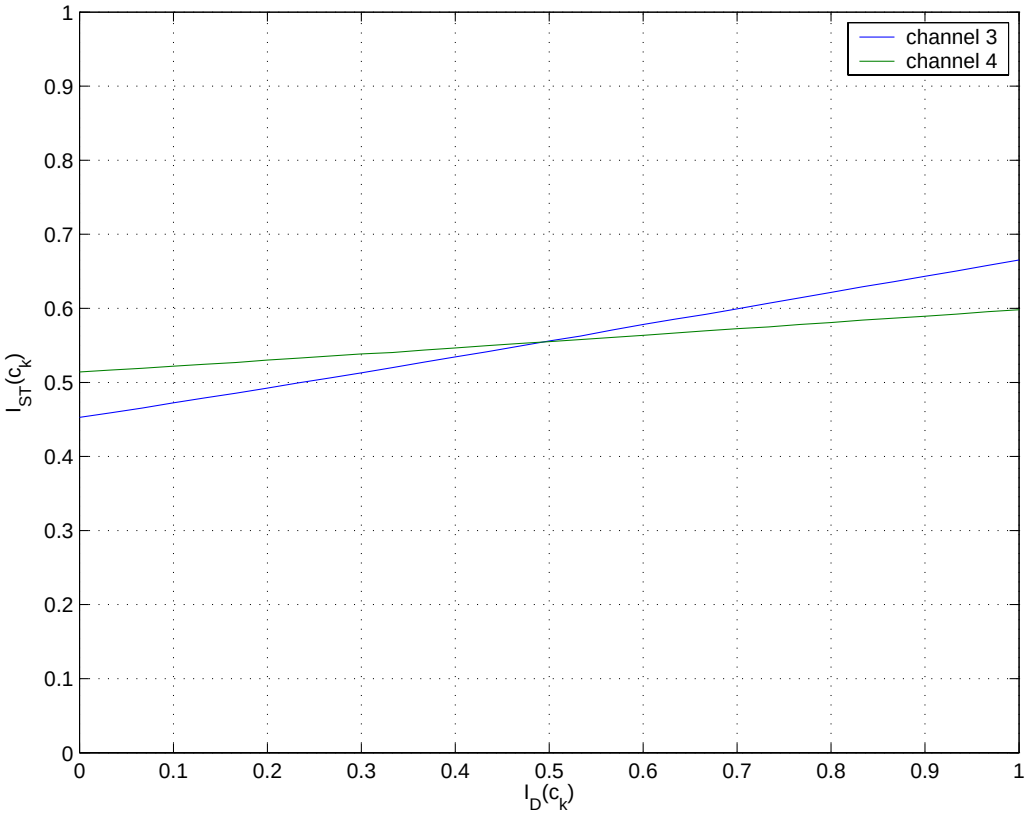

Fig. 26. EXIT charts for $N_t = N_r = 2$ system with QPSK modulation, $\rho = 9$ dB. The two channels H_3 and H_4 have a constrained i.i.d. QPSK mutual information ≈ 2 b/s/Hz.

and

$$H_2 = \begin{bmatrix} -0.3993 - j0.0498 & 0.5103 + j0.1865 \\ 0.2888 - j0.0364 & -0.6031 + j0.2014 \end{bmatrix},$$

which have an iid BPSK mutual information rate of ≈ 1 b/s/Hz at an SNR of $\rho = 6$ dB. In Fig. 26, we show the EXIT charts when the modulation is QPSK. The two channels

$$H_3 = \begin{bmatrix} 0.6529 + j0.2634 & 0.4807 - j0.0366 \\ 0.3382 + j0.1496 & 0.1139 + j0.2011 \end{bmatrix},$$

and

$$H_4 = \begin{bmatrix} -0.1296 - j0.0018 & 0.1205 - j0.7093 \\ 0.3612 - j0.0392 & -0.3088 + j0.4162 \end{bmatrix},$$

have an iid QPSK mutual information rate of ≈ 2 b/s/Hz at an SNR of $\rho = 9$ dB.

We observe that the area under these plots exceeds the corresponding mutual information rate by about 11 to 15 percent. Thus, inspite of the assumptions, these curves are reasonably accurate to draw meaningful conclusions about the systems we are considering.

3. Limits on Iterative Decoding and Demodulation

Recall from Chapter II the way the unconstrained outage capacity is computed. We simulate a large number of channels and compute the information rate for each channel realization at a given the SNR. Given the desired outage probability, the supported rate is found from the quantile plot of the information rates.

This procedure also suggests that the outage-capacity achieving outer code has the following property: This code will be decoded successfully for all channel realizations that have an information rate, $C(\mathbf{H})$, higher than the desired rate, R.

If we assume that the outer code mimics this behavior of the unconstrained case

in the constrained modulation case also, then ideally we need to find an outer code whose EXIT chart (when plotted with the axes interchanged) has the following two properties:

- It lies strictly below the EXIT charts of *all* the channels with the same mutual information rate.
- It has the maximum possible area below it.

This is illustrated in Fig. 27. The boundary of the lower shaded area is how the transfer curve of the "optimal" outer code should look like. This shaded area is equal to the rate of the outer code (to within the accuracy of these EXIT charts).

The upper shaded area represents the loss in rate that is incurred because the receiver has separated demodulation and decoding. This is to be contrasted to a joint maximum-likelihood demodulation and decoding wherein the receiver would be able to recover codewords transmit over both H_1 and H_2, without this rate-loss.

The real importance of recognizing this rate-loss is in the following conclusion. Given any scenario where the (equivalent) channel seen by the iterative decoder is non-ergodic and varies from one codeword to the next, the system employing the iterative decoder cannot be *universal*. All quasi-static channels fall into this category. Also note that this conclusion is not dependent on what the inner and the outer codes are. So, for example, if we were to concatenate an outer LDPC code with an inner space-time trellis code (such as has been done in section III.D), the same conclusion about non-universality will hold.

To corroborate this claim for space-time trellis codes, consider Fig. 28. We plot the EXIT charts for a specific space-time trellis code [1] for two channel realizations which have about the same constrained information rates. Once again, we point out that the overlap in the EXIT charts is the cause of rate-loss and hence iterative

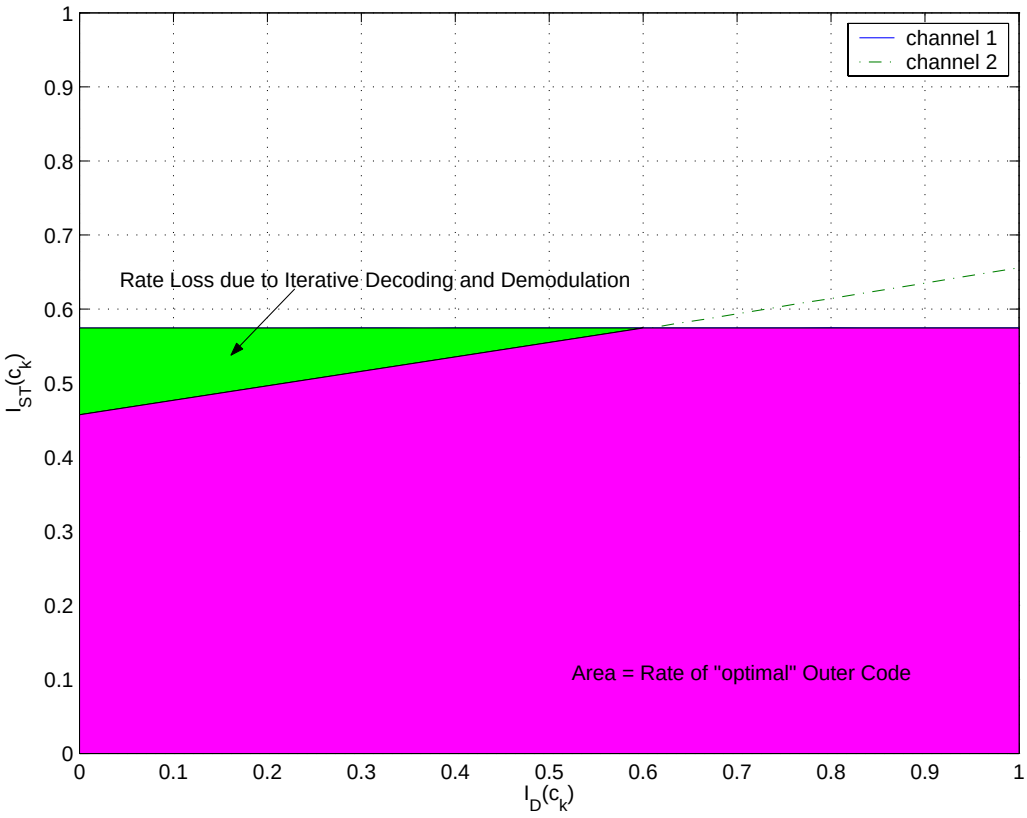

Fig. 27. An example of the sub-optimality of iterative decoding.

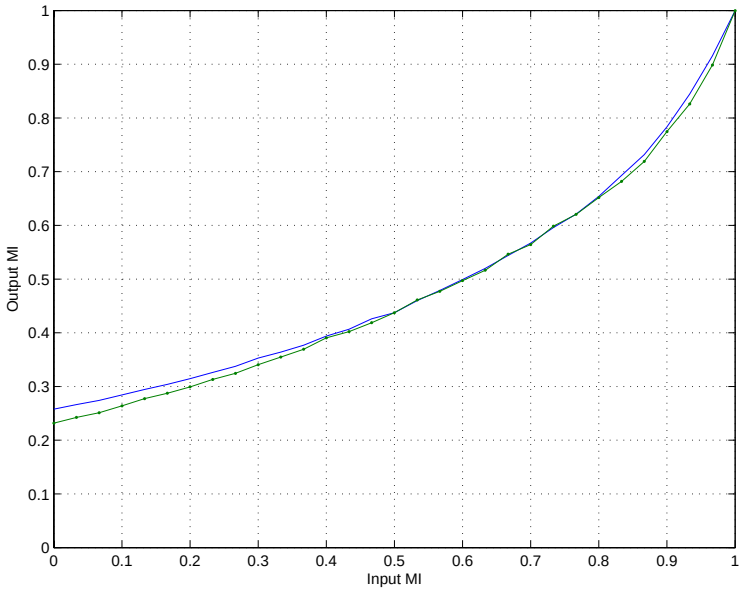

Fig. 28. EXIT charts for $N_t = N_r = 2$, 4-state, 4-PSK space-time trellis code [1] at SNR = 4 dB. Both channel realizations have a constrained information rate of about 1 b/s/Hz.

decoding and demodulation will be sub-optimal. This explains the gap between the performance and the outage limit in Fig. 18-19.

C. Conclusion

In this chapter, we studied the i.i.d. transmission of constrained modulation symbols over the multiple-input multiple-output (MIMO) flat quasi-static fading channel using extrinsic information transfer (EXIT) charts. We showed that the pdf of the received log likelihoods is well represented by a mixture of Gaussian random variables, each with a variance that is twice its mean. We have also shown that iterative decoding and demodulation has an inherent rate-loss. In other words, no matter how good the design of the outer code is, iterative decoding will not be able to achieve the constrained modulation information rate for any non-ergodic channel model.

CHAPTER V

SPACE-TIME TRELLIS CODES: A BCJR-DECISION FEEDBACK DECODING BASED SCHEME

We have seen before that space-time trellis codes (STTCs) offer a higher mutual information rate than, for example, space-time block codes and spatial multiplexing schemes. This higher rate comes at the price of decoding complexity.

The more important issue is that of designing practical schemes which will come close to the outage information rate of the STTC. We have the insight from Chapter IV that iterative decoding is not the best strategy. In this chapter, we consider a system where an outer binary error-correction code is concatenated to an inner STTC. The structure of the transmit data block allows for a step-by-step near-optimal BCJR decision-feedback decoder to decode the STTC. The outer decoder in turn passes near-optimal decisions to the STTC decoder for the next step of equalization. This serial-parallel decoding achieves near outage performance.

In the following sections, we explore the various components of this scheme – the theory behind the scheme, the transmitter structure, the decoder structure and the achieved performance. We also present a variant of the scheme which also achieves close to capacity performance.

A. Background

The structure that we consider here is an extension of what is proposed in [68] for use in ISI channels. There are two main components to this structure. First, a computationally efficient BCJR-decision feedback decoder (BCJR-DFD) for the space-time trellis code which produces optimal soft estimates of the input to the STTC, given the past P symbols exactly (where P is the memory of the STTC) and the observation

from the channel. Second, the encoder and decoder for the outer codes are used in such a manner that allows the past P symbols to be fed to the BCJR-DFD in an error free manner, through the use of a code that achieves capacity on the memoryless channel.

A system with multiple-transmit antennas presents a channel which looks very similar to the inter-symbol interference (ISI) channel and the multi-user channel . There has been a lot of work in these areas that can be applied in some manner to the MIMO case. It is well known that MMSE feedback equalization with error free feedback is a canonical structure and can be used to predict the performance for any ISI channel accurately [69, 70]. In the multi-user scenario, the optimality of ideal decision feedback is also well known [71]. This has been used to prove the optimality of D-BLAST in the diversity-multiplexing trade-off sense [12].

Recently it has been shown that if the MMSE is replaced by a BCJR algorithm, ideal feedback can lead to achieving capacity for any SNR and any input constellation [68]. Further, this scheme does not require the knowledge of the ISI channel at the transmitter and performs close to capacity for several ISI channel realizations. This scheme, therefore, is a good candidate for use in the MIMO quasi-static flat fading case where the channel realization is not known apriori and we want the scheme to perform well on a number of channel realizations.

In order to adapt the scheme to the case we are considering, we replace the ISI channel with an STTC. The binary decision feedback equalizer (DFE) is replaced by a symbol-based decision feedback decoder (DFD). Since the STTC operates in a non-binary domain, we use a BICM-like approach for the outer code. In other words, the output of a binary code is interleaved and mapped on to symbols, which are then fed to the STTC. This is explained in more details in the following sections.

B. Transmitter Structure

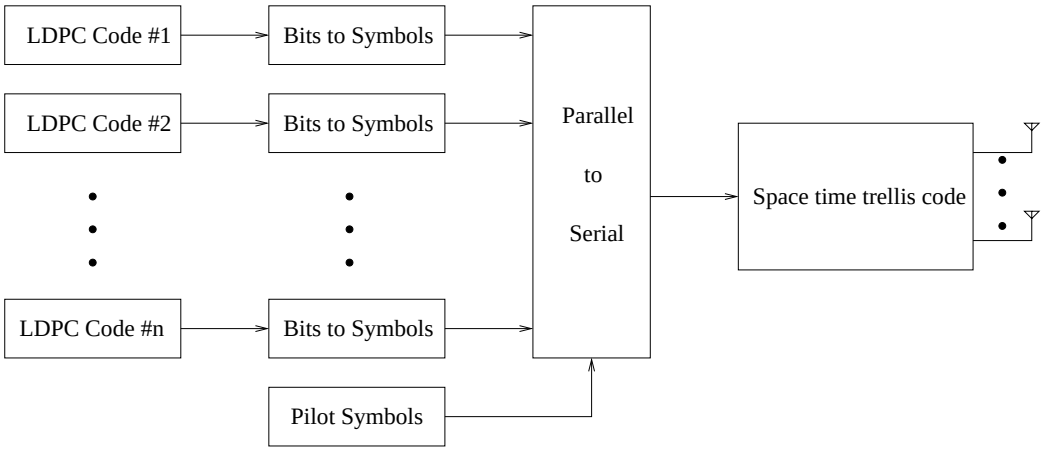

Fig. 29. Transmitter structure for the proposed BCJR-DFD based scheme.

The transmitter structure is shown in Fig. 29. It consists of n LDPC encoders arranged in parallel followed by a symbol-mapper for each of the LDPC codes. Let c_{ij} denote the j-th coded bit of the $i-$th LDPC encoder. The mapper combines m consecutive coded bits to form the k-th symbol $x_{ik} \in \mathcal{A}_x$, where $|\mathcal{A}_x| = M$, for example, $\mathcal{A}_x$ is the M-PSK alphabet.

Next, the output of these mappers is converted to a serial stream. So, the $k-$th block of data input to the space-time trellis code, $\mathbf{U}_k$, consists of symbols $x_{ik}, i = 1, 2, \ldots, n$ and p pilot symbols $x^p_{ik}, i = 1, 2, \ldots, p$. The purpose of the pilots is to enforce that the space-time trellis code is terminated in a known state (say, state S_*). Therefore, p equals P, the memory of the space-time trellis code. The space time trellis code encodes this data $\mathbf{U}_k$ to form the $k-$th transmission block $\mathbf{T}_k$. If N be the length of each LDPC code, the number of such transmission blocks is $K = (N/m)$. The codeword, $\mathbf{C}$ for this overall scheme is the concatenation of the K transmission blocks $\mathbf{T}_k, k = 1, 2, \ldots, K$.

We can graphically represent the data blocks as shown in Fig. 30. Since, for a system with N_t transmit antennas and N_r receive antennas, the STTC outputs a vector of N_t symbols for every x_{ik} input to it, we can write the transmission blocks $\mathbf{T}_k$ if we replace each x_{ik} by $\mathbf{x}_{ik}$ in Fig. 30. For each $\mathbf{x}_{ik}$, the transmission through the channel may be represented as:

$$\mathbf{y}_{ik} = \sqrt{\frac{\rho}{N_t}}\mathbf{H}\mathbf{x}_{ik} + \mathbf{n}_{ik}, \tag{5.1}$$

where $\mathbf{y}_{ik}$ is a vector of N_r received symbols, rho is the SNR, $\mathbf{H}$ is the channel realization and $\mathbf{n}_{ik}$ is the noise vector with zero mean complex Normal entries. For ease of notation, we can collect the samples corresponding to a row in Fig. 30 into a vector/matrix and write:

$$\begin{aligned} X_k &= \{x_{ik}\};\ i = 1, \ldots, n \\ \mathbf{X}_k &= \{\mathbf{x}_{ik}\};\ i = 1, \ldots, n \\ \mathbf{Y}_k &= \{\mathbf{y}_{ik}\};\ i = 1, \ldots, n \end{aligned}$$

etc.

C. Receiver Structure

The receiver structure is shown in Fig. 31. The underlying algorithm used in this block is the maximum *a posteriori* probability algorithm of Bahl, Cocke, Jelinek and Raviv [72]. Since the STTC is a Markov chain, if the previous state is known, the a posteriori probability of a symbol depends only on the signal received in that time instant:

$$\Pr(x_{ik}|\mathbf{Y_k}, S_{i,k}) = \Pr(x_{ik}|\mathbf{y}_{ik}, S_{i,k}). \tag{5.2}$$

Note that due to the presence of pilots, the states $S_{0,k}$ and $S_{n+1,k}$ are known and fixed (equal to S_*). Consequently, the received values of the symbols for row $k+1$ do not

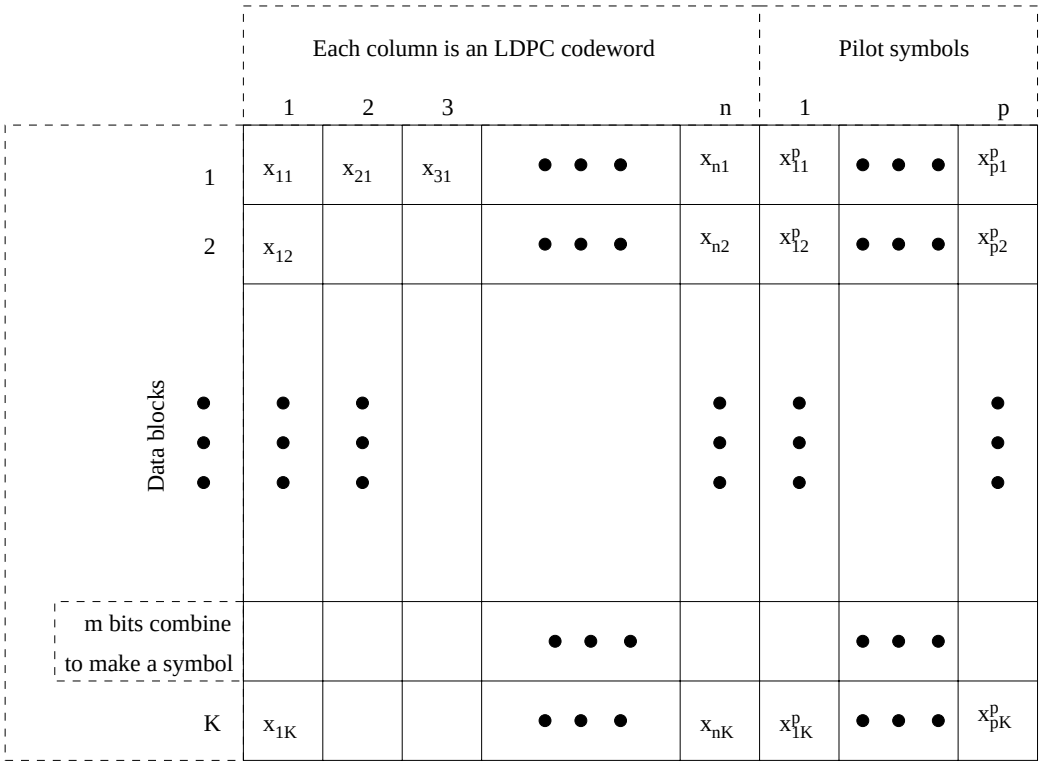

Fig. 30. Data matrix representing the operations at the transmitter of the proposed BCJR-DFD based scheme.

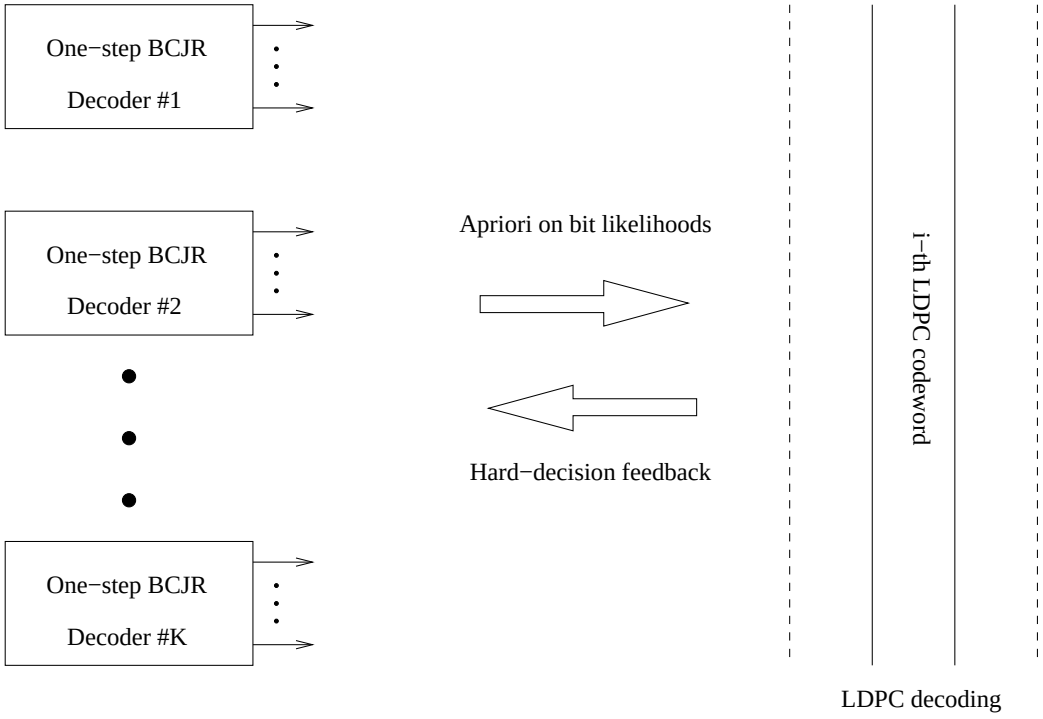

Fig. 31. The BCJR-DFD receiver structure for the proposed scheme has a serial-parallel decoding rule.

affect the soft outputs of symbols in row k.

As a result of (5.2), we can run K BCJR decoders in parallel, one for each row in Fig. 30. Without loss of generality, consider the k-th BCJR module. In order to describe the working of this modified decoder, we define the following quantities [73]:

1. The a priori probability on the input to the decoder:

$$P(X_k; I) = (P_i(x_{ik}; I))_{1 \le i \le n},$$

where $P_i(x_{ik}; I)$ is the probability that the i-th symbol in the k-th data block is x_{ik}.

2. The a priori probability of the coded symbols:

$$P(\mathbf{X}_k; I) = (P_i(\mathbf{x}_{ik}; I))_{1 \le i \le n},$$

where $P_i(\mathbf{x}_{ik}; I)$ is the probability that the i-th symbol in the k-th transmission block is $\mathbf{x}_{ik}$. Since the channel is additive white Gaussian (5.1), we have:

$$\begin{aligned} P_i(\mathbf{x}_{ik}; I) &= \Pr(\mathbf{x}_{ik} | \mathbf{y}_{ik}, \mathbf{H}) \\ &= \frac{1}{\sqrt{\pi}} \exp\left(-\frac{1}{2} \left| \mathbf{y}_{ik} - \sqrt{\frac{\rho}{N_t}} \mathbf{H} \mathbf{x}_{ik} \right|^2 \right). \end{aligned} \tag{5.3}$$

3. Corresponding to the a priori probabilities (input to the decoder), we define the a posteriori probabilities (output of the decoder) by replacing I with O as below:

$$P(X_k; O) \quad P(\mathbf{x}_k; O),$$

which are computed by the decoder.

4. For a given edge e on the trellis, define its starting state $s^S(e)$, the ending state $s^E(e)$, the corresponding input $x(e) \in \mathcal{A}_x$ and the corresponding output

$\mathbf{x}(e) \in \mathcal{A}_x^{N_t}$. Let the set of states be denoted by S so that $s^S(e), s^E(e) \in S$.

The usual BCJR algorithm is then described by these equations [72, 73]:

$$P_i(x_{ik}; O) = N_u \sum_{e:x(e)=x_{ik}} \alpha_{i-1}[s^S(e)] P_i(\mathbf{x}(e); I) \beta_i[s^E(e)]; \tag{5.4}$$

$$\alpha_i[s] = \sum_{e:s^E(e)=s} \alpha_{i-1}[s^S(e)] P_i(x(e); I) P_i(\mathbf{x}(e); I); \quad i = 1, 2, \ldots, n; \tag{5.5}$$

$$\beta_i[s] = \sum_{e:s^S(e)=s} \beta_{i+1}[s^E(e)] P_{i+1}(x(e); I) P_{i+1}(\mathbf{x}(e); I); \quad i = 1, 2, \ldots, n. \tag{5.6}$$

The normalization factor N_u ensures that $\sum_{x_{ik} \in \mathcal{A}_x} P_i(x_{ik}; O) = 1$. At any time i, if the state is known (say $s = \hat{S}$), we can write:

$$\alpha_i[s] = \begin{cases} 1 & s = \hat{S} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

$$\beta_i[s] = \begin{cases} 1 & s = \hat{S} \\ 0 & \text{otherwise} \end{cases} \tag{5.8}$$

So, the presence of pilots allows us to initialize both α_0 and β_n since the initial and final states of the trellis are S_*. Since the output of the LDPC code is approximately iid, the input to the STTC encoder is also iid. Thus, $P_i(x_{ik}; I) = 1/M$.

Assuming that we have the symbol extrinsic $P_i(x_{ik}; O)$ for all $x_{ik} \in \mathcal{A}_x$, we can compute the bit extrinsic probabilities easily. Assume that the symbol $x \in \mathcal{A}_x$ is associated with bits $u_j, 1 \leq j \leq m$. We have [73]:

$$P_{ij}(u_{ikj}; O) = N_{ub} \sum_{x:u_k=u_{ikj}} P_i(x_{ik}; O) \prod_{l \neq j} P_{ij}(u_{ikl}; I) \tag{5.9}$$

where the a priori bit probabilities $P_{ij}(u_{ikl}; I)$ are all equal to $1/2$ (since the bits are iid) and u_{ikj} represents the j-th bit of the symbol x_{ik}. The normalization factor N_{ub} ensures that the resulting bit probabilities sum up to one as they should. The log

likelihood ratio is then computed as:

$$\lambda_{ikj} = \log \frac{P_{ij}(u_{ikj} = 1; O)}{P_{ij}(u_{ikj} = 0; O)}, \tag{5.10}$$

and for ease of notation we re-index the λs to λ_{ik} such that $1 \leq i \leq N$ and $1 \leq k \leq K$.

The computations in (5.3)-(5.9) can also be all done in the log domain by working with logarithm of probabilities. This makes the computations numerically less intensive and this is what we implement.

The decoding algorithm can now be summarized as below:

1. *Initialization:* From the received signal $\mathbf{Y}_k; 1 \leq k \leq K$, compute $P_i(\mathbf{x}_{ik}; I)$ using (5.3). Initialize the variable $S_p(k) = S_*$ for $1 \leq k \leq K$.

2. *Backward recursion:* For each of the K equalizers, compute the backward recursion via (5.6) for all time instants $1 \leq i \leq n$.

3. *Serial-parallel decoding:* For each time instant $1 \leq i \leq n$, do the following:

   (a) For each of the K decoders $1 \leq k \leq K$:

   - *One-step BCJR DFD:* Compute one step of the forward recursion via (5.5) and (5.7) with $\hat{S} = S_p(k)$. Compute the extrinsic information on the symbols via (5.4).
   - *Bit estimates:* From the symbol extrinsic information $P_i(x_{ik}; O)$, compute the bitwise extrinsic LLRs λ_{ik} using (5.9) and (5.10).

   (b) *LDPC decoding:* Using the λ_{ik}, run the LDPC decoding for the i-th LDPC code. Make hard decisions on the output.

   (c) *Symbol decisions:* From the LDPC decoder output bits, compute hard estimates of the symbols $\hat{x}_{ik}; 1 \leq k \leq K$.

   (d) *Decision feedback:* For each equalizer k:

- Compute the edge e on the trellis that corresponds to the (known) previous state $S_p(k)$ and the input $\hat{x}_{ik}$.
- Update $S_p(k) = s^E(e)$.

An important implication of this encoding and decoding strategy is that at each time instant i, the trellis section is reduced to only one state $S_p(k)$ and so there are only $|\mathcal{A}_x|$ transitions to consider. This set of transitions or trellis edges is given by $E = \{e : s^S(e) = S_p(k)\}$. The output of the BCJR DFD depends on the set of distances $D[S_p(k)] = \{|\mathbf{x}(e1) - \mathbf{x}(e2)| : e1, e2 \in E\}$. If this set is identical for all $S_p(k) \in S$, we expect a uniform error protection and therefore good frame error rate performance. Otherwise, the performance of the scheme will depend on the input sequence and since we are interested in frame-error rates, the performance may suffer. This is verified by simulation results presented in the next section.

D. Simulation Results

We consider a system with $N_t = N_r = 2$. The modulation used is 4-PSK. The outer binary LDPC code is chosen to be a rate-1/2 code, whose profile is optimized for the AWGN channel (Table II, code with $d_v = 20$ in [63]). The threshold for this profile is 0.11 dB away from AWGN capacity. A length $N = 8192$ code constructed using the bit-filling method [74] is used. The LDPC decoder performs 100 iterations.

We use two different 4-state, 4-PSK space-time trellis codes – the delay diversity (DD) code of [2] and the code (CYV) from [1]. The latter code is the best known 4-state 4-PSK code space-time trellis code. The set of distances $D[s]$ is identical for all the states for DD code but not for the CYV code. Only one pilot column $P = 1$ is needed for each of these codes. We use a block length of $n = 10$ for simulations. The performance does not change if the value of n is made larger and therefore the

bandwidth loss due to the presence of pilots is neglected.

The scheme is simulated for an ensemble of 10000 channels. In order to find the threshold performance of this scheme (i.e. performance in the limiting case of $N \to \infty$ and infinite number of iterations), we look at the subset of channels where the system with the finite length code results in an error. For each channel in this subset, we first compute the probability density function (pdf) of the received bit likelihood ratios (i.e. pdf of λ_{ik}) assuming perfect decision feedback. Density evolution [63] is then used to determine whether the chosen LDPC profile will be able to correctly decode or not, given the channel realization. We assume that during the actual construction of the code, the degree 2 variable nodes can all be made parity bits and so the density evolution looks only at the bit error rate in non-degree-2 nodes. If this error rate is low (less than 1e-6), we consider the frame error free. The number of channel realizations still in error determine the threshold performance of the system.

The results are presented in Figs. 32-33. The code performance is within a dB of the constrained outage for both the DD and the CYV codes. The thresholds is 0.25 dB for the DD code and 0.4 dB for the CYV code. This difference arises since the set of distances $D[s]$ for the CYV code depends on the state s.

E. A Variation Based on Multilevel Encoding

We can also look at a variant of the transmitter where the bits that are input to the symbol mapper come from different LDPC codes, in a manner similar to traditional multi-level coding [75]. The key difference here is that we do not use different codes for the various bits that enter the symbol mapper. In comparison to the previous scheme, now we have $m \times n$ LDPC encoders and we index them as $\{(1,1),\ldots,(1,m),\ldots,(n,1),\ldots,(n,m)\}$. This scheme is shown in Fig. 34 and the

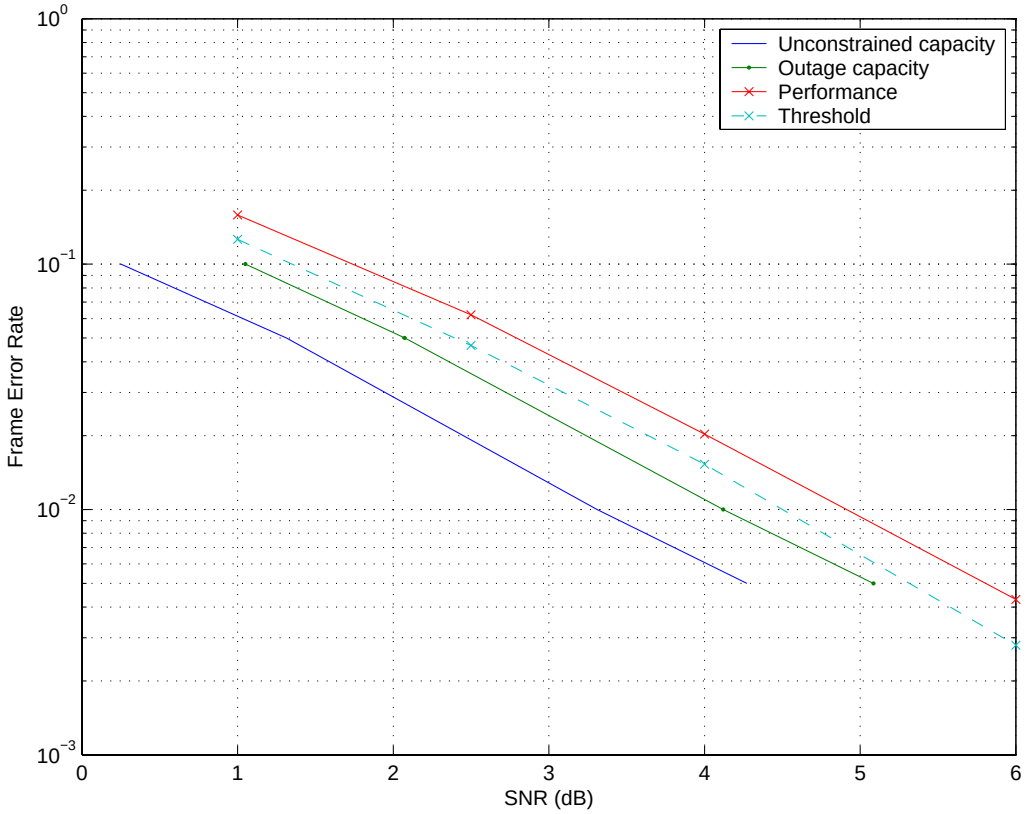

Fig. 32. Performance of the BCJR-DFD scheme with $N_t = N_r = 2$, 4-state delay diversity, QPSK with a rate 1/2 outer LDPC code. The overall rate is 1 b/s/Hz.

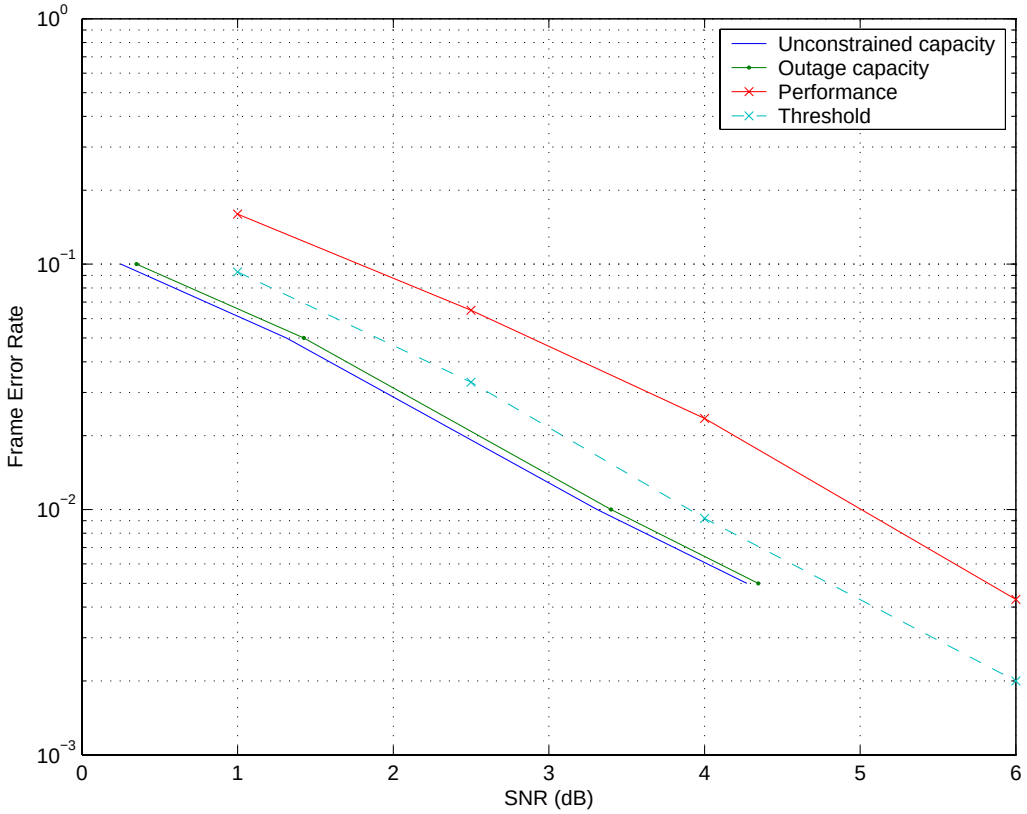

Fig. 33. Performance of the BCJR-DFD scheme with $N_t = N_r = 2$, 4-state code from [1], QPSK with a rate 1/2 outer LDPC code. The overall rate is 1 b/s/Hz.

corresponding data matrix is shown in Fig. 35.

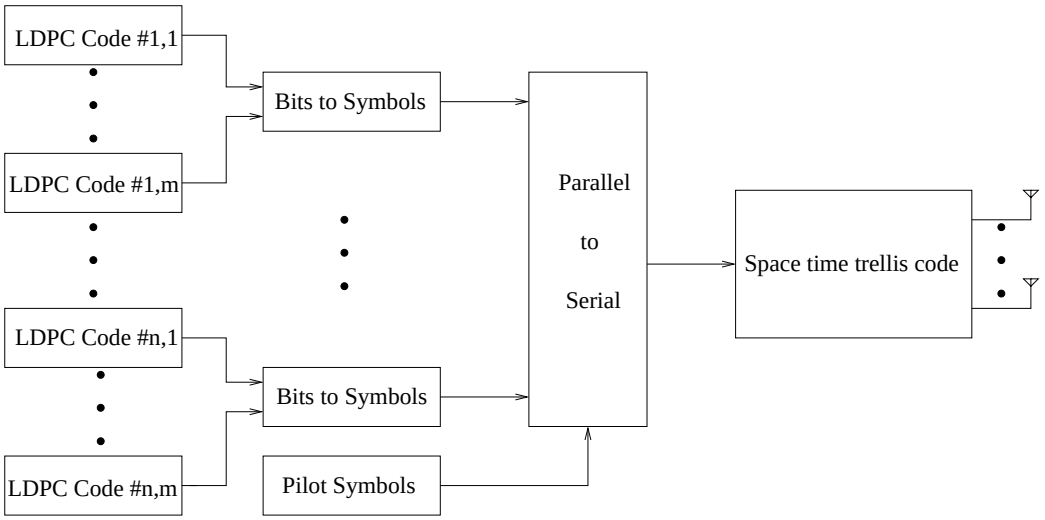

Fig. 34. A variation in the transmitter structure for the BCJR-DFD based scheme.

The decoding algorithm can be easily modified and is summarized below:

1. *Initialization:* From the received signal $\mathbf{Y}_k; 1 \leq k \leq K$, compute $P_i(\mathbf{x}_{ik}; I)$ using (5.3). Initialize the variable $S_p(k) = S_*$ for $1 \leq k \leq K$.

2. *Backward recursion:* For each of the K equalizers, compute the backward recursion via (5.6) for all time instants $1 \leq i \leq n$.

3. *Serial-parallel decoding:* For each time instant $1 \leq i \leq n$, do the following:

   (a) For each of the K decoders $1 \leq k \leq K$:

   - *One-step BCJR DFD:* Compute one step of the forward recursion via (5.5) and (5.7) with $\hat{S} = S_p(k)$. Compute the extrinsic information on the symbols via (5.4).
   - *Bit estimates:* From the symbol extrinsic information $P_i(x_{ik}; O)$, compute the bitwise extrinsic LLRs λ_{ikj} using (5.9) and (5.10).

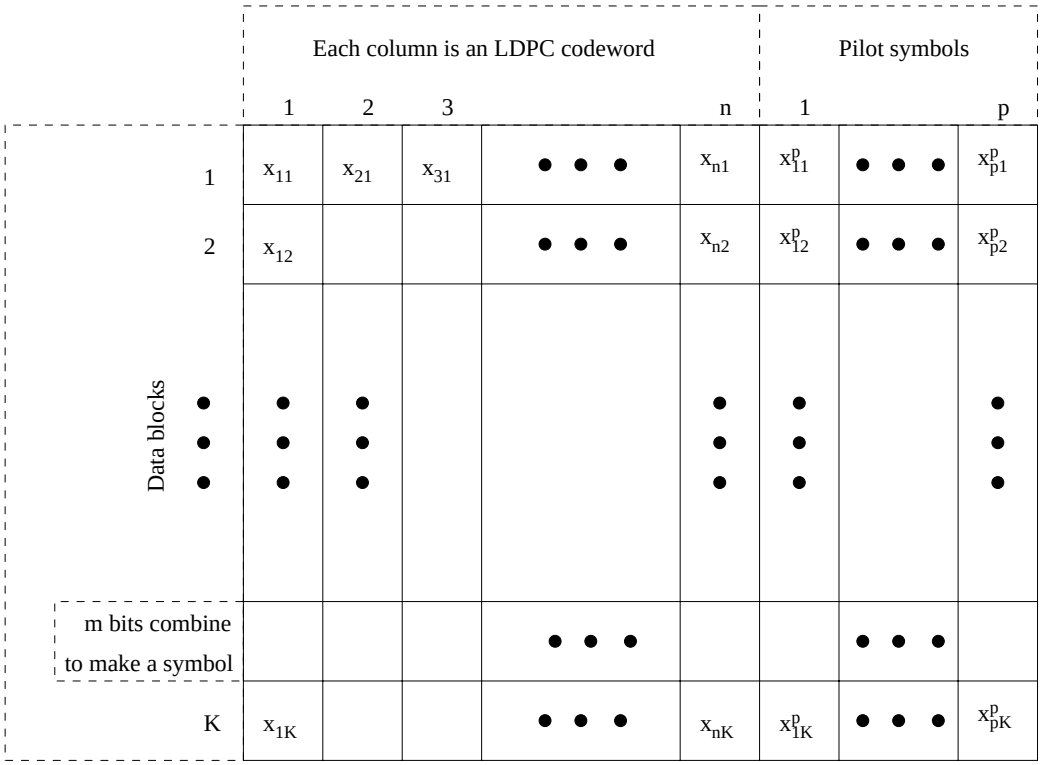

Fig. 35. Data matrix representing the operations at the transmitter of the proposed variation to the BCJR-DFD based scheme.

(b) *LDPC decoding:* For each $1 \leq j \leq m$:

- Using the λ_{ikj}, run the LDPC decoding for the (i, j)-th LDPC code. Make hard decisions on the output.

(c) *Symbol decisions:* From the outputs of the m LDPC decoders $\{(i, 1), \ldots, (i, m)\}$, compute hard estimates of the symbols $\hat{x}_{ik}; 1 \leq k \leq K$.

(d) *Decision feedback:* For each equalizer k:

- Compute the edge e on the trellis that corresponds to the (known) previous state $S_p(k)$ and the input $\hat{x}_{ik}$.
- Update $S_p(k) = s^E(e)$.

The performance of this scheme for the delay diversity code is shown in Fig. 36. In order to keep the simulation complexity low, we reduce the length of the frame n to 5 (instead of 10). It should be noted that even though the two schemes have similar performance, for fixed n, the original BICM based scheme has an advantage over the MLC based variant. In the former, we can afford to increase the length of the LDPC code to m times that of the latter scheme and still keep the same complexity. In general, the performance of an LDPC code improves with length. Further, the hardware for the MLC based scheme requires m LDPC decoders to work in parallel whereas the BICM-based scheme requires only one LDPC decoder.

F. Conclusions

In this chapter we looked at a novel encoding and decoding structure for use with space-time trellis codes. The structure is inspired by the fact that error-free decision feedback achieves capacity in the ISI channel case. The decoder for the space-time trellis code is implemented as a decision-feedback BCJR algorithm (BCJR-DFD).

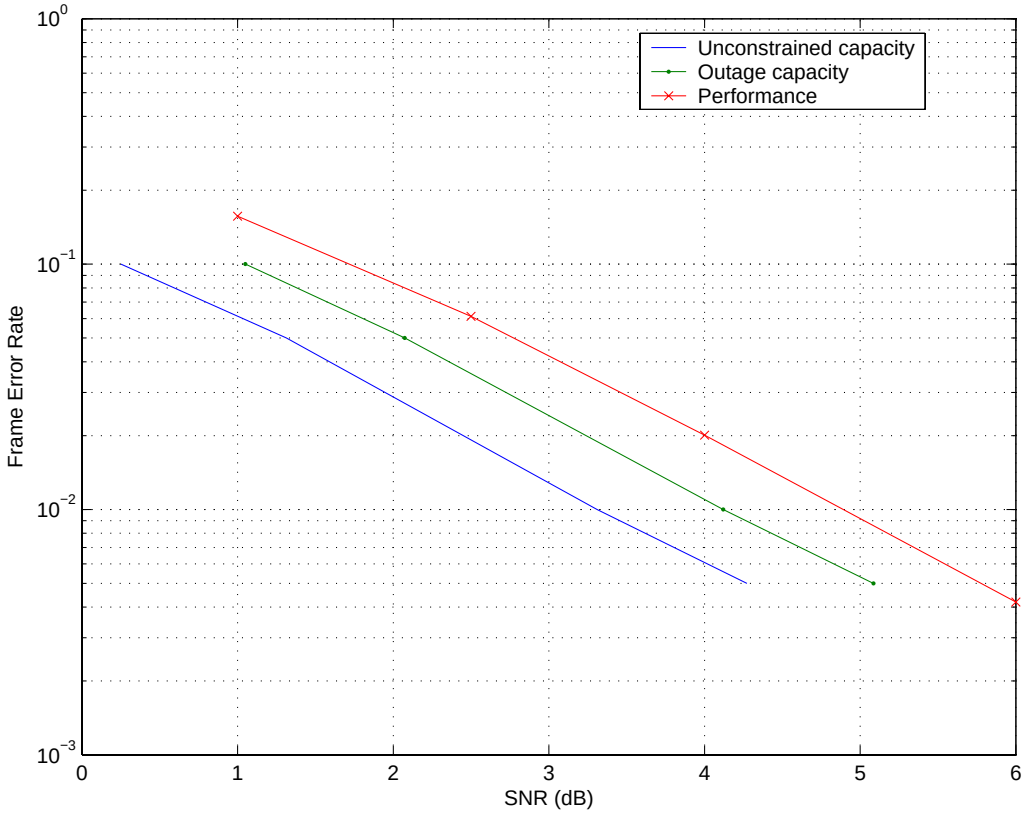

Fig. 36. Performance of the variation of the BCJR-DFD scheme with $N_t = N_r = 2$, 4-state delay diversity, QPSK with a rate 1/2 outer LDPC code. The overall rate is 1 b/s/Hz.

The receiver structure ensures that the decisions being fed back are error-free (due to the use of a capacity achieving LDPC code). The performance is within a dB of the constrained outage of the given space-time trellis code. The thresholds are within 0.25 dB of the constrained outage for the delay diversity scheme and within 0.4 dB of the CYV code.

We also propose a multi-level variant of this scheme. The performance is very similar to the original proposal, although this variant has a higher receiver complexity.

CHAPTER VI

SPACE-TIME TRELLIS CODES: A SERIAL CONCATENATION SCHEME BASED ON RECURSIVE REALIZATIONS[†]

In this chapter, we consider a serial-concatenation/iterative-decoding approach to achieving constrained outage information rates of space-time trellis codes. We propose and analyze a class of codes which are a serial concatenation of an outer binary code with an inner *recursive* space-time code such as considered in [76]. These codes can be decoded using an iterative (turbo) decoding procedure and, hence, we call this family of concatenated codes as *Turbo Space-time Codes (TSTC)*. Two solutions are proposed - codes to primarily improve power efficiency where the outer code is of fairly low rate, and codes to improve power efficiency with minimal loss in spectral efficiency (when the outer code is of very high rate). We show that several classes of space-time codes have equivalent recursive realizations. Since the inner code is still essentially a space-time code, the overall code delivers the promised spatial diversity. Analysis of serial concatenated space-time codes such as what are considered here has recently appeared in [43]. The analysis here and in [43] are similar since they are both based on deriving the union bound for SCCC over the ensemble of all interleavers. We also provides more specific results when the inner code is a differential encoder based space-time code. In addition to the above results, we show that single parity check based turbo codes are a good candidate for high rate outer codes.

A. System Model

We consider the transmission of data in frames of length N. As shown in Fig. 37, a block of data is encoded using a rate k/n outer code, interleaved, and then encoded using a space-time inner code. In the next section, we will show that many space-time trellis encoders can be realized using recursive encoders. The output of the space-time encoder at every epoch is a group of n 2^p-ary symbols which are transmitted using $N_t = n$ transmit antennas thereby making the overall spectral efficiency (pk/n) bits/sec/Hz. At the receiver, each of the m antennas receive a noisy super-position of the transmitted signals. Specifically, the signal received by the j-th antenna is given by:

$$r_j(t) = \sqrt{\rho_t} \sum_{i=1}^{n} \alpha_{i,j}(t) s_i(t) + n(t) \tag{6.1}$$

where $s_i(t)$ is the signal transmitted from the i-th antenna, $\alpha_{i,j}(t)$ is the channel gain from the i-th transmit antenna to the j-th receive antenna and $n(t)$ is an additive white Gaussian noise process. The channel gains and the noise process are normalized to unit power so that $\rho_t = \rho/n$ is the average SNR per receive antenna, independent of the number of transmit antennas. All the results in this chapter, with the exception of the figure on page 113, are for two transmit antennas and one receive antenna ($n = 2$ and $m = 1$).

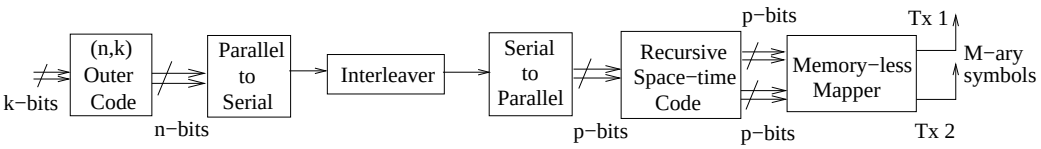

Fig. 37. Encoder structure for the serial concatenation scheme.

We use an iterative procedure [77,78] for decoding. First, the log-likelihood ratios

(LLRs) of the $2p$ bits generated by the recursive space-time encoder are computed. A general procedure for this computation from complex received signals is described in [38]. These LLRs are then fed to the inner decoder. The space-time code, being a trellis based code, can be decoded using the MAP decoding rule [72,78,79]. Extrinsic information is then passed to the outer decoder. For the outer codes considered in this chapter, computationally efficient soft output decoding algorithms exist. The outer decoder then processes the extrinsic information of the inner decoder (which is the *apriori* information for this outer decoder) to generate extrinsic information about both the information bits and the coded bits. This serves as the *apriori* information for the inner decoder for the next iteration.

In the case of parallel concatenation, a block of N symbols and its (symbol) interleaved version are encoded using identical recursive space-time codes. The outputs of these codes are transmitted at two successive time instants. The overall rate of this code is 1/2. The decoding procedure is similar to that described above. In this case, since we use a symbol interleaver, all operations are on symbol likelihoods.

For any particular channel realization, the samples of $\alpha_{i,j}(t)$ are complex, zero-mean, Gaussian distributed random variables with a variance of 1/2 in each dimension. Further, throughout this chapter, we assume knowledge of perfect channel state information at the receiver (perfect CSI-R).

B. Recursive Realization of Space-time Trellis Codes

Space-time trellis codes can be represented using a linear shift register (LSR) followed by a signal mapper or directly in terms of a trellis diagram which shows the output signals along each branch in the trellis. We refer to the former case as LSR based space-time trellis codes. For LSR based space-time trellis codes, the concept of a

recursive realization is well understood. In order to include the Non-LSR based space-time trellis codes, we define recursive realization as follows:

Definition 1 (Recursive Space-Time Trellis Codes) *A space-time trellis code is said to be recursive if the minimum Hamming distance between two input words (sequences) that produces a finite length error event is 2.*

For the case of LSR based codes, this means that feedback shift registers are used in the encoder. For any space-time trellis code (including non-LSR space-time codes), following two conditions on the structure of the trellis are sufficient for the code to be recursive:

1. All transitions originating from a particular state are caused by different inputs.

2. All transitions terminating at a particular state are caused by different inputs.

These conditions ensure that any finite length error event includes atleast two transitions where the input symbols are different. The first condition is always true; otherwise the code is not uniquely decodable. In most cases, the second condition is easy to satisfy by a suitable re-labeling of the mapping from inputs to outputs.

In the following discussion we first show how a LSR based space-time trellis code may be realized in a recursive fashion. Then, we show an example of how re-labeling the trellis can make it recursive.

1. LSR (Linear Shift Register) Based Space-time Trellis Codes

A space-time trellis encoder takes in an M-ary ($M = 2^p$) symbol and outputs a set of n M-ary symbols. For the class of LSR based codes, this can also be viewed as an (np, p) *binary* convolutional encoder followed by a memory-less mapper that maps the np bits to n M-ary symbols. Given this perspective, we can show that *all* space-time

trellis codes based on LSR's can be realized using recursive encoders. This includes the space-time trellis codes proposed by Baro [80], El Gamal [24] etc.

Let x_k be the symbol input to the space-time trellis encoder at time k which can be de-multiplexed into a set of p bits: $x_k = [x_{1k}\ x_{2k}\ \ldots\ x_{pk}]$. At each time instant this group of p bits selects one of the 2^p possible branch transitions. The outputs along the selected transition are transmitted using n antennas. The encoder is equivalent to n separate binary rate-1 encoders followed by a memory-less mapper. The output of the i-th encoder at time k is a p-tuple denoted by $y_k = [y_{1,i,k}\ y_{2,i,k}\ \ldots\ y_{p,i,k}]$. The D transform of y_i for $1 \leq i \leq n$ can be expressed in matrix form as

$$Y_i(D) = X(D)\ G_i(D) \tag{6.2}$$

where

$$\begin{aligned} Y_i(D) &= [y_{1,i}(D)\ y_{2,i}(D)\ \ldots\ y_{p,i}(D)] \\ X(D) &= [x_1(D)\ x_2(D)\ \ldots\ x_p(D)] \\ G_i(D) &= \begin{bmatrix} g_{1,1,i}(D) & g_{1,2,i}(D) & \cdots & g_{1,n,i}(D) \\ g_{2,1,i}(D) & g_{2,2,i}(D) & \cdots & g_{2,n,i}(D) \\ \vdots & \vdots & & \vdots \\ g_{p,1,i}(D) & g_{p,2,i}(D) & \cdots & g_{p,n,i}(D) \end{bmatrix} \end{aligned} \tag{6.3}$$

where $g_{r,s,t}(D)$ is the generator polynomial linking the r-th input stream to the s-th output stream for the t-th transmit antenna. Let the degree of $g_{r,s,t}(D)$ be denoted by $\deg(g_{r,s,t}(D))$ and let $\nu_r = \max_{s,t} \deg(g_{r,s,t}(D))$. Clearly, $\sum_{r=1}^{p} \nu_r = J$, where the states in the trellis diagram of the space-time code denote the content of the J shift registers. Let $G(D) = [G_1(D)\ G_2(D)\ \ldots\ G_i(D)\ \ldots\ G_n(D)]$ denote the overall generator matrix which is the concatenation of the individual $G_i(D)$. Now, consider a generator matrix $G_R(D)$ obtained by dividing the j-th row of $G(D)$ by any

polynomial $T_j(D)$. Then, the new generator matrix represents a recursive encoder for generating the *same* set of codewords as that of the original space-time code. It is easy to see that if the input $X(D) = [x_1(D)\ x_2(D)\ \ldots\ x_p(D)]$ generates a codeword $y(D)$ with the generator matrix $G(D)$, then the input sequence $X_R(D) = [T_1(D)x_1(D)\ T_2(D)x_2(D)\ \ldots\ T_p(D)x_p(D)]$ generates the same codeword using the recursive encoder with generator matrix $G_R(D)$. If the degree of $T_j(D)$ is chosen to be less than ν_j corresponding to every G_i, then the resulting encoder has exactly the same number of states as that of the non-recursive encoder and, hence, there is no increase in decoding complexity. For this recursive encoder a weight one input sequence generates an infinite weight output sequence and every finite output weight error event corresponds to an input weight of at least 2. In general, the recursive codes generated thus are not systematic. Since it is the recursive nature of the encoders that is important in a concatenated scheme, we do not bother about the non-systematic nature of these realizations.

These ideas are explained more clearly using the following two examples. First, consider the 4-state, 4-PSK space-time[1] code employed by a 2-antenna transmitter as described in [2] and shown in Fig. 38. It may be represented by the following transfer function matrix:

$$G(D) = \begin{bmatrix} D & 0 & 1 & 0 \\ 0 & D & 0 & 1 \end{bmatrix}$$

The transfer function matrix of the recursive form (Fig. 38) of this code is given by:

$$G_R(D) = \begin{bmatrix} \frac{D}{1+D} & 0 & \frac{1}{1+D} & 0 \\ 0 & \frac{D}{1+D} & 0 & \frac{1}{1+D} \end{bmatrix}$$

[1]The term "space-time code," in this chapter, would refer to space-time trellis codes unless otherwise stated.

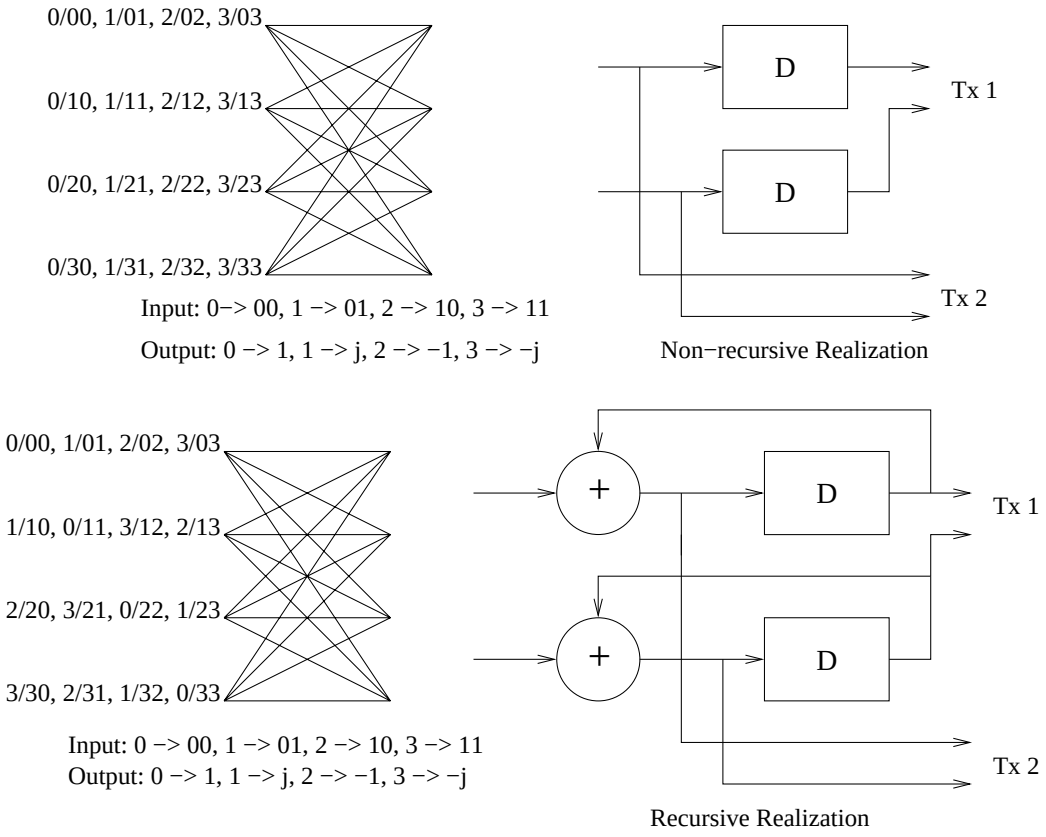

Fig. 38. 4-state, 4-PSK space-time code: Trellises, non-recursive and recursive realizations.

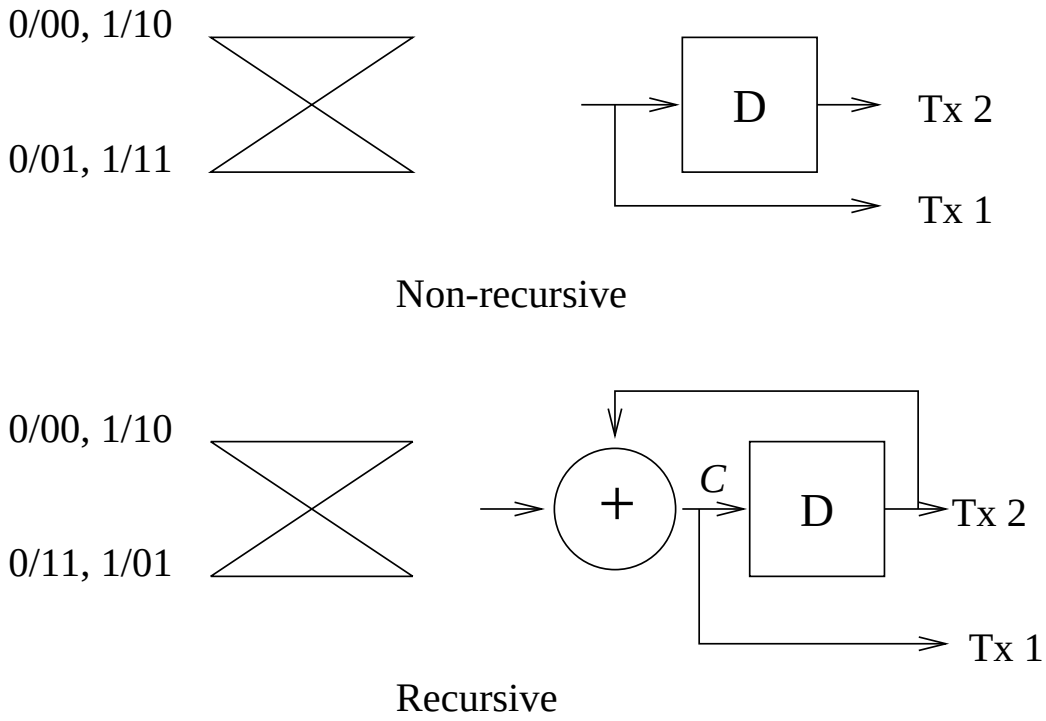

Fig. 39. 2-state, BPSK delay diversity: Trellises, non-recursive and recursive realizations.

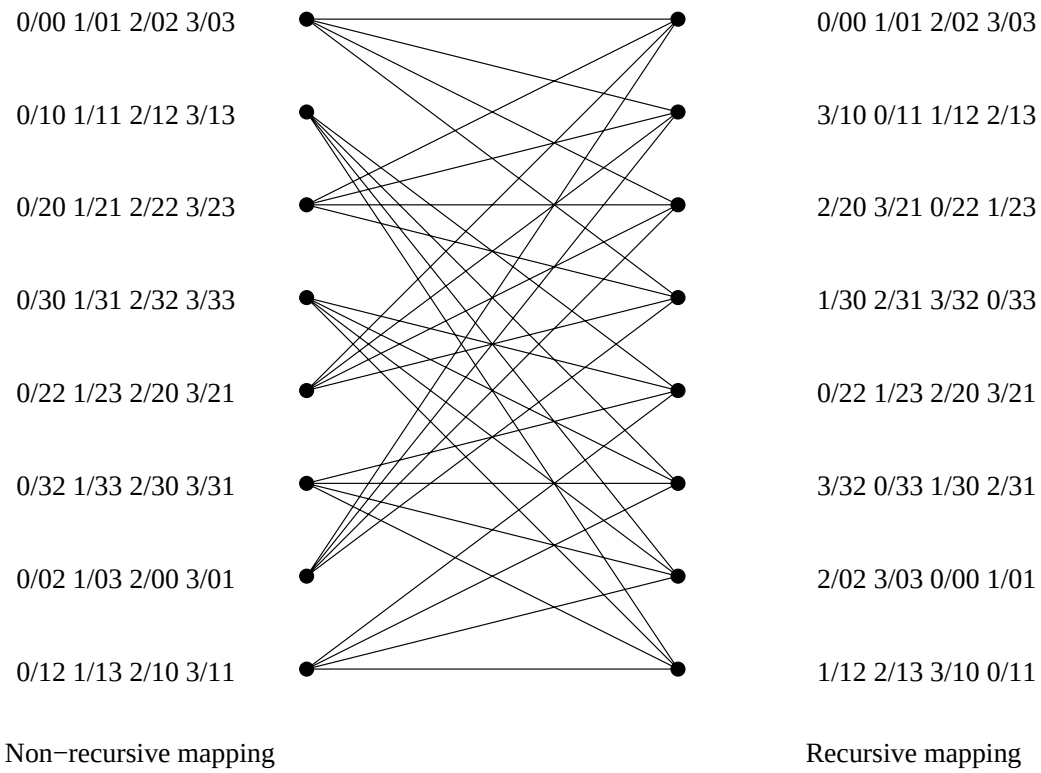

Fig. 40. 8-state, 4-PSK space-time code: Non-recursive and recursive labeling of the trellis.

Both these codes produce exactly the same set of codewords, though the input producing a particular codeword may be different. Similarly, a simple BPSK based two antenna delay diversity scheme [81] (Fig. 39) has the following non-recursive and recursive generator matrices:

$$G(D) = \begin{bmatrix} D & 1 \end{bmatrix}$$
$$G_R(D) = \begin{bmatrix} \frac{D}{1+D} & \frac{1}{1+D} \end{bmatrix}$$

Several known space-time codes are LSR based and hence their recursive versions are easy to find.

2. Trellis Re-labeling for Non-LSR Based STTC (Space-time Trellis Codes)

Some space-time codes such as those in [2] are directly specified in terms of the trellis. That is, the output symbols to be transmitted along each branch in the trellis specify the code, without specifying a particular encoder structure. We refer to these as Non-LSR based space-time codes. In some cases, it is possible to obtain an equivalent LSR based structure for a Non-LSR based space-time code. However, the technique we discuss here is more general and encompasses such codes as well. Recall that for a given trellis, we wish to redefine the mapping from the input symbols to the set of output symbols such that the resulting code is recursive. As an example, consider the 8-state, 4-PSK space-time code of [2]. In reference [2], the mapping from the input to the trellis transition has not been defined explicitly. Fig. 40 shows both the non-recursive and recursive representations of this code. It should also be noted that there may exist more than one recursive versions of a given code. In this chapter we do not consider the effect of this choice on the performance of the proposed system.

C. Performance Analysis

In this section we derive the union bound on the codeword error probability for a simple code belonging to the class of serial Turbo space-time codes. The code consists of a concatenation of an outer convolutional code and an inner space-time code which is a recursive delay-diversity scheme and the modulation is BPSK. Further, two transmit antennas and one receive antenna are considered. The recursive delay diversity scheme is nothing but a differential encoder followed by a delay diversity scheme as shown in Fig. 39. Let C^S denote the overall concatenated code i.e. the set of all possible codewords at the output of the differential encoder (sequence C in Fig. 39).

We start with the pairwise error probability between two codewords C and E for a quasi-static fading channel for any space-time code given by [2, Eqn 10]:

$$P(\mathbf{C} \to \mathbf{E}) \leq \left(\prod_{i=1}^{r} \lambda_i \right)^{-m} \left(\frac{E_s}{4N_0} \right)^{-rm} \tag{6.4}$$

where

$$\mathbf{C} = \begin{pmatrix} c_1^1 & c_2^1 & \cdots & c_N^1 \\ c_1^2 & c_2^2 & \cdots & c_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^n & c_2^n & \cdots & c_N^n \end{pmatrix}$$

is the transmit codeword,

$$\mathbf{E} = \begin{pmatrix} e_1^1 & e_2^1 & \cdots & e_N^1 \\ e_1^2 & e_2^2 & \cdots & e_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_1^n & e_2^n & \cdots & e_N^n \end{pmatrix}$$

is the (erroneous) decision made by the maximum-likelihood decoder, m is the number of receive antennas, λ_i are the non-zero eigenvalues of the matrix $A(\mathbf{C},\mathbf{E})$ and $r = \min(n,m)$ is the rank of the matrix $B(\mathbf{C},\mathbf{E})$, where:

$$\begin{aligned} B(\mathbf{C},\mathbf{E}) &= \begin{pmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \cdots & \cdots & e_N^1 - c_N^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \cdots & \cdots & e_N^2 - c_N^2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ e_1^n - c_1^n & e_2^n - c_2^n & \cdots & \cdots & e_N^n - c_N^n \end{pmatrix} \\ &= \mathbf{E} - \mathbf{C} \\ A(\mathbf{C},\mathbf{E}) &= (\mathbf{E} - \mathbf{C})(\mathbf{E} - \mathbf{C})^H \\ &= \mathbf{C}\mathbf{C}^H + \mathbf{E}\mathbf{E}^H - \mathbf{E}\mathbf{C}^H - \mathbf{C}\mathbf{E}^H \end{aligned}$$

For the case of a delay-diversity code, as is what is of interest here, it is easy to see that the rows of $B(\mathbf{C},\mathbf{E})$ are linearly independent and, hence, the matrix A has rank n, thus guaranteeing full diversity. For the particular case of $n = 2$, $m = 1$ and BPSK, the product of the determinant of the Eigenvalues can be reduced to certain parameters of the codewords as shown in [82] and explained below. For BPSK modulation, the Hermitian is just the transpose. The elements of $\mathbf{C}$ and $\mathbf{E}$ are either $+1$ or -1. Let $c^{(i)}(i = 1, 2)$ be the binary codeword to be transmitted over the i-th antenna if the codeword $\mathbf{C}$ is being transmitted. Further, since the inner code is a linear code, we can assume that the binary transmit codeword $\mathbf{C}$ is the all zeros codeword, i.e. $c^{(i)} = 0, (i = 1, 2)$. We define the following quantities:

$$\begin{aligned} d_1(\mathbf{E}) &= d_H(c^{(1)} \oplus e^{(1)}) = d_H(e^{(1)}) \\ d_2(\mathbf{E}) &= d_H(e^{(2)}) \\ d_3(\mathbf{E}) &= d_H(e^{(1)} \oplus e^{(2)}) \end{aligned}$$

where $d_H(a)$ refers to the Hamming weight of a and $\oplus$ is the modulo two sum operator. Thus, we may write:

$$\begin{aligned}
\mathbf{C}\mathbf{C}^H &= N \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
\mathbf{E}\mathbf{E}^H &= \begin{bmatrix} N & N-2d_3(\mathbf{E}) \\ N-2d_3(\mathbf{E}) & N \end{bmatrix} \\
\mathbf{E}\mathbf{C}^H &= \begin{bmatrix} N-2d_1(\mathbf{E}) & N-2d_1(\mathbf{E}) \\ N-2d_2(\mathbf{E}) & N-2d_2(\mathbf{E}) \end{bmatrix} \\
\mathbf{C}\mathbf{E}^H &= \begin{bmatrix} N-2d_1(\mathbf{E}) & N-2d_2(\mathbf{E}) \\ N-2d_1(\mathbf{E}) & N-2d_2(\mathbf{E}) \end{bmatrix} \\
\Rightarrow\ \mathbf{A}(\mathbf{C},\mathbf{E}) &= 2 \begin{bmatrix} 2d_1(\mathbf{E}) & (d_1(\mathbf{E})+d_2(\mathbf{E})-d_3(\mathbf{E})) \\ (d_1(\mathbf{E})+d_2(\mathbf{E})-d_3(\mathbf{E})) & 2d_2(\mathbf{E}) \end{bmatrix}
\end{aligned}$$

The product of the eigenvalues of the matrix $\mathbf{A}(\mathbf{C},\mathbf{E})$:

$$\begin{aligned}
\lambda_1\lambda_2 &= 2\left(2d_1(\mathbf{E}) \times 2d_2(\mathbf{E}) - (d_1(\mathbf{E})+d_2(\mathbf{E})-d_3(\mathbf{E}))^2\right) \\
&= 2\left[(d_1(\mathbf{E})+d_2(\mathbf{E})+d_3(\mathbf{E}))^2 - d_1^2(\mathbf{E}) - d_2^2(\mathbf{E}) - d_3^2(\mathbf{E})\right]
\end{aligned}$$

For the case of delay-diversity, since $e^{(2)}(D) = D\ e^{(1)}(D)$ (that is, the sequence transmitted from one antenna is a time-shifted version of the other for all codewords), $d_1(\mathbf{E}) = d_2(\mathbf{E})$. The union bound on the probability of word error is given by

$$P(e) \leq \sum_{\mathbf{E}} P(\mathbf{C} \rightarrow \mathbf{E}) \tag{6.5}$$

which can be expressed as

$$P(e) \leq \sum_{d} \sum_{\mathbf{E}:d_1(\mathbf{E})=d} P(\mathbf{C} \rightarrow \mathbf{E}) \tag{6.6}$$

Since each codeword $\mathbf{E}$ is a serially concatenated codeword it corresponds to an outer-codeword, say $\mathbf{E}^o$ of Hamming weight $d^o(\mathbf{E}) = l$. Then, the union bound can be re-expressed as

$$P(e) \leq \sum_{l} \sum_{d} \sum_{\mathbf{E}:d_1(\mathbf{E})=d, d^o(\mathbf{E})=l} P(\mathbf{C} \rightarrow \mathbf{E}) \tag{6.7}$$

Now, we will show that for the particular case of a recursive delay diversity inner code, $d^o(\mathbf{E})$ is directly related to $d_3(\mathbf{E})$. The differential encoder has a transfer function

$$T^i(X, Y, Z) = \frac{X^2 Y Z^2}{1 - YZ} = X^2 Y Z^2 [1 + YZ + Y^2 Z^2 + Y^3 Z^3 + \ldots +] \tag{6.8}$$

where the powers of X, Y and Z denote the input weight, output weight and length of the error event. It is clear that every error event for the inner code corresponds to an input-weight of 2. That is, an outer codeword of weight l produces exactly $l/2$ error events if l is even and an infinite output weight if l is odd. Since the blocks are terminated, we can assume that the number of error events is $\lceil l/2 \rceil$. It can also be seen from (6.8) that the output weight of an error event is the same as the length. That is, all error events are of the form $(1, 1, 1, \ldots, 1)$. Since each codeword $\mathbf{E}$ can be thought of as the concatenation of one or more error events (since the code is linear), it follows that $\mathbf{E}$ and $e^{(2)}$ have the form:

$$\mathbf{E} = e^{(1)} = (0, 0, \ldots, 1, 1, \ldots, 1, 0, 0, \ldots, 1, 1, 1, \ldots, 1, \ldots, 0, 0) \tag{6.9}$$

$$e^{(2)} = (0, 0, \ldots, 0, 1, 1, \ldots, 1, 0, 0, \ldots, 1, 1, 1, \ldots, 1, \ldots, 0, 0) \tag{6.10}$$

where there are exactly $l/2$ concatenations of strings of 1s. The second equality follows since $e^{(2)}(D) = De^{(1)}(D)$. Recalling that $d_3(\mathbf{E})$ is the Hamming distance between the coded streams transmitted from the two antennas, it follows that $d_3(\mathbf{E}) = 2 \times (l/2) = l$. Here we have excluded those codewords which are concatenations of error events

such that one error event begins at the next time instant another error event ends. In that case, both error events together would appear to be a longer error event. A simple counting exercise will show that the probability of such error events occurring decreases with increasing N and tends to zero for large N and, hence, can be ignored. Therefore, the union bound can be rewritten as

$$\begin{aligned} P(e) &\leq \sum_{l}\sum_{d} A^{C_s}(l,d)\frac{1}{(2d+l)^2-2d^2-l^2}\left(\frac{E_s}{4N_o}\right)^{-2} \\ &= \sum_{l}\sum_{d} A^{C_s}(l,d)\frac{1}{2d(d+2l)}\left(\frac{E_s}{4N_o}\right)^{-2} \end{aligned} \tag{6.11}$$

where $A^{C_s}(l,d)$ is the number of codewords with outer code weight l and overall codeword weight d. For a serially concatenated code with a uniform interleaver, $A^{C_s}(l,d)$ is given by

$$A^{C_s}(l,d) = \frac{A^{C_o}(l) \times A^{C_i}(l,d)}{\binom{N}{l}} \tag{6.12}$$

where $A^{C_o}(l)$ is the number of codewords of the outer code with weight l and $A^{C_i}(l,d)$ is the number codewords of the inner code (differential encoder) with weight d corresponding to an input of weight l. Using the result in [77] this can be expressed as:

$$A^{C_s}(l,d) \leq \sum_{l}\sum_{d}\sum_{n^o}\sum_{n^i} N^{n^o+n^i-l-1}\frac{l^l l!}{n^o! n^i!}A^{C_o}(l,n^o)A^{C_i}(l,d,n^i) \tag{6.13}$$

where $A^{C_o}(l,n^o)$ is the number of codewords of the outer code with weight l corresponding to exactly n^o concatenations of error events and $A^{C_i}(l,d,n^i)$ is the number of codewords of the inner code with input weight l, output weight d and n^i concatenations of error events. This when substituted in (6.11) yields:

$$P(e) \leq \sum_{l}\sum_{d}\sum_{n^o}\sum_{n^i} N^{n^o+n^i-l-1}\frac{l^l l!}{n^o! n^i!}\frac{A^{C_o}(l,n^o)A^{C_i}(l,d,n^i)}{4d(d+2l)}\left(\frac{E_s}{4N_o}\right)^{-2} \tag{6.14}$$

Following the derivations in [77], we consider the maximum exponent of the codeword length N. For a non-recursive code, $n^i = l$, and so the exponent of N is always positive. Since $n^i \leq l/2$ for a recursive code and $n^o \leq l/d^o_{min}$ for any code, the maximum exponent of N is negative only if $min(l) = d^o_{min} \geq 3$. This key result can be summarized as follows.

If the inner code is a recursive delay diversity scheme with BPSK modulation and the outer code has a minimum distance $d^o_{min} \geq 3$, the concatenation scheme is guaranteed to achieve full diversity and the union bound on the probability of word error decreases exponentially in the length N. Though we have not derived a closed form expression for the union bound when other recursive space-time codes are used or a higher modulation format is used, the proof above provides an indication of the expected results.

D. Simulation Results

For simulations, we use the recursive form of the 4-state code (Fig. 38) and the 8-state code (Fig. 40) with QPSK modulation, as the inner code. An s-random interleaver [83] is used to interleave the bits between the outer and the inner code. Before transmission, the symbols to be transmitted on a particular antenna are interleaved using a block interleaver. In order to demonstrate that it is the recursive nature of these codes that provides the interleaving gains, results for the non-recursive version of the 4-state code are also included.

1. Convolutional Outer Code

We study the performance of the proposed scheme with a 2-state ($[1, 1 + D]$) and a 4-state ($[1 + D^2, 1 + D + D^2]$), rate-1/2 convolutional codes as outer codes. Since the

rate of the outer code is 1/2, this setup achieves a spectral efficiency of 1 bits/sec/Hz with QPSK modulation. All simulation setups have 2 transmit antennas, 1 receive antenna, QPSK modulation and transmission block length of 1024 symbols.

Figure 41 shows the performance of the proposed scheme over a quasi-static fading channel. Seven iterations were used in the decoder. The recursive inner code performs about 4 dB better than the non-recursive scheme. No significant improvement was observed by using a more complex outer code or a more complex inner code.

In Fig. 42, we show the performance of the serial concatenation scheme when a recursive realization of the CYV code [1] is used. We choose $N_t = N_r = 2$ so that we can compare these results to those in Figs. 18,33. Once again, observe that the non-recursive inner code does not give good performance. A simpler outer code is better and increasing the information block length from 1024 to 32678 gives only a slight improvement in performance. The performance of this scheme is about 2 dB away from the constrained limit.

2. Single Parity Check Turbo Product Outer Code

In this section, we consider the concatenation of recursive space-time codes with a very high rate outer code (say 0.9 or higher). The objective is to get improved power efficiency with minimal sacrifice in data rate. From Section C, we can see that in order to get an interleaving gain, the minimum distance of the outer code in a concatenation scheme should be at least 3. Since the decoder is an iterative decoder, we need to find codes which have $d_{min} \geq 3$ for very high rates (0.9 or higher) and that are easily decodable. One approach is to puncture convolutional codes to rates around 0.9. However, obtaining $d_{min} \geq 3$ with such high puncturing rates would require the constraint length to be very high. This, in turn, will result in an exponential increase

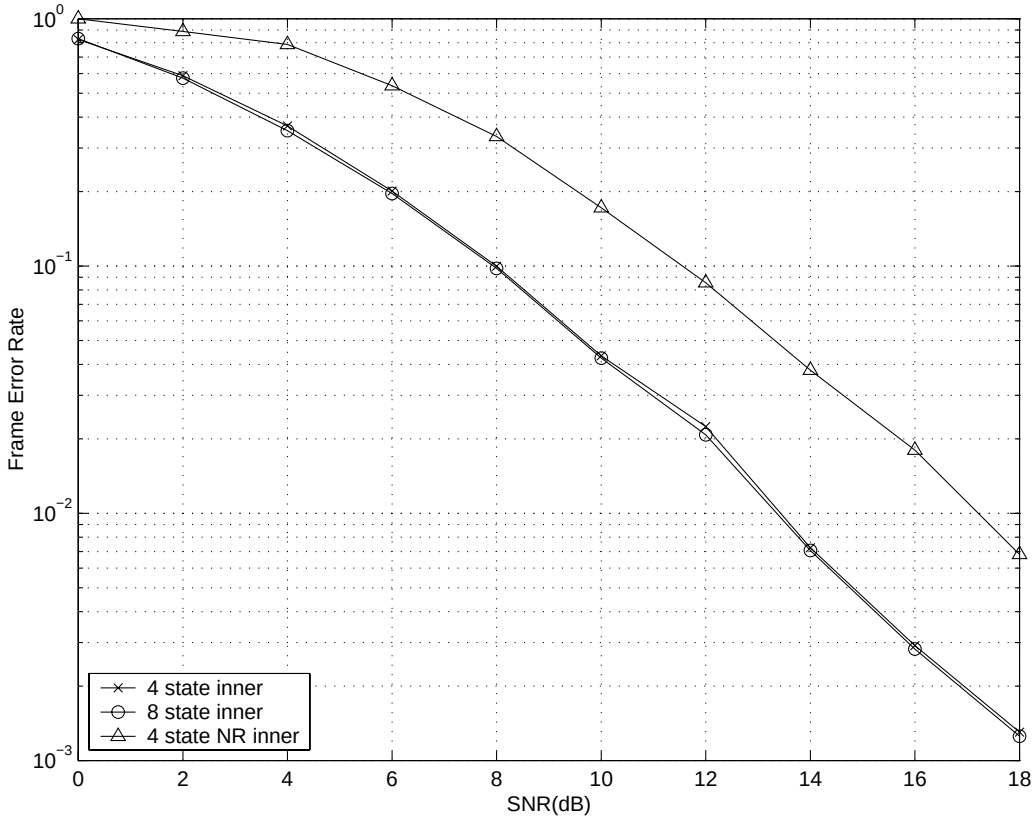

Fig. 41. Performance of a serial concatenation of $[1, 1 + D]$ convolutional outer code with a recursive realization of the 4-state delay diversity code. $N_t = 2, N_r = 1$, QPSK, 1 b/s/Hz.

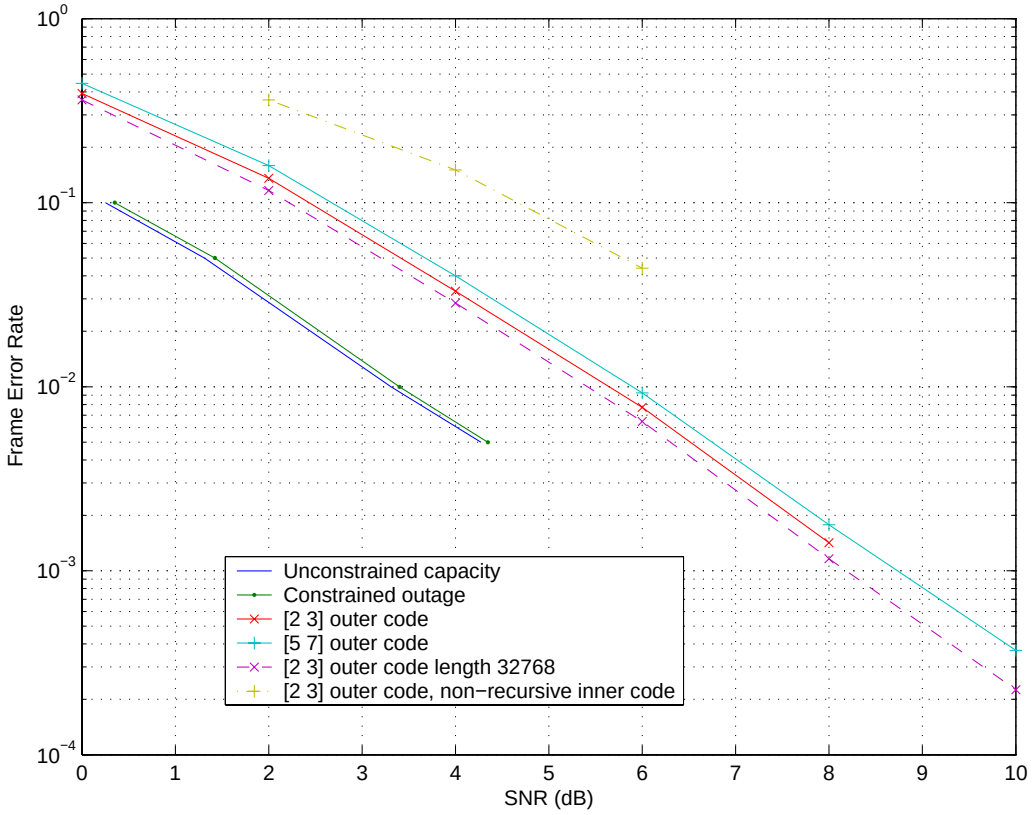

Fig. 42. Performance of a serial concatenation of different convolutional outer codes with a recursive realization of the CYV code. $N_t = N_r = 2$, QPSK, 1 b/s/Hz.

in the decoding complexity.

As an alternative, we propose to use two dimensional product codes where each of the dimensions is a $K/K+1$ single parity check code. Hence, the overall rate is $(K/K+1)^2$. In general, the input stream can be divided into several such blocks of length K^2 to obtain a higher interleaving gain while maintaining the same rate. We consider an example with $K = 31$ and concatenation of two such blocks. Hence, the length of the product code is 2048 bits. A spectral efficiency of 0.94 b/s/Hz is achieved with BPSK or 1.88 b/s/Hz with 4-PSK assuming independent coding on the I and Q channels. This is a very small reduction in data rate. The motivation for the choice of single parity check based product codes are that (i) For any rate the minimum distance of these codes is 4 [84] and, hence, we expect an interleaving gain (ii) they can be soft-decoded using a belief propagation algorithm with very little decoding complexity since each of the dimensions in this code is a single parity check. Here, each iteration of the decoder consists of 3 iterations within the product code. This is sufficient since the outer code is 2-dimensional [84].

Fig. 43 compare this scheme with the 16-state and 32-state 4-PSK based space-time trellis codes from [2] which achieve a spectral efficiency of 2 b/s/Hz. Although we are comparing two codes with marginally different spectral efficiencies, the main point is that for a small sacrifice in spectral efficiency (9% here, for QPSK), significant improvement in power efficiency can be achieved with the proposed scheme. Further, it should be noted that the decoding complexity of the proposed scheme is lesser than that for Viterbi decoding of the 32-state 4-PSK codes. The Alamouti scheme does not perform well in comparison to these higher complexity trellis codes and so we omit plotting its performance.

The proposed scheme performs better than the 8-state and 32-state codes by about 2.5 dB and 2 dB respectively.

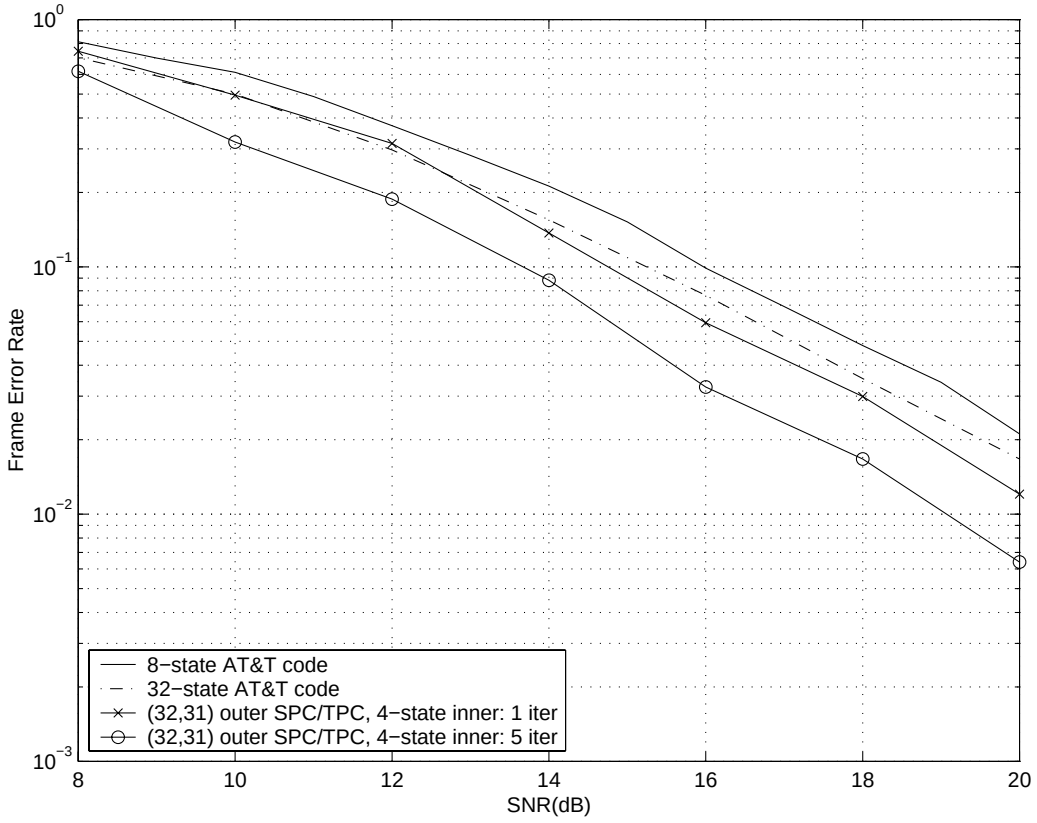

Fig. 43. Performance of the serial concatenation scheme with a SPC/TPC outer code.

E. Conclusions

We have proposed concatenation schemes based on recursive realizations of space-time codes. The concatenation scheme is based on serial concatenation of convolutional codes where the inner code is a space-time code. We have shown that on the quasi-static fading channel full diversity can be achieved and an increase in coding gain results from the concatenation. We will later show that this scheme provides significant gains on time varying channels (even when there are as low as 4 independent fading blocks per codeword) and independent fading channels. We have also shown that turbo product codes based on single parity check codes can be used as high rate codes that provide these advantages at very minimal reduction in data rate.

From the application point of view, the proposed family of codes offers a wide variety of choices. Consider, for example, the mobile users receiving data from a base station. Different users have different priorities. Some users may be willing to sacrifice both data rate and performance to keep the complexity of the receiver low. Other users may require a better performance at the cost of reduced data rate and/or a more complex receiver. The proposed class of codes offers an easy and natural way to provide such differentiated services with the same encoder structure. In another scenario, a transmitter can dynamically change the outer code being used to make a more effective use of the channel conditions. The scheme also provides a natural differentiation among various users based upon the decoding complexity they can afford. For example, if a user cannot afford the latency of 4 iterations, just one iteration may be used. Finally, since most space-time trellis codes have recursive realizations, users who are able to afford the complexity may use space-time codes with larger number of states, higher constellation size and/or more than one receive antennas to obtain improved performance.

CHAPTER VII

EXTENSIONS[†]

In this chapter, we consider various extensions of the ideas presented in the previous chapters. First, we extend the idea of recursive space-time trellis codes to the case of parallel concatenated codes and apply it to an automatic repeat-request (ARQ) system. In the second half of the chapter, we present results for the independent (ergodic) and block fading channels.

A. ARQ Scheme Using Recursive Space-time Trellis Codes

Data transmission systems will most likely use re-transmission schemes (ARQ) to reduce the frame error rate and, hence, it is important to design efficient ARQ schemes for use with space-time codes. Recursive realizations of space-time encoders provide a convenient way to design ARQ schemes based on the turbo coding principle. In the proposed ARQ scheme the first transmission uses a recursive space-time encoder to encode the information $\mathbf{x}$. The output of the space-time encoder from the ith transmit antenna $S_{i,k}$ is transmitted over a fading channel with instantaneous gain $\alpha^1_{i,k}$. The signal at the receive antenna is given by:

$$r^1_k = \sum_i \alpha^1_{i,k} S_{i,k} + n_k \tag{7.1}$$

The received signal vector $\mathbf{r}^1 = (r^1_0, r^1_1, \ldots, r^1_{N-1})$ is decoded and checked for errors (an error detection code such as cyclic redundancy check code is assumed). If the resulting vector has any errors, a re-transmission is requested. At the encoder, the data

[†]

sequence $\mathbf{x}$ is interleaved and encoded using the same recursive encoder as shown in Fig. 44. It can be readily seen that the two transmissions together represent a parallel concatenated convolutional code with *recursive* component codes. The receiver utilizes both the received frames $\mathbf{r}^1$ and $\mathbf{r}^2$ to iteratively decode the data [76,78]. Decoding proceeds in an iterative fashion until either the frame is correctly decoded or the number of iterations crosses a threshold. In the latter case, another re-transmission is requested and the set-up now becomes similar to multi-level parallel concatenated codes of [85]. Thus, after each re-transmission, the overall code corresponds to a lower rate code.

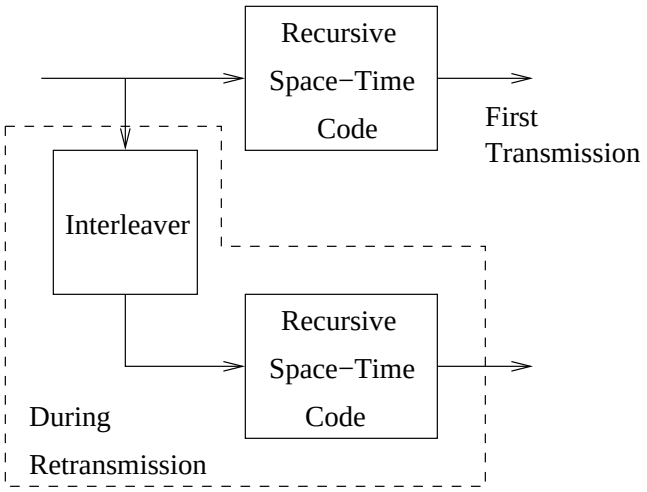

Fig. 44. Encoder structure for the ARQ system.

B. Performance of the ARQ Scheme

The set of codewords for the recursive and non-recursive space-time encoder are identical and, hence, the probability of frame error (or probability of a re-transmission) is identical to both encoders. That is, the use of recursive space-time encoder does not

affect the performance during the first transmission whereas after two transmissions, provides significant improvement in performance due to the parallel concatenated code structure. Since the component encoders are recursive, a significant interleaving gain results. Without the recursive realization, after two transmissions, the parallel concatenated code has non-recursive component encoders and, hence, no interleaving gain will result. In order to show the efficacy of the proposed scheme, we compare this scheme to ARQ schemes with non-recursive space-time encoders and also to a system where the re-transmission is identical to the original transmission. The latter system allows for a maximum-likelihood combining of the two transmissions using a Viterbi decoder with a modified metric. The received signal during the k-th stage of the trellis during the two transmissions may be written in a matrix form as:

$$\begin{bmatrix} r_k^1 \\ r_k^2 \end{bmatrix} = \begin{bmatrix} \alpha_{0,k}^1 & \alpha_{1,k}^1 & \cdots & \alpha_{m-1,k}^1 \\ \alpha_{0,k}^2 & \alpha_{1,k}^2 & \cdots & \alpha_{m-1,k}^2 \end{bmatrix} \begin{bmatrix} S_{0,k} \\ S_{1,k} \\ \vdots \\ S_{m-1,k} \end{bmatrix} + \begin{bmatrix} n_k^1 \\ n_k^2 \end{bmatrix} \tag{7.2}$$

$$\mathbf{R}_k = \mathbf{A}_k \, \mathbf{S}_k + \mathbf{N}_k \tag{7.3}$$

The noise terms n_k^1 and n_k^2 are i.i.d. complex, zero-mean, Gaussian random variables with variance $N_0/2$ in each dimension. So, the covariance of $\mathbf{N}_k$ is given by:

$$\mathbf{M} = \begin{bmatrix} N_0 & 0 \\ 0 & N_0 \end{bmatrix} \tag{7.4}$$

The maximum-likelihood (ML) decoding rule may be written as:

$$\tilde{\mathbf{S}}_k = \arg\max_S \frac{1}{(2\pi)|\mathbf{M}|} \exp\left(-\frac{1}{2} (\mathbf{R}_k - \mathbf{A}_k \, S)' \, \mathbf{M}^{-1} \, (\mathbf{R}_k - \mathbf{A}_k \, S) \right) \tag{7.5}$$

In the case of a two-transmit, one-receive antenna system, (7.5) simplifies to:

$$(\tilde{S}_{0,k}, \tilde{S}_{1,k}) \;=\; \arg\min_{s_0,s_1} \left[\left(r_k^1 - \alpha_{0,k}^1 s_0 - \alpha_{1,k}^1 s_1\right)^2 + \left(r_k^2 - \alpha_{0,k}^2 s_0 - \alpha_{1,k}^2 s_1\right)^2 \right] \quad (7.6)$$

This is the modified metric used in the Viterbi algorithm for jointly decoding the two received frames.

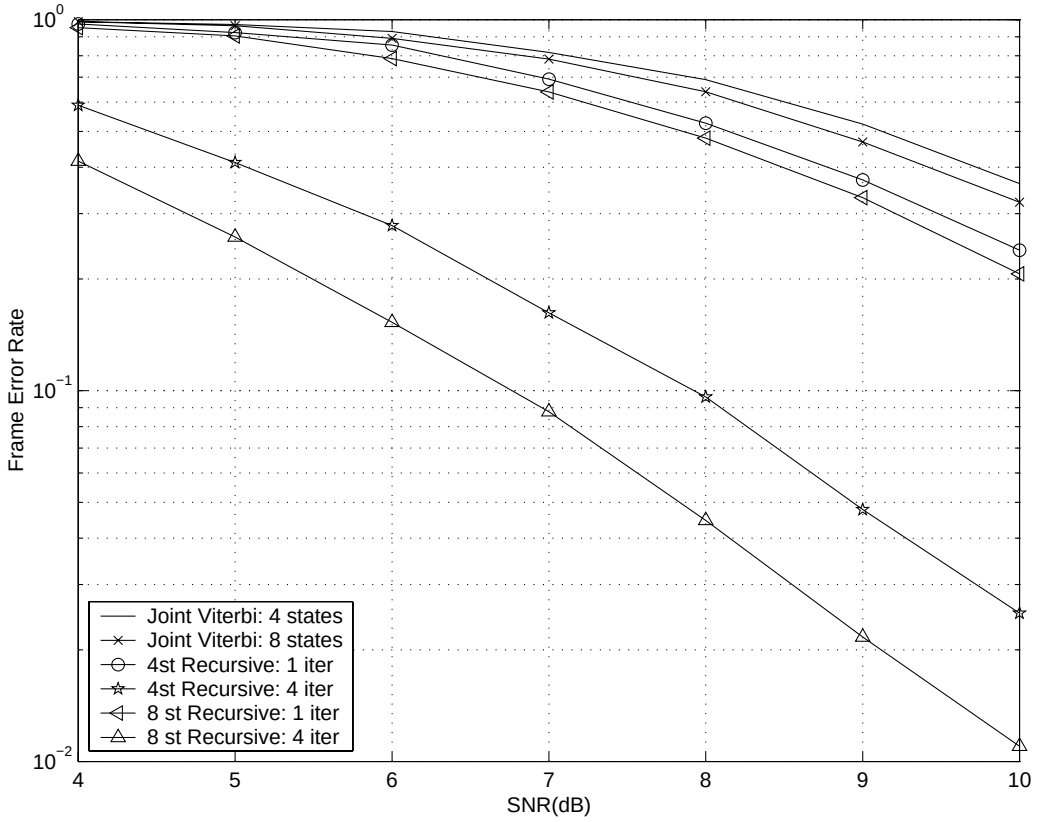

Fig. 45. ARQ scheme over quasi-static fading channel.

The proposed scheme has been simulated using the recursive realizations of the 4-state, 4-PSK code (Fig. 38) and the 8-state, 4-PSK code (Fig. 40) proposed in [2]. The data frame length, and hence the interleaver length as well as the transmitted frame length, is 4096 symbols. An s-random interleaver operating on a symbol basis

(as opposed to a bit interleaver) is used. The proposed scheme is compared to the ARQ scheme where no interleaver is used and the Viterbi algorithm is used to jointly decode the received frames. In order to demonstrate that it is the recursive nature of the code being used that yields interleaving gains, comparison with an identical scheme that uses the non-recursive realization of the same space-time codes is also done.

The performance measure used to make these comparison is the frame error rate after two transmissions (that is, one original transmission and one re-transmission). In case of iterative decoding, frame errors are compared after the first, the fourth iteration and the tenth iteration. Simulation results indicate that all the performance is achieved in about 4 iterations for quasi-static fading and in about 10 iterations for the independently fading channel. Figure 45 shows that on quasi-static fading channels, the joint Viterbi decoding performs worse than even the first iteration of the recursive codes. The non-recursive code (not shown in Fig. 45) performs almost the same as the joint Viterbi decoder and, as expected, their performance does not improve with iterations. At a frame error rate of 10^{-1}, the proposed scheme outperforms the Viterbi decoder by 1 dB after 1 iteration and by 4 dB after 4 iterations.

C. Serial Concatenation of Recursive Space-time Trellis Code over Independent Fading Channel

The independent fading channel is characterized by fast fading, such that fading experienced by adjacent symbols is independent. In practice such a channel model is useful when the channel is varying with time and a channel interleaver is used to break any correlation in channel variations. In the following we present a fast-fading analysis of the system presented in Chapter VI.sec:perfanal. The analysis is then

supported by simulation results.

1. Performance Analysis

When fading is rapid that two adjacent transmitted bits experience independent fading, then the pairwise error probability is given by [2, Eqn 17]:

$$P(\mathbf{C} \to \mathbf{E}) \leq \prod_{t \in \nu(\mathbf{C},\mathbf{E})} \left(|c_t - e_t|^2 \frac{E_s}{4N_o} \right)^{-m} \tag{7.7}$$

where $\nu(\mathbf{C}, \mathbf{E})$ denotes the set of time instants where the code matrices $\mathbf{C}$ and $\mathbf{E}$ differ. For linear codes, we can choose $\mathbf{C}$ to be the all-zeroes codeword. For BPSK constellations $|c_t - e_t|^2 = 4$ for the starting and ending time periods of each error event and $|c_t - e_t|^2 = 8$ for all time instants in between. To get an upper bound we simply consider $|c_t - e_t|^2 = 4$ for all $t \in \nu(\mathbf{C}, \mathbf{E})$. Also, the cardinality of the set $\nu(\mathbf{C}, \mathbf{E})$ in this case is bounded below by $d_1(\mathbf{E}) + 1$ and bounded above by $2d_1(\mathbf{E})$. Therefore, the pairwise error probability can be expressed as:

$$P(\mathbf{C} \to \mathbf{E}) \leq \left(\frac{E_s}{N_o} \right)^{-(d_1(\mathbf{E})+1)m} \tag{7.8}$$

Therefore, the union bound on the probability of error can be shown to be:

$$P_e \leq \sum_d \left(\frac{E_s}{N_o} \right)^{-(d+1)m} \tag{7.9}$$

$$= \sum_d \sum_l A^{C_s}(l,d) \left(\frac{E_s}{N_o} \right)^{-(d+1)m} \tag{7.10}$$

$$P_e = \left(\frac{E_s}{N_o} \right)^{-m} \sum_{d=d_{min}} \sum_l \sum_{n^o} \sum_{n^i} N^{n^o+n^i-l-1} \frac{l^l l!}{n^o! n^i!} \times A^{C_o}(l, n^o) A^{C_i}(l, d, n^i) \left(\frac{E_s}{N_o} \right)^{-md} \tag{7.11}$$

We see that the diversity order is m $(d_{min} + 1)$, where d_{min} is the minimum distance of the overall code. The interleaving gain assures us that the number of

codewords at a given distance d (and hence a diversity order $m(d+1)$) decreases exponentially in N if the inner code is recursive and $d^o_{min} \geq 3$. Equivalently, if a particular interleaver is chosen, the probability of obtaining a large diversity order will be high. In [86], Kahale and Urbanke have shown that the minimum distance of a serially concatenated code with a recursive inner code increases with the length N by the factor:

$$d_{min} \propto N^{\frac{d^o_{min}-2}{d^o_{min}}} \tag{7.12}$$

Therefore, as N increases we can conclude that the diversity order also increases at the same rate, namely $N^{\frac{d^o_{min}-2}{d^o_{min}}}$, resulting in significant improvement in performance. It should be noted here that increasing N does not increase the decoding complexity per decoded bit and, hence, diversity advantage results at the expense of latency.

Similar to the case of quasi-static fading, we have shown that in independent fading, the concatenated scheme with a recursive delay diversity as the inner code, an outer code with a large minimum distance and BPSK as the modulation format, it is possible to obtain asymptotically good performance. Since quasi-static and independent fading represents two extreme cases of block fading, these codes can be expected to perform well on the general block fading channel also.

The above results are useful when a maximum likelihood decoder is used and the performance with an actual iterative decoder is quite hard to characterize mathematically. Therefore, we study the performance through simulations in the following subsection.

2. Simulation Results

We study the performance of the proposed scheme with a 2-state ($[1, 1+D]$) and a 4-state ($[1+D^2, 1+D+D^2]$), rate-1/2 convolutional codes as outer codes. Since the

rate of the outer code is 1/2, this setup achieves a spectral efficiency of 1 bits/sec/Hz with QPSK modulation. All simulation setups have 2 transmit antennas, 1 receive antenna, QPSK modulation and transmission block length of 1024 symbols.

Fig. 46 shows the performance over the independently fading channel. Ten turbo iterations are needed to realize all the gains. It is observed that for the 2-state "weaker" outer code, the performance improves by having a stronger inner code. On the other hand, the performance of the 4-state outer code degrades when a stronger inner code is used. This is due to the nature of the iterative decoding algorithm rather than the code structure.

In Fig. 47, we demonstrate the performance over the block fading channel. Seven turbo iterations are used. While the recursive code has about a 4 dB gain over the non-recursive code, the performance of the iterative decoder is between the quasi-static fading performance and the independent fading performance. When the 4-state outer code is used, a weaker inner code performs better at low SNR whereas a stronger inner code performs better at high SNR. Again, this is an artifact of the iterative decoding algorithm.

Another advantage of the proposed class of codes is that it can be easily extended to more than two transmit antennas. We demonstrate this for three transmit antennas by considering three concatenated schemes based on delay diversity scheme. In all three cases, the outer code is a 2-state convolutional code with generator polynomial $[1, 1 + D]$. The inner codes for the three cases are - delay diversity scheme (non-recursive), delay diversity with recursive realization with feedback polynomial $[1+D^2]$, and, delay diversity with recursive realization with feedback polynomial $[1+D+D^2]$, respectively. The performance of these concatenation schemes for QPSK modulation and data frame size of 1024 bits, over independently fading channel is shown in Fig. 48. It is easily seen that the recursive delay diversity schemes are able to achieve greater

diversity benefit (steeper slope) than their non-recursive counterpart. Among the two recursive realizations, the $[1+D^2]$ does slightly better (about 0.5 dB) better. Note that for the 3-transmit antenna case, it is not possible to obtain simple orthogonal codes such as the Alamouti's scheme [26] and, therefore, it is not possible to concatenate a simple orthogonal block code with a more complex outer code, for example as in [87], to obtain good performance.

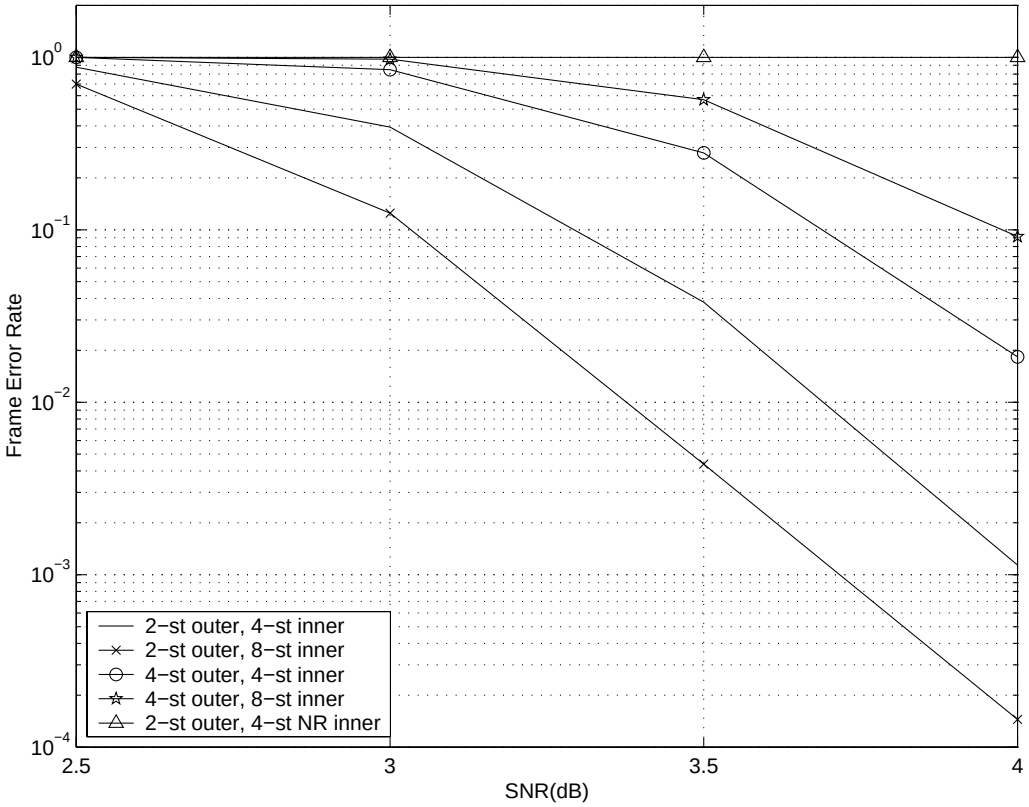

Fig. 46. Convolutional outer code: Performance over independent fading channel.

We also consider the concatenation of recursive space-time trellis codes with single-parity check turbo-product codes (SPC/TPC) in independent and block fading channels. We consider an example with $K = 31$ and concatenation of two such blocks.

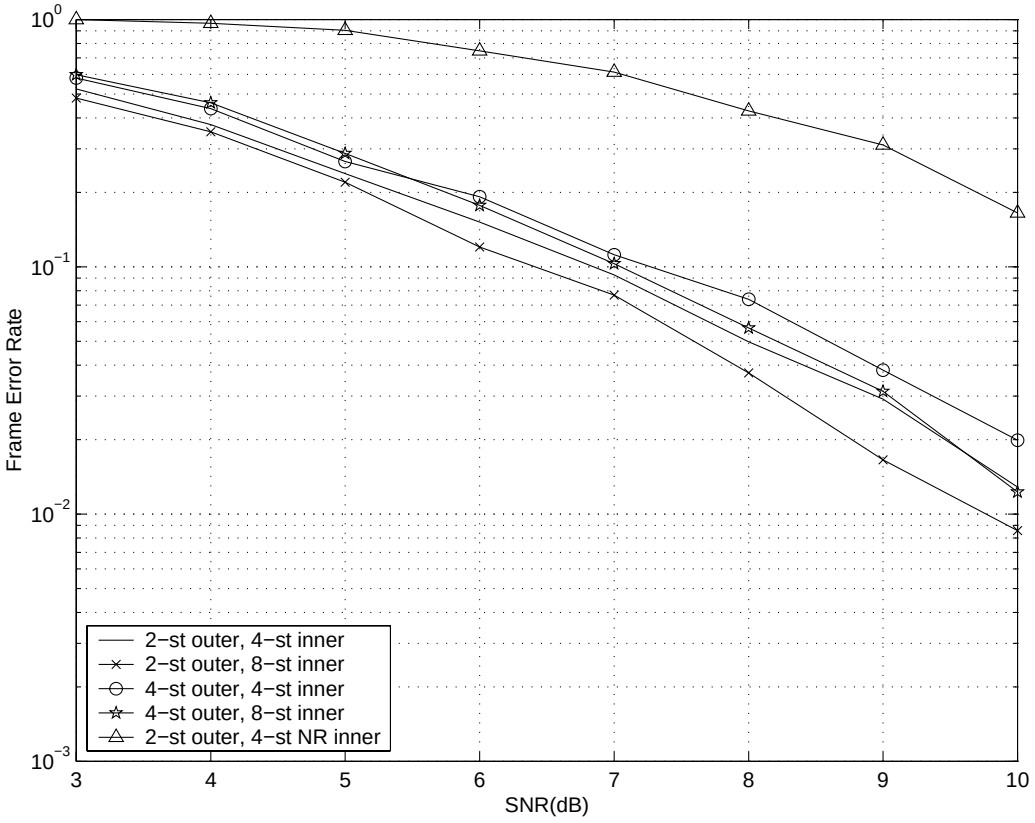

Fig. 47. Convolutional outer code: Performance over block fading channel - 4 fading blocks per codeword.

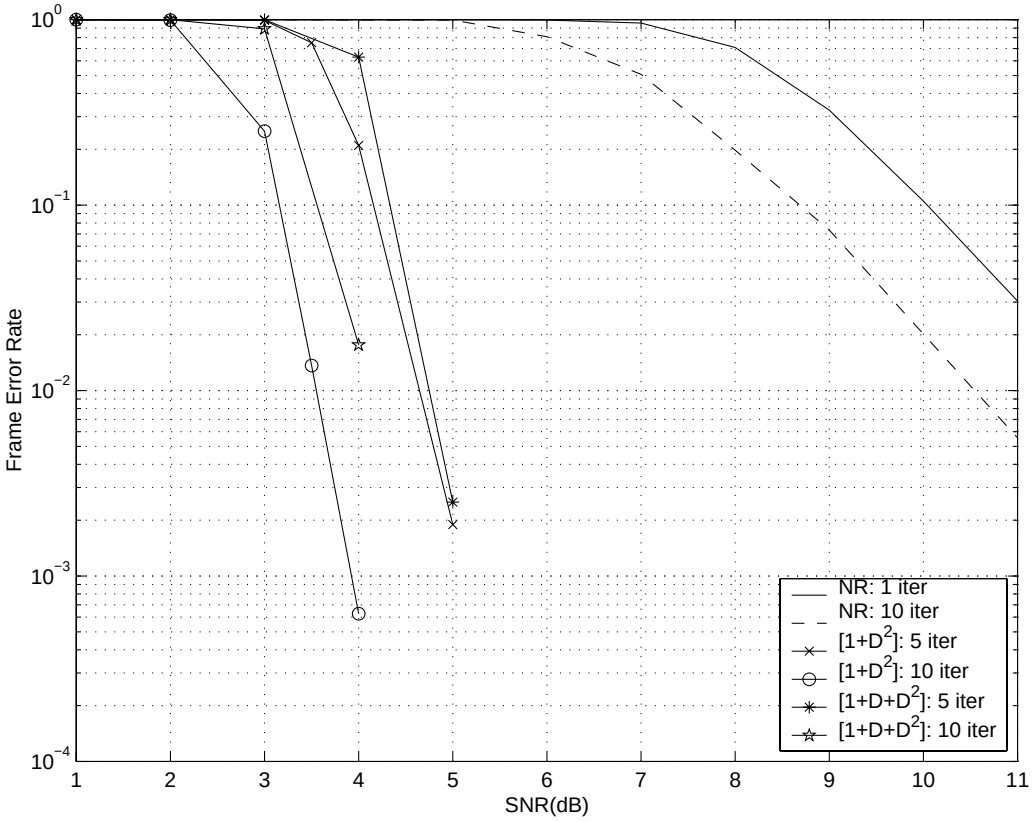

Fig. 48. Convolutional outer code: Three transmit antenna delay diversity inner code, $[1, 1 + D]$ 2-state outer code; over independent fading channel.

Hence, the length of the product code is 2048 bits. A spectral efficiency of 0.94 b/s/Hz is achieved with BPSK or 1.88 b/s/Hz with 4-PSK assuming independent coding on the I and Q channels. This is a very small reduction in data rate. The motivation for the choice of single parity check based product codes are that (i) For any rate the minimum distance of these codes is 4 [84] and, hence, we expect an interleaving gain (ii) they can be soft-decoded using a belief propagation algorithm with very little decoding complexity since each of the dimensions in this code is a single parity check. Here, each iteration of the decoder consists of 3 iterations within the product code. This is sufficient since the outer code is 2-dimensional [84].

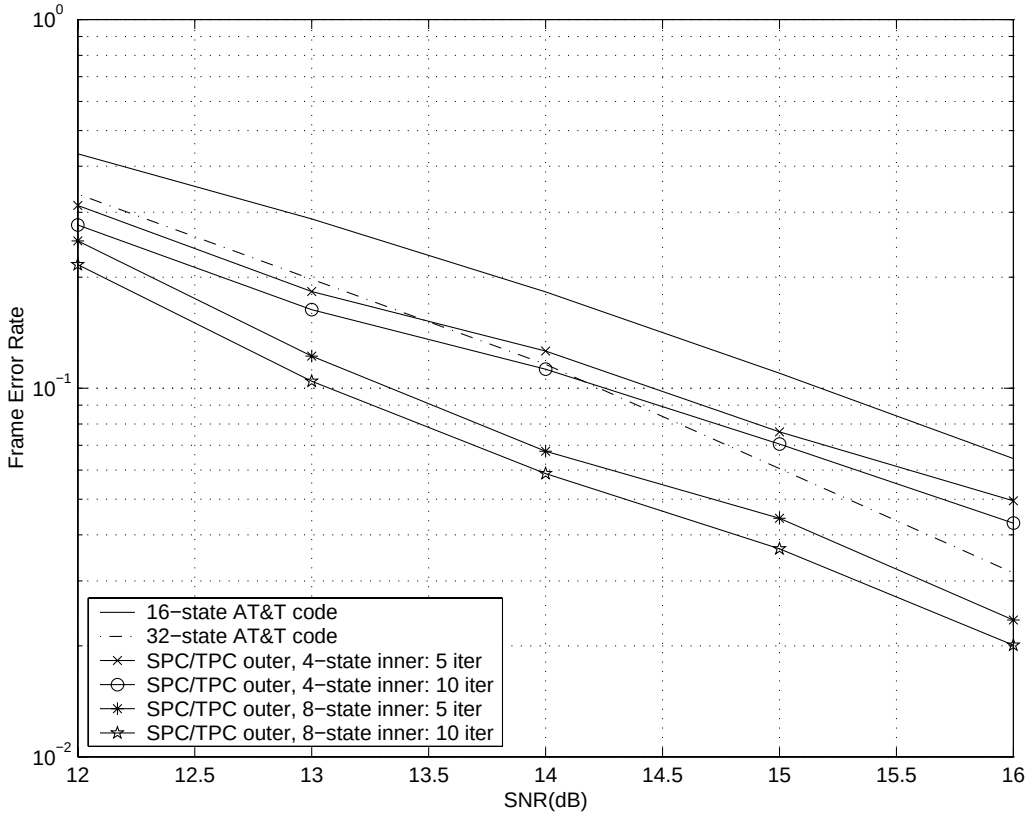

Fig. 49. SPC/TPC outer code: Performance over block-fading channel – 4 fading blocks per frame.

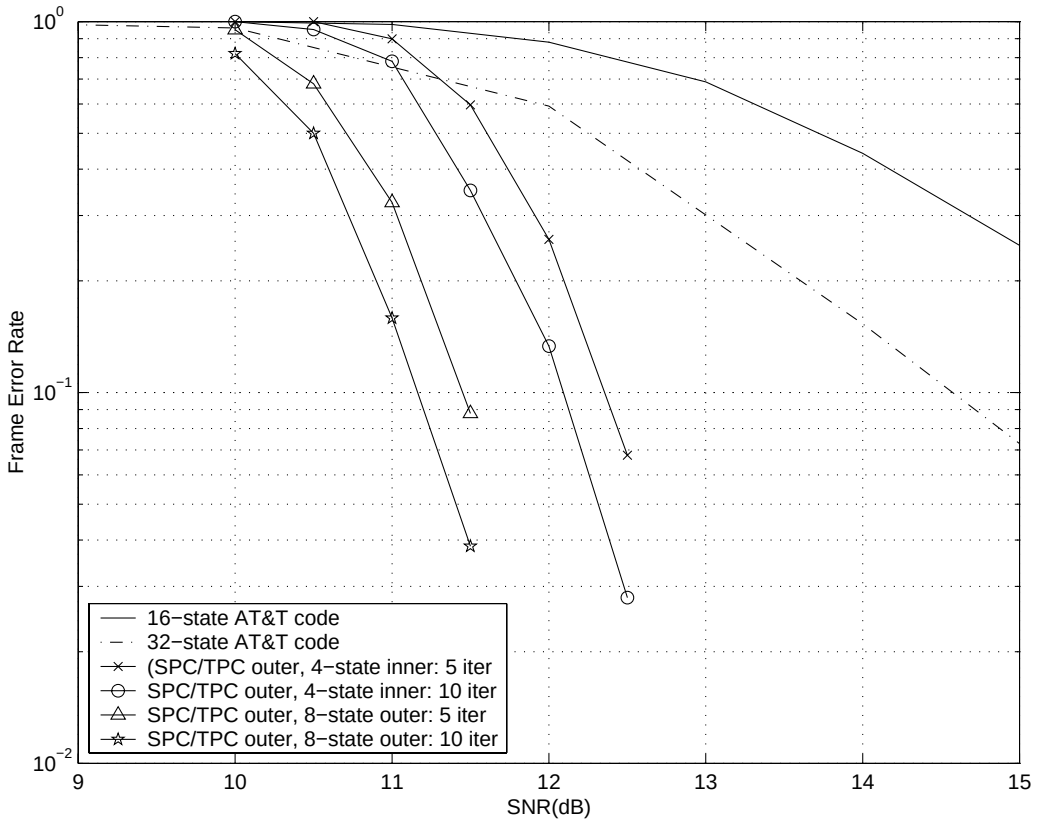

Fig. 50. SPC/TPC outer code: Performance over the independent fading channel.

Figures 49-50 compare this scheme with the 16-state and 32-state 4-PSK based space-time trellis codes from [2] which achieve a spectral efficiency of 2 b/s/Hz. Although we are comparing two codes with marginally different spectral efficiencies, the main point is that for a small sacrifice in spectral efficiency (9% here, for QPSK), significant improvement in power efficiency can be achieved with the proposed scheme. Further, it should be noted that the decoding complexity of the proposed scheme is lesser than that for Viterbi decoding of the 32-state 4-PSK codes. The Alamouti scheme does not perform well in comparison to these higher complexity trellis codes and so we omit plotting its performance.

For a block-fading channel with 4 fading blocks per frame, at a frame error rate of 0.1, the proposed scheme outperforms the 16-state code by about 3 dB after 7 iterations. In Fig. 50, the performance is compared over independent fading channels. The concatenated scheme achieves more diversity than the 32-state code after 5 iterations and the gain is between 2 dB (4-state inner code) to 3 dB at a frame error rate of 0.1. It is clear from the slope that even higher gains will be achievable at lower FERs.

D. Conclusions

In this chapter we have shown that serial concatenation of recursive space-time trellis codes with convolutional and SPC/TPC codes performs well in independent and block fading channels as well. We have shown before that on the quasi-static fading channel full diversity can be achieved and an increase in coding gain results from the concatenation. However, this is not the main advantage of the proposed scheme. The main advantage is that in addition to providing marginally improved performance to other more complex concatenated schemes on quasi-static fading channels, these

schemes provide significant gains on time varying channels (even when there are as low as 4 independent fading blocks per codeword) and independent fading channels. We have also shown that turbo product codes based on single parity check codes can be used as high rate codes that provide these advantages at very minimal reduction in data rate.

In general, the performance of the proposed codes will improve with increase in the length, whereas the performance of other schemes such as concatenation with an outer convolutional code, Reed-Solomon code or that of a simple space-time code will deteriorate with increase in length. Hence, the proposed scheme will offer higher advantage as the length of the codewords increases.

We have also proposed the use of recursive space-time codes for use in a turbo-type ARQ scheme when employing multiple transmit antennas. The advantage of this scheme is that iterative decoding is used only when needed. For example, if a more complex code is used to bring down the number of re-transmissions required and the channel happens to be "good", it will not justify the complex decoding required. On the other hand, if the channel happens to be "bad", a re-transmission will be needed anyway. Again the complexity is not justified. However, the proposed scheme would do better (in terms of complexity) in both these cases. Also, the proposed scheme is able to effectively utilize any time variations in the channel. For example, we observe that the gains over independent fading channels are very high. The scheme also provides a natural differentiation among various users based upon the decoding complexity they can afford. For example, if a user cannot afford the latency of 4 iterations, just one iteration may be used. Finally, since most space-time trellis codes have recursive realizations, users who are able to afford the complexity may use space-time codes with larger number of states, higher constellation size and/or more than one receive antennas to obtain improved performance.

CHAPTER VIII

CONCLUSIONS

This work focused on the multiple-input multiple-output (MIMO) flat quasi-static fading channel (QSFC). Prior to this work, the unconstrained outage capacity (with iid Gaussian signaling) was known [5, 6]. The code design for MIMO-QSFC was limited to designing schemes that achieve full spatial diversity. Such schemes cover a wide variety from space-time trellis codes [1, 2, 17–24] to space-time block codes [25, 26, 29–31] to spatial multiplexing and its variations [24, 34–38] There had been sporadic attempts to achieve higher rate by sacrificing diversity, for example in [32,33], or achieving lower rate by concatenating a space-time code with an outer code code, for example in [39–41, 43, 44].

The present work is the first to analyze the capacity of the MIMO-QSFC in a constrained modulation setting. We have shown why it is hard to compute this capacity exactly. We have presented achievable lower bounds on this capacity. These lower bounds also serve as fundamental limits on the performance of some known space-time codes. We have shown that the optimality of iid signaling does not carry over from the infinite alphabet (Gaussian signaling) case to the finite alphabet case. The suitability of orthogonal space-time block codes in low-rate, low-complexity setting has been established from an information theoretic perspective. We have shown that space-time trellis codes are near optimal for a wide range of rates.

This work has established an important negative result – systems that employ iterative decoding between the space-time demodulator/decoder and an outer code cannot have performance arbitrarily close to constrained modulation outage limit. This conclusion is especially significant since over the past few years iterative decoding has been considered the panacea to all communication problems. At the same time,

a word of caution. The technique used to arrive at this conclusion (namely, EXIT charts) is an approximate one. Whereas it can predict the trends, specific conclusions must be drawn carefully and cross-checked with actual simulations.

We have proposed a non-iterative transceiver structure based on the fact that perfect decision feedback is optimal in terms of mutual information. The encoding and decoding structures ensure that perfect decisions are available thanks to a capacity achieving low-density parity check (LDPC) code. These decisions are then utilized by a low-complexity space-time trellis decoder based on the BCJR algorithm. The system performance is within a dB of the constrained limit of the trellis code while the threshold (performance in the limit of infinite decoding complexity and decoding delay) is within 0.25 dB of the limit.

Finally, we have proposed and analyzed a serial concatenation scheme consisting of an outer convolutional code or a single-parity-check turbo-product-code (SPC/TPC) and an inner *recursive* space-time trellis code. Even though this scheme is 1.5-2 dB away from the outage limit, this proposed system is robust and flexible. It offers multiple choices in terms of rate and decoding complexity. It provides good performance even if the channel has very little time variations. The system is scalable in terms of the number of transmit as well as receive antennas. The idea of recursive trellis codes is also useful when implementing automatic repeat request (ARQ) in packet-data systems.

1. Thesis Contributions

The main contributions of this work are summarized below:

- Achievable lower bounds to the constrained modulation outage capacity of MIMO systems in a flat quasi-static fading channel have been computed.

- Fundamental limits on some known space-time systems, viz. spatial multiplexing, space-time block codes and space-time trellis codes have been established.
- Non-optimality of iid signaling has been established in the constrained modulation scenario.
- Optimality of space-time trellis codes for a range of rates and SNRs has been established.
- Non universality of iterative decoding based systems has been established.
- Two variants of a non-iterative transceiver have been proposed. Both of these have performance close to the outage limit.
- It has been shown that almost all space-time trellis codes have recursive realizations.
- Serial concatenation scheme with a recursive realization of a space-time trellis code has been proposed and analyzed. Its flexibility in terms of rate and receiver complexity has been demonstrated.
- The idea of recursive space-time trellis codes has been extended to parallel concatenated codes and to automatic repeat request systems.

2. Future Work

We propose the following future research:

- Capacity achieving schemes for spatial multiplexing, especially in the high rate regime.
- Extending the decision-feedback scheme to higher constellations.

- The gap from constrained capacity of the decision-feedback scheme depends on the exact trellis code being used. A detailed study can be done with an aim to find if and how we can predict this gap.

- Even though the two schemes – spatial multiplexing and the concatenation of a trellis code with an LDPC code – use iterative decoding, the former suffers a smaller loss from its constrained capacity than the latter. A detailed study can be done to explain this phenomenon.

- Studying the effect of recursive encoders on the EXIT charts of space-time trellis codes and the performance analysis of serial concatenation scheme using EXIT charts.

- Constrained modulation capacity analysis for MIMO block fading and delay limited channels.

- The effect of channel knowledge at the transmitter upon the constrained modulation outage capacity.

- Good system designs for very short block lengths.

REFERENCES

[1] Z. Chen, J. Yuan, and B. Vucetic, "Improved space-time trellis coded modulation scheme on slow Rayleigh fading channels," *Electronics Letters*, vol. 37, no. 7, pp. 440–441, Mar. 2001.

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

[3] J. G. Proakis, *Digital Communications*, 3rd ed. New York City, NY: McGraw Hill, 1995.

[4] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Tech.*, vol. 43, no. 2, pp. 359–378, May 1994.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999, available for download at http://mars.bell-labs.com/cm/ms/what/mars/papers/proof/, last accessed Oct. 2004.

[7] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Raykeigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[8] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[9] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. Intl. Conf. Commun.*, vol. 1, New York, NY, USA, 2002, pp. 406–410.

[10] R. W. Heath, Jr. and A. Paulraj, "Characterization of MIMO channels for spatial multiplexing," in *Proc. Intl. Conf. Commun.*, vol. 2, Helsinki, Finland, 2001, pp. 591–595.

[11] A. Paulraj, D. Gore, and R. U. Nabar, "Performance limits in fading MIMO channels," in *Proc. 5th Intl. Symp. Wirless Personal Multimedia Commun.*, vol. 1, Honolulu, Hawaii, USA, 2002, pp. 7–11.

[12] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in mutiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[13] V. Gulati and K. Narayanan, "Concatenated space-time codes for quasi-static fading channels: Constrained capacity and code design," in *Proc. Global Commun. Conf.*, vol. 2, Taipei, Taiwan, 2002, pp. 1202–1206.

[14] J. F. Cheng, "Performance of MIMO space-time coded discrete modulations - part I: Introduction and flat-fading channels," in *Proc. Global Commun. Conf.*, vol. 2, Taipei, Taiwan, 2002, pp. 1177–1181.

[15] ——, "Performance of MIMO space-time coded discrete modulations - part II: Extensions and frequency-selective fading channels," in *Proc. Global Commun.*

Conf., vol. 2, Taipei, Taiwan, 2002, pp. 1182–1186.

[16] E. Baccarelli, "Evaluation of the reliable data rates supported by multiple-antenna coded wireless links for QAM transmissions," *IEEE Journal Selected Areas Commun.*, vol. 19, no. 2, pp. 295–304, Feb. 2001.

[17] J. C. Guey, M. P. Fitz, M. R. Bell, and W. Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.

[18] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis coded modulation," *IEEE Commun. Letters*, vol. 4, no. 1, pp. 20–22, Jan. 2000.

[19] Q. Yan and R. S. Blum, "Optimum space-time convolutional codes," in *Proc. Wireless Commun. and Networking Conf.*, vol. 3, Chicago, IL, USA, 2000, pp. 1351–1355.

[20] E. Biglieri and A. M. Tulino, "Designing space-time codes for large number of receiving antennas," *IEEE Electronics Letters*, vol. 37, pp. 1073–1074, Aug. 2001.

[21] E. Biglieri, G. Taricco, and A. M. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1794–1803, Jul. 2002.

[22] M. Tao and R. S. Cheng, "Improved design criteria and new trellis codes for space-time coded modulation in slow flat fading channels," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 313–315, Jul. 2001.

[23] D. Aktas, H. E. Gamal, and M. P. Fitz, "On the design and maximum-likelihood decoding of space-time trellis codes," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 854–859, Jun. 2003.

[24] A. R. Hammons, Jr. and H. E. Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 524–542, Mar. 2000.

[25] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal Selected Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[26] V. Tarokh, R. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[27] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul 2002.

[28] S. Sandhu and A. Paulraj, "Space-time block codes: A capacity perspective," *IEEE Commun. Letters*, vol. 4, no. 12, pp. 384–386, Dec. 2000.

[29] B. A. Sethuraman and B. S. Rajan, "Optimal STBC over PSK signal sets from cyclotomic field extensions," in *Proc. IEEE Intl. Conf. Commun.*, vol. 3, New York, NY, USA, 2002, pp. 1783–1787.

[30] M. O. Damen, A. Tewfik, and J. C. Belfiore, "A construction of a space-time code based on number theory," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 753–760, Mar. 2002.

[31] M. O. Damen, K. A. Meraim, and J. C. Belfiore, "Diagonal algebraic space-time block codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 628–636, Mar. 2002.

[32] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[33] O. Tirkkonen and A. Hottinen, "Tradeoffs between rate, puncturing and orthogonality in space-time block codes," in *Proc. IEEE Intl. Conf. Commun.*, vol. 4, Helsinki, Finland, 2001, pp. 1117–1121.

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

[35] J. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal Selected Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[36] H. E. Gamal and R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2321–2334, Sep. 2001.

[37] G. Caire and G. Colavolpe, "On low-complexity space-time coding for quasi-static channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 6, pp. 1400–1416, Jun. 2003.

[38] A. Stefanov and T. M. Duman, "Turbo coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations," *IEEE Journal Selected Areas Commun.*, vol. 19, no. 5, pp. 958–968, May 2001.

[39] G. Bauch, "Concatenation of space-time block codes and "turbo"-TCM," in

Proc. IEEE Intl. Conf. Commun., vol. 2, Vancouver, BC, Canada, Jun. 1999, pp. 1202–1206.

[40] G. Bauch and J. Hagenauer, "Analytical evaluation of space-time transmit diversity with FEC-coding," in *Proc. Global Telecomm. Conf.*, San Antonio, TX, USA, Nov. 2001, pp. 435–439.

[41] X. Lin and R. Blum, "Improved space-time codes using serial concatenation," *IEEE Commun. Letters*, vol. 4, no. 7, pp. 221–223, Jul. 2000.

[42] V. Gulati and K. R. Narayanan, "Concatenated codes for fading channels based on recursive space-time trellis codes," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 118–128, Jan. 2003.

[43] X. Lin and R. S. Blum, "Guidelines for serially concatenated space-time code design in Rayleigh fading channels," in *IEEE Third Wkshp. Sig. Proc. Adv. in Wireless Comm.*, Taiwan, China, 2001, pp. 247–250.

[44] L. Goulet and H. Leib, "Serially concatenated space-time codes with iterative decoding and performance limits of block-fading channels," *IEEE Journal Selected Areas Commun.*, vol. 21, no. 5, pp. 765–773, Jun. 2003.

[45] V. Gulati, X. Zhang, K. Narayanan, M. P. Fitz, and A. Matache, "Constrained modulation capacity of MIMO systems in quasi-static fading channels," Nov 2003, under review for publication in IEEE Trans. Wireless Commun.

[46] V. Gulati and K. Narayanan, "Code design for the MIMO flat quasi-static fading channel," 2004, in preparation.

[47] V. Gulati, X. Zhang, K. R. Narayanan, M. P. Fitz, and A. Matache, "Constrained modulation capacity of mimo systems in quasi-static fading channels," Nov 2003,

submitted to IEEE Trans. Wireless Commun.

[48] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[49] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge University Press, 1985.

[50] J. Zheng and S. L. Miller, "Analysis of power and rate allocation of coded OFDM systems over frequency selective fading channels," Feb. 2003, submitted to *IEEE Trans. Inform. Theory*.

[51] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 1, pp. 55–67, Jan. 1982.

[52] S. Siwamogsatham and M. P. Fitz, "Improved high-rate space-time codes from expanded STB-MTCM construction," Feb. 2002, submitted to *IEEE Trans. Inform. Theory*. Available for download at http://www.ee.ucla.edu/~fitz/Papers_folder/stcmbc.pdf, last accessed Oct. 2004.

[53] H. Jafarkhani and N. Seshadri, "Super-orthogonal space-time trellis codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 937–950, Apr. 2003.

[54] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.

[55] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[56] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct 2001.

[57] R. Storn and K. Price, "Differential Evolution – a simple and efficient adaptive scheme for global optimization over continuous spaces," Univ. California, Berkeley, CA, Tech. Rep. TR-95-012, 1995, available online at http://citeseer.nj.nec.com/article/storn95differential.html, last accessed Oct. 2004.

[58] A. Shokrollahi and R. Storn, "Design of efficient erasure codes with differential evolution," 1999, avaliable online at http://shokrollahi.com/amin/pub.html, last accessed Oct., 2004.

[59] J. Ha and S. W. McLaughlin, "Low-density parity-check codes over Gaussian channels with erasures," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1801–1809, Jul 2003.

[60] S. Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Letters*, vol. 5, no. 2, pp. 58–60, Feb 2001.

[61] R. Narayanswami, "Coded modulation with low-density parity-check codes," Master's thesis, Department of Electrical Engg., Texas A&M University, Jun. 2001.

[62] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[63] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb 2001.

[64] G. Colavolpe, G. Ferrari, and R. Raheli, "Extrinsic information in iterative decoding: A unified view," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2088–2094, Dec 2001.

[65] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," 2004, to appear in *IEEE Trans. Inform. Theory*. Available online http://cm.bell-labs.com/cm/ms/who/gkr/Papers/exitIT04.pdf, last accessed Oct. 2004.

[66] S. ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection decoding," *IEEE Trans. Sig. Proc.*, vol. 51, no. 11, pp. 2764–2762, Nov. 2003.

[67] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 657–670, Feb 2001.

[68] K. R. Narayanan and N. Nangare, "A BCJR-DFE based reciver for achieving near capacity performance on inter-symbol interference channels," in *Proceedings 42-nd Allerton Conf. Commun.*, Monticello, IL, USA, 2004.

[69] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and J. G. D. Forney, "MMSE decision feedback equalizers and coding – part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct 1995.

[70] ——, "MMSE decision feedback equalizers and coding – part II: Coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct 1995.

[71] M. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple ac-

cess channel," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, 1997, pp. 1405–1409.

[72] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[73] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output maximum a posteriori (MAP) module to decode parallel and serial concatenated codes," Jet Propulsion Lab, NASA, Tech. Rep. TDA Progress Report 42-127, Nov 15 1996.

[74] J. Campello, D. S. Modha, and S. Rajagopalan, "Designing ldpc codes using bit-filling," in *Proc. IEEE Intl. Conf. Commun.*, vol. 1, Helsinki, Finland, 2001, pp. 55–59.

[75] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul 1999.

[76] K. R. Narayanan, "Turbo decoding of concatenated space-time codes," in *Proc. 37-th Annual Allerton Conf. on Comm., Control and Computing*, Monticello, IL, USA, 1999, pp. 899–900.

[77] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[78] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445,

Mar. 1996.

[79] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log-domain," in *Proc. IEEE Intl. Conf. Comm.*, vol. 2, Seattle, WA, USA, Jun. 1995, pp. 1009–1013.

[80] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis coded modulation," *IEEE Comm. Letters*, vol. 4, no. 1, pp. 20–22, Jan. 2000.

[81] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE Intl. Conf. Comm.*, vol. 3, Geneva, Switzerland, 1993, pp. 1630–1634.

[82] T. Muharemovic, A. Gatherer, W. Ebel, S. H. ad D. Hocevar, and E. Huang, "Space-time codes with bit interleaving," in *Proc. Global Telecomm. Conf.*, vol. 2, San Antonio, TX, USA, Nov. 2001, pp. 1088–1092.

[83] D. Divsalar and F. Pollara, "Turbo codes for PCS applications," in *Proc. IEEE Intl. Conf. Commun.*, vol. 1, Seattle, WA, USA, Jun. 1995, pp. 54–59.

[84] D. M. Rankin and T. A. Gulliver, "Single parity check product codes," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1354–1362, Aug 2001.

[85] D. Divsalar and F. Pollara, "Multiple turbo codes," in *Proc. IEEE MILCOM*, San Diego, CA, USA, 1995, pp. 279–285.

[86] N. Kahale and R. Urbanke, "On the minimum distance of parallel and serial concatenated codes," in *Proc. IEEE Intl. Symp. Inform. Theory*, Cambridge, MA, USA, 1998, p. 31.

[87] G. Bauch, "Concatenation of space-time block codes and turbo-TCM," in *Proc. IEEE Intl. Conf. Comm.*, vol. 2, Vancouver, BC, Canada, Jun. 1999, pp. 1202–1206.

[88] D. Arnold and H. A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. IEEE Intl. Conf. Commun.*, vol. 9, Helsinki, Finland, Jun. 2001, pp. 2692–2695.

[89] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Intl. Symp. Inform. Theory*, Washington, DC, USA, 2001, p. 283.

[90] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. Global Telecomm. Conf.*, vol. 5, San Antonio, TX, USA, 2001, pp. 2992–2996.

[91] L. R. Bahl, J. Cocke, F. Jelenik, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, Mar 1974.

APPENDIX A

DIAGONAL CORRELATION SUFFICES TO OPTIMIZE OUTAGE FOR GAUSSIAN SIGNALING

Let the $N_t \times N_f$ transmit matrix $\mathbf{X}$ be re-written as a vector:

$$\vec{x} = [x_{11}\ x_{12}\ \ldots\ x_{1N_t}\ \ldots\ x_{N_fN_t}]^t = [\bar{x}_1^t\ \bar{x}_2^t\ \ldots\ \bar{x}_{N_f}^t]^t.$$

The correlation matrix of $\vec{x}$ is given by:

$$\mathbf{Q} = \mathcal{E}[\vec{x}\vec{x}^H] = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} & \ldots & \mathbf{Q}_{1N_f} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \ldots & \mathbf{Q}_{2N_f} \\ \vdots & & & \vdots \\ \mathbf{Q}_{N_f1} & \mathbf{Q}_{N_f2} & \ldots & \mathbf{Q}_{N_fN_f} \end{bmatrix},$$

where $\mathbf{Q}_{ij} = \mathcal{E}[\bar{x}_i\bar{x}_j^H]$ is the correlation matrix between the vectors $\bar{x}_i$ and $\bar{x}_j$ transmit at times i and j. Clearly, $\mathbf{Q}_{ij} = \mathbf{Q}_{ji}^H$.

The baseband receive signal may now be written as $\vec{y} = \mathbf{A}\vec{x} + \vec{n}$, where the matrix $\mathbf{A} = I_{N_f} \otimes \mathbf{H}$ is the Kronecker product of the channel matrix $\mathbf{H}$ with the identity matrix. With this formulation and assuming Gaussian distributions for $\vec{x}$ and $\vec{y}$, the instantaneous mutual information between $\vec{x}$ and $\vec{y}$ may be written as $\Psi(\mathbf{Q}, \mathbf{A}) = \log\det(I_{N_rN_f} + \mathbf{AQA}^H)$ [5,6]. The quantity $\mathbf{AQA}^H$ has a block structure:

$$\mathbf{AQA}^H = \begin{bmatrix} \mathbf{HQ}_{11}\mathbf{H}^H & \mathbf{HQ}_{12}\mathbf{H}^H & \ldots & \mathbf{HQ}_{1N_f}\mathbf{H}^H \\ (\mathbf{HQ}_{12}\mathbf{H})^H & \mathbf{HQ}_{22}\mathbf{H}^H & \ldots & \mathbf{HQ}_{2N_f}\mathbf{H}^H \\ \vdots & & & \vdots \\ (\mathbf{HQ}_{1N_f}\mathbf{H})^H & (\mathbf{HQ}_{2N_f}\mathbf{H})^H & \ldots & \mathbf{HQ}_{N_fN_f}\mathbf{H}^H \end{bmatrix},$$

since $\mathbf{Q}_{ij} = \mathbf{Q}_{ji}^H$. This form is exactly the one required by Fischer's inequality for semi-positive definite matrices [49], and so we have:

$$\begin{aligned}\frac{1}{N_f}\Psi(\mathbf{Q},\mathbf{A}) &= \frac{1}{N_f}\log\det(I_{N_r N_f} + \mathbf{A}\mathbf{Q}\mathbf{A}^H) \\ &\leq \frac{1}{N_f}\sum_{i=1}^{N_f}\log\det(I_{N_r} + \mathbf{H}\mathbf{Q}_{ii}\mathbf{H}^H) \\ &= \frac{1}{N_f}\sum_{i=1}^{N_f}\Psi(\mathbf{Q}_{ii},\mathbf{H})\end{aligned}$$

Note that the upper bound corresponds to the mutual information when the signal vectors are chosen to be independent in time. Hence, it is optimal to choose the components of $\mathbf{X}$ to be vectors which are independent in time. Further, within each time instant, we can choose the symbols to be independent of each other (no spatial correlation). This can be shown as follows. The mutual information between the transmit vector $\vec{x}$ and the received vector $\vec{y}$ at a given time instant is [5, 6]: $\Psi(\mathbf{Q},\mathbf{H}) = \log\det(I_{N_r} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)$, where $\mathbf{Q} = \mathcal{E}[\vec{x}\vec{y}^H]$. In this case, the distribution of $\Psi(\mathbf{Q},\mathbf{H})$remains the same if $\mathbf{Q}$ is replaced by a diagonal matrix $\mathbf{V} = \mathbf{U}\ \mathbf{Q}\ \mathbf{U}^H$, where $\mathbf{U}$ is unitary transformation. Since $\mathbf{Q}$ is the correlation matrix, it is always diagonalizable [6]. Hence, a diagonal correlation matrix $\mathbf{Q}$ (independent elements in the vector $\bar{x}$) suffices to optimize the outage probability.

APPENDIX B

INFORMATION RATE COMPUTATION FOR A SPACE-TIME TRELLIS CODE

The information rate computation is based on the ideas in [88–90] and is described in [13]. Recall that the space-time trellis codes output N_t symbols per branch of the trellis and the corresponding received signal, at time k, is given by:

$$\vec{y}[k] = \mathbf{H}\vec{x}[k] + \vec{n}[k] = \vec{a}[k] + \vec{n}[k]$$

When the channel is known, the noiseless received signal is $\vec{a} = [a_1,\ a_2,\ \ldots\ a_{N_r}]^t = \mathbf{H}\vec{x}$ and the received signal at the i-th receive antenna is $y_i = a_i + n_i$, where n_i is the i-th component of the complex Gaussian noise vector $\vec{n}$. We now wish to compute the mutual information between the time series $\vec{x}$ and $\vec{y}$ given $\mathbf{H}$, or equivalently, between $\vec{A}$ and $\vec{y}$: $I^{sttc}_{N_t,N_r}(\mathbf{h}) = I(\vec{X};\vec{Y}|\mathbf{h}) = h(\vec{Y}|\mathbf{h}) - h(\vec{N}|\mathbf{h})$, where the superscript $sttc$ refers to a space-time trellis code. The quantity $h(\vec{N}|\mathbf{h}) = \frac{N_r}{2}\log(2\pi e)$ since the N_r entries of $\vec{n}$ have unit variance. In order to estimate $h(\vec{Y}|\mathbf{h})$, we extend the ideas from [88–90]. For any given block length N_f and any given channel output $\vec{\mathbf{y}} \triangleq [\vec{y}[1], \vec{y}[2], \ldots, \vec{y}[N_f]]$, the probability of $\Pr(\vec{\mathbf{y}})$ can be computed using the forward recursion of the BCJR algorithm [91] which operates on the trellis of the code. To emphasize this similarity we define:

$$\alpha_k(s) = \Pr\left(\sigma[k] = s|\vec{y}(k-1)\right)$$

and use the shorthand notation $\{\sigma[k] = s\} = \{s[k]\}$. The forward recursion to compute $\alpha_k(s)$ is:

$$\alpha_k(s) = \sum_{s' \in \Omega_\sigma} \alpha_{k-1}(s') f\left(\vec{y}[k-1]|s[k], s'[k-1]\right)$$

The summation can be evaluated by noting that:

$$p\left(s[k], \vec{y}[k-1] | s'[k-1]\right) = N_\alpha p\left(\vec{y}[k] | s'[k-1], s[k]\right),$$

where N_α is a normalization factor. The trellis representation simplifies this to:

$$Pr\left(s[k] | s'[k-1], \vec{y}(k-1)\right) = N_\alpha p\left(\vec{y}[k] | s[k], x(k-1)\right). \tag{B.1}$$

Note that the RHS of (B.1) is a Gaussian density. Finally observing that:

$$\begin{aligned}
p\left(\vec{y}(N_f)\right) &= \prod_{k=0}^{N_f} p\left(\vec{y}[k] | \vec{y}(k-1)\right) \\
\log p\left(\vec{y}(N_f)\right) &= \sum_{k=0}^{N_f} \log p\left(\vec{y}[k] | \vec{y}(k-1)\right) \\
p\left(\vec{y}[k] | \vec{y}(k-1)\right) &= \sum_{s \in \Omega_\sigma} \alpha_k(s) p\left(\vec{y}[k] | s[k]\right) \\
p\left(y(k) | s[k]\right) &= \sum_{x(k-1)} p\left(y(k) | s(k), x(k-1)\right) \\
&= \sum_{x(k-1)} p\left(y(k) | \vec{a}[k]\right),
\end{aligned}$$

the recursion can be completed. Hence the calculation of $\log p\left(\vec{\mathbf{y}}\right)$ has exactly the same structure as implementing the BCJR algorithm.

The BCJR algorithm operating on long blocks of data can approximate the required ensemble average with a time averaging (ergodicity). An estimate of $h(\vec{\mathbf{y}}) = -\mathcal{E}[\log\left(\Pr(\vec{\mathbf{y}})\right)]$ is thus obtained by simulating the channel N times, each time starting with a stationary state distribution, simulating N_f inputs $\vec{\mathbf{x}}_i$ and corresponding outputs $\vec{\mathbf{y}}_i$ and computing $\Pr(\vec{\mathbf{y}}_i)$ using the BCJR algorithm. Now, the quantity $-\frac{1}{N}\sum_i \Pr(\vec{\mathbf{y}}_i) \log\left(\Pr(\vec{\mathbf{y}}_i)\right)$ is an estimate of $h(\vec{\mathbf{Y}})$ that converges with probability 1 to the true value for $N_f \to \infty$. Since the Markov process is stationary and ergodic, $-\frac{1}{N_f} \log \Pr(\vec{\mathbf{y}})$ also converges to $h(\vec{\mathbf{Y}})$ and hence a single long simulation of $\vec{\mathbf{y}}$ and the

corresponding single forward BCJR recursion also gives a good estimate of $h(\vec{\mathbf{y}}|\mathbf{h})$.

APPENDIX C

INFORMATION RATE COMPUTATION FOR PUNCTURED SPACE-TIME TRELLIS CODES

In this appendix we show that the Shannon-McMillan-Briemann theorem can also be applied to the information rate computation of punctured space-time trellis codes. However, with puncturing the trellis must be modified to make the Markov process stationary. This is explained in more details below with the help of an example. It should be noted that by puncturing of space-time codes we mean the omission the transmission of signals (from all transmit antennas) at a given time instant. There are other ways of puncturing but we do not consider them here.

Consider the simple example of the 4-state AT&T code [2] for $N_t = 2$ transmit antennas and QPSK modulation. The trellis structure of this code is shown in Fig. 5. At each time instant one QPSK symbol is input to the encoder and a vector of two QPSK symbols is transmitted. The overall rate of this code is 2 b/s/Hz. Now suppose that we want to achieve 3 b/s/Hz with the same code by omitting the transmission of output vector once every three time instants. In the trellis representation, now we consider the input of the encoder to be of length 3 and the output to be of length 4 (2 transmit antennas, 2 time instants). The resulting (stationary) trellis also has the same number of states (four in this case) as the original code but the number of paths out of each state increases to 64 (instead of 4 in the unpunctured case). Similarly, if we puncture the symbols every other time instant, the number of paths per state is 16. Note, again, the parallel transitions do not affect the algorithm to compute information rates. The resulting trellis is also shown in Fig. 5. In general, if the puncturing rate is once every p-time instants, the number of paths in the modified

trellis would be M^p, where M is the cardinality of the signal set being used. It should be noted, though, that all this complexity is in the computation of information rates only. The decoding is still only as complex as that for the unpunctured trellis.

Once we have the modified trellis, we can use the method of Appendix B to compute the i.i.d. information rates for punctured space-time trellis codes. While conceptually this is pretty straightforward, the computations become numerically very intensive, even for $p \geq 3$.

VITA

Vivek Gulati obtained a B. Tech.(Hons.) degree in electronics and electrical communications engineering from Indian Institute of Technology, Kharagpur, India in May 1999.

He may be reached at H. No. 5, Professor's Colony, Yamuna Nagar, Haryana, India-135001. His email address is gulati@ieee.org.