Space–Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction

Vahid Tarokh, *Member, IEEE*, Nambi Seshadri, *Senior Member, IEEE*, and A. R. Calderbank, *Fellow, IEEE*

***Abstract*— We consider the design of channel codes for improving the data rate and/or the reliability of communications over fading channels using multiple transmit antennas. Data is encoded by a channel code and the encoded data is split into n streams that are simultaneously transmitted using n transmit antennas. The received signal at each receive antenna is a linear superposition of the n transmitted signals perturbed by noise. We derive performance criteria for designing such codes under the assumption that the fading is slow and frequency nonselective. Performance is shown to be determined by matrices constructed from pairs of distinct code sequences. The minimum rank among these matrices quantifies the *diversity gain*, while the minimum determinant of these matrices quantifies the *coding gain*. The results are then extended to fast fading channels. The design criteria are used to design trellis codes for high data rate wireless communication. The encoding/decoding complexity of these codes is comparable to trellis codes employed in practice over Gaussian channels. The codes constructed here provide the best tradeoff between data rate, diversity advantage, and trellis complexity. Simulation results are provided for 4 and 8 PSK signal sets with data rates of 2 and 3 bits/symbol, demonstrating excellent performance that is within 2–3 dB of the outage capacity for these channels using only 64 state encoders.**

***Index Terms*— Array processing, diversity, multiple transmit antennas, space–time codes, wireless communications.**

I. Introduction

A. Motivation

CURRENT cellular standards support circuit data and fax services at 9.6 kb/s and a packet data mode is being standardized. Recently, there has been growing interest in providing a broad range of services including wire-line voice quality and wireless data rates of about 64–128 kb/s (ISDN) using the cellular (850-MHz) and PCS (1.9-GHz) spectra [2]. Rapid growth in mobile computing is inspiring many proposals for even higher speed data services in the range of 144 kb/s (for microcellular wide-area high-mobility applications) and up to 2 Mb/s (for indoor applications) [1].

The majority of the providers of PCS services have further decided to deploy standards that have been developed at cellular frequencies such as CDMA (IS-95), TDMA (IS-54/IS-136), and GSM (DCS-1900). This has led to considerable effort in developing techniques to provide the aforementioned new services while maintaining some measure of backward compatibility. Needless to say, the design of these techniques is a challenging task.

Band-limited wireless channels are narrow pipes that do not accommodate rapid flow of data. Deploying multiple transmit and receive antennas broadens this data pipe. Information theory [14], [35] provides measures of capacity, and the standard approach to increasing data flow is linear processing at the receiver [15], [44]. We will show that there is a substantial benefit in merging signal processing at the receiver with coding technique appropriate to multiple transmit antennas. In particular, the focus of this work is to propose a solution to the problem of designing a physical layer (channel coding, modulation, diversity) that operate at bandwidth efficiencies that are twice to four times as high as those of today's systems using multiple transmit antennas.

B. Diversity

Unlike the Gaussian channel, the wireless channel suffers from attenuation due to destructive addition of multipaths in the propagation media and due to interference from other users. Severe attenuation makes it impossible for the receiver to determine the transmitted signal unless some less-attenuated replica of the transmitted signal is provided to the receiver. This resource is called *diversity* and it is the single most important contributor to reliable wireless communications. Examples of diversity techniques are (but are not restricted to)

- *Temporal Diversity:* Channel coding in conjunction with time interleaving is used. Thus replicas of the transmitted signal are provided to the receiver in the form of redundancy in temporal domain.
- *Frequency Diversity:* The fact that waves transmitted on different frequencies induce different multipath structure in the propagation media is exploited. Thus replicas of the transmitted signal are provided to the receiver in the form of redundancy in the frequency domain.
- *Antenna Diversity:* Spatially separated or differently polarized antennas are used. The replicas of transmitted signal are provided to the receiver in the form of redundancy in spatial domain. This can be provided with no penalty in bandwidth efficiency.

When possible, cellular systems should be designed to encompass all forms of diversity to ensure adequate performance

Manuscript received December 15, 1996; revised August 18, 1997. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, June 29–July 4, 1997.

The authors are with the AT&T Labs–Research, Florham Park, NJ 07932 USA.

Publisher Item Identifier S 0018-9448(98)00933-X.

[26]. For instance, cellular systems typically use channel coding in combination with time interleaving to obtain some form of temporal diversity [28]. In TDMA systems, frequency diversity is obtained using a nonlinear equalizer [4] when multipath delays are a significant fraction of symbol interval. In DS-CDMA, RAKE receivers are used to obtain frequency diversity. Antenna diversity is typically used in the up-link (mobile-to-base) direction to provide the link margin and cochannel interference suppression [40]. This is necessary to compensate for the low power transmission from mobiles.

Not all forms of diversity can be available at all times. For example, in slow fading channels, temporal diversity is not an option for delay-sensitive applications. When the delay spread is small, frequency (multipath) diversity is not an option. In macrocellular and microcellular environments, respectively, this implies that the data rates should be at least several hundred thousand symbols per second and several million symbols per second, respectively. While antenna diversity at a base-station is used for reception today, antenna diversity at a mobile handset is more difficult to implement because of electromagnetic interaction of antenna elements on small platforms and the expense of multiple down-conversion RF paths. Furthermore, the channels corresponding to different antennas are correlated, with the correlation factor determined by the distance as well as the coupling between the antennas. Typically, the second antenna is inside the mobile handset, resulting in signal attenuation at the second antenna. This can cause some loss in diversity benefit. All these factors motivate the use of multiple antennas at the base-station for transmission.

In this paper, we consider the joint design of coding, modulation, transmit and receive diversity to provide high performance. We can view our work as combined coding and modulation for multi-input (multiple transmit antennas) multi-output (multiple receive antennas) fading channels. There is now a large body of work on coding and modulation for single-input/multi-output channels [5], [10], [11], [29], [30], [38], and [39], and a comparable literature on receive diversity, array processing, and beamforming. In light of these research activities, receive diversity is very well understood. By contrast, transmit diversity is less well understood. We begin by reviewing prior work on transmit diversity.

C. Historical Perspective on Transmit Diversity

Systems employing transmit fall into three general categories. These are

- schemes using feedback,
- those with feedforward or training information but no feedback, and
- blind schemes.

The first category uses implicit or explicit feedback of information from the receiver to the transmitter to configure the transmitter. For instance, in time-division duplex systems [16], the same antenna weights are used for reception and transmission, so feedback is implicit in the appeal to channel symmetry. These weights are chosen during reception to maximize the signal-to-noise ratio (SNR), and during transmission to weight the amplitudes of the transmitted signals. Explicit feedback includes switched diversity with feedback [41] as well as techniques that use spatiotemporal-frequency water pouring [27] based on the feedback of the channel response. However, in practice, vehicle movements or interference causes a mismatch between the state of the channel perceived by the transmitter and that perceived by receiver.

Transmit diversity schemes mentioned in the second category use linear processing at the transmitter to spread the information across the antennas. At the receiver, information is obtained by either linear processing or maximum-likelihood decoding techniques. Feedforward information is required to estimate the channel from the transmitter to the receiver. These estimates are used to compensate for the channel response at the receiver. The first scheme of this type was proposed by Wittneben [43] and it includes the delay diversity scheme of Seshadri and Winters [32] as a special case. The linear processing approach was also studied in [15] and [44]. It has been shown in [42] that delay diversity schemes are indeed optimal in providing diversity in the sense that the diversity advantage experienced by an optimal receiver is equal to the number of transmit antennas. We can view the linear filter as a channel code that takes binary data and creates real-valued output. It is shown that there is significant gain to be realized by viewing this problem from a coding perspective rather than purely from the signal processing point of view.

The third category does not require feedback or feedforward information. Instead, it uses multiple transmit antennas combined with channel coding to provide diversity. An example of this approach is to combine phase sweeping transmitter diversity of [18] with channel coding [19]. Here a small frequency offset is introduced on one of the antennas to create fast fading. An appropriately designed channel code/interleaver pair is used to provide diversity benefit. Another scheme is to encode information by a channel code and transmit the code symbols using different antennas in an orthogonal manner. This can be done either by frequency multiplexing [9], time multiplexing [32], or by using orthogonal spreading sequences for different antennas [37]. A disadvantage of these schemes over the previous two categories is the loss in bandwidth efficiency due to the use of the channel code. Using appropriate coding, it is possible to relax the orthogonality requirement needed in these schemes and obtain the diversity as well as coding advantage offer without sacrificing bandwidth. This is possible when the whole system is viewed as a multiple-input/multiple-output system and suitable codes are used.

Information-theoretic aspects of transmit diversity were addressed in [14], [25], and [35]. We believe that Telatar [35] was the first to obtain expressions for capacity and error exponents for multiple transmit antenna system in the presence of Gaussian noise. Here, capacity is derived under the assumption that fading is independent from one channel use to the other. At about the same time, Foschini and Gans [14] derived the outage capacity under the assumption that fading is quasistatic; i.e., constant over a long period of time, and then changes in an independent manner. A particular layered space–time architecture was shown to have the potential to achieve a substantial fraction of capacity. A major conclusion

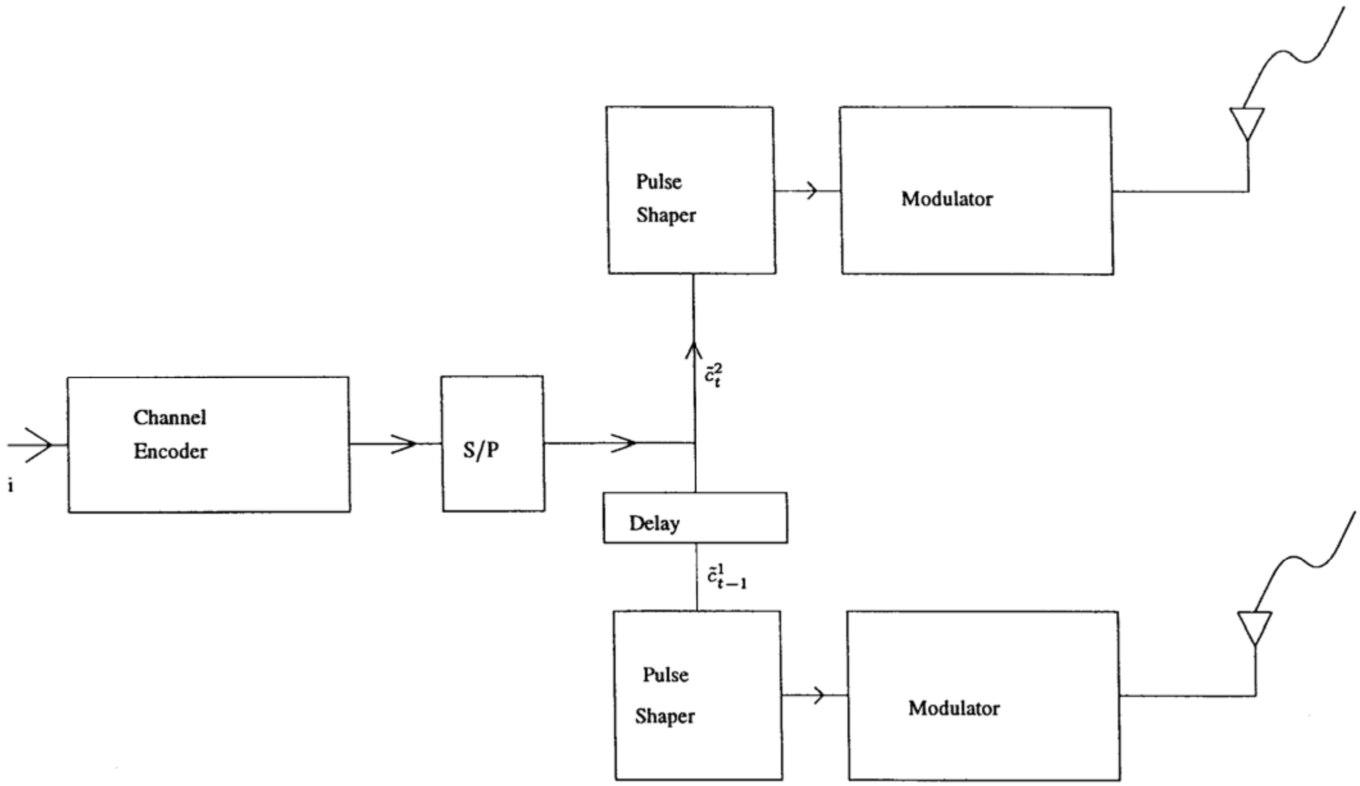

Fig. 1. The block diagram of a delay diversity transmitter.

of these works is that the capacity of a multi-antenna systems far exceeds that of a single-antenna system. In particular, the capacity grows at least linearly with the number of transmit antennas as long as the number of receive antennas is greater than or equal to the number of transmit antennas. A comprehensive information-theoretic treatment for many of the transmit diversity schemes that have been studied before is presented by Narula, Trott, and Wornell [25].

D. Space–Time Codes

We consider the delay diversity scheme as proposed by Wittneben [44]. This scheme transmits the same information from both antennas simultaneously but with a delay of one symbol interval. We can view this as a special case of the arrangement in Fig. 1, where the information is encoded by a channel code (here the channel code is a repetition code of length 2). The output of the repetition code is then split into two parallel data streams which are transmitted with a symbol delay between them. Note that there is no bandwidth penalty due to the use of the repetition code, since two output-channel symbols are transmitted at each interval.

It was shown in [32], via simulations, that the effect of this technique is to change a narrowband purely frequency-nonselective fading channel into a frequency-selective fading channel. Simulation results further demonstrated that a maximum-likelihood sequence estimator at the receiver is capable of providing dual branch diversity.

When viewed in this framework, it is natural to ask if it is possible to choose a channel code that is better than the $R = 1/2$ repetition code in order to provide improved performance while maintaining the same transmission rate?

We answer the above question affirmatively and propose a new class of codes for this application referred to as the *Space–Time Codes*. The restriction imposed by the delay element in the transmitter is first removed. Then performance criteria are established for code design assuming that the fading from each transmit antenna to each receive antenna is Rayleigh or Rician. It is shown that the delay diversity scheme of Seshadri and Winters [32] is a specific case of space–time coding.

In Section II, we derive performance criteria for designing codes. For quasistatic flat Rayleigh or Rician channels, performance is shown to be determined by the diversity advantage quantified by the rank of certain matrices and by the coding advantage that is quantified by the determinants of these matrices. These matrices are constructed from pairs of distinct channel codewords. For rapidly changing flat Rayleigh channels, performance is shown to be determined by the diversity advantage quantified by the generalized Hamming distance of certain sequences and by the coding advantage that is quantified by the generalized product distance of these sequences. These sequences are constructed from pairs of distinct codewords. In Section III, this performance criterion is used to design trellis codes for high data rate wireless communication. We design coded modulation schemes based on 4-PSK, 8-PSK, and 16-QAM that perform extremely well and can operate within 2–3 dB of the outage capacity derived by Foschini and Gans [14]. For a given data rate, we compute the minimal constraint length, the trellis complexity required to achieve a certain diversity advantage, and we establish an upper bound

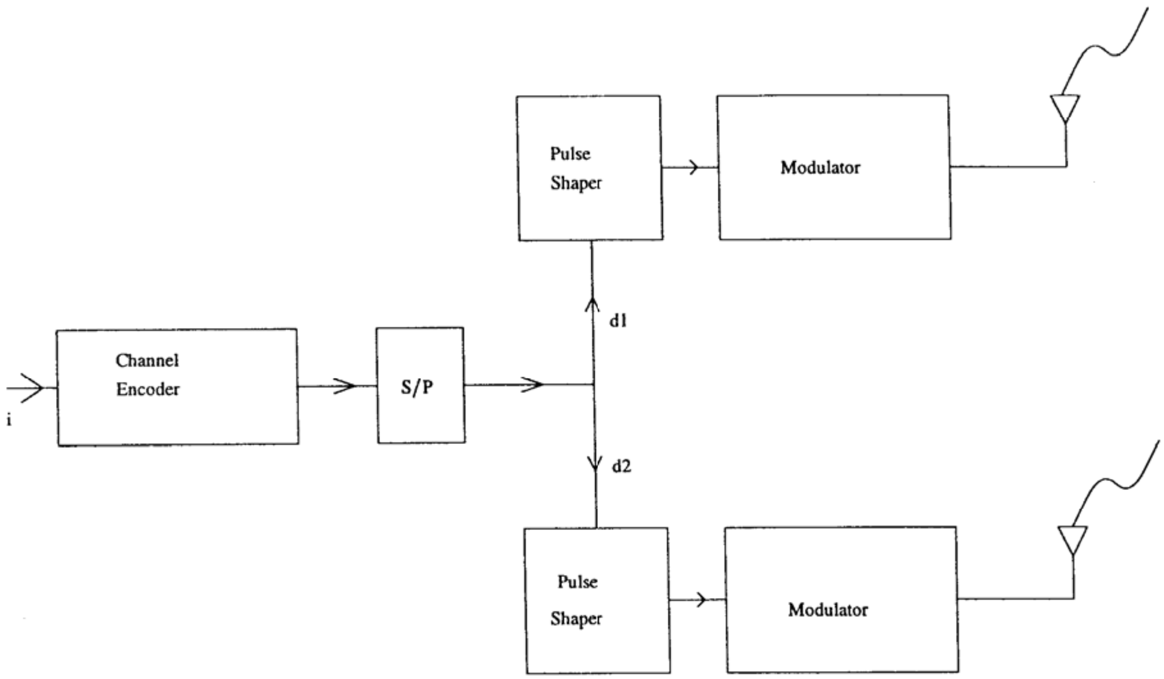

Fig. 2. The block diagram of the transmitter.

on the data rate as a function of the constellation size and diversity advantage. For a given diversity, we provide explicit constructions of trellis codes that achieve the minimum trellis complexity as well as the maximum data rate. Then, we revisit delay diversity and show that some of the codes constructed before have equivalent delay diversity representations. This section also includes multilevel constructions which provide an efficient way to construct and decode codes when the number of antennas is large (4–8). It is further shown that it is not possible for block-coded modulation schemes to outperform trellis codes constructed here at a given diversity advantage and data rate. Simulation results for many of the codes that we have constructed and comparisons to outage capacity for these channels are also presented. We then consider design of space–time codes that guarantee a diversity advantage of r_1 when there is no mobility and a diversity advantage of $r_2 \geq r_1$ when the channel is fast-fading. In constructing these codes, we combine the design criteria for rapidly changing flat Rayleigh channels with that of quasistatic flat Rayleigh channels to arrive at a hybrid criteria. We refer to these codes as *smart greedy codes* which also stands for *low-rate multidimensional space–time codes for both slow and rapid fading channels*. We provide simulation results indicating that these codes are ideal for increasing the frequency reuse factor under a variety of mobility conditions. Some conclusions are made in Section IV.

II. Performance Criteria

A. The System Model

We consider a mobile communication system where the base-station is equipped with n antennas and the mobile is equipped with m antennas. Data is encoded by the channel encoder, the encoded data goes through a serial-to-parallel converter, and is divided into n streams of data. Each stream of data is used as the input to a pulse shaper. The output of each shaper is then modulated. At each time slot t, the output of modulator i is a signal c_t^i that is transmitted using transmit antenna (Tx antenna) i for $1 \leq i \leq n$. We emphasize that the n signals are transmitted simultaneously each from a different transmit antenna and that all these signals have the same transmission period T. The signal at each receive antenna is a noisy superposition of the n transmitted signals corrupted by Rayleigh or Rician fading (see Fig. 2). We assume that the elements of the signal constellation are contracted by a factor of $\sqrt{E_s}$ chosen so that the average energy of the constellation is 1.

At the receiver, the demodulator computes a decision statistic based on the received signals arriving at each receive antenna $1 \leq j \leq m$. The signal d_t^j received by antenna j at time t is given by

$$d_t^j = \sum_{i=1}^{n} \alpha_{i,j} c_t^i \sqrt{E_s} + \eta_t^j \tag{1}$$

where the noise η_t^j at time t is modeled as independent samples of a zero-mean complex Gaussian random variable with variance $N_0/2$ per dimension. The coefficient $\alpha_{i,j}$ is the path gain from transmit antenna i to receive antenna j. It is assumed that these path gains are constant during a frame and vary from one frame to another (quasistatic flat fading).

B. The Case of Independent Fade Coefficients

In this subsection, we assume that the coefficients $\alpha_{i,j}$ are first modeled as independent samples of complex Gaussian random variables with possibly nonzero complex mean $E\alpha_{i,j}$ and variance 0.5 per dimension. This is equivalent to the assumption that signals transmitted from different antennas undergo independent fades.

We shall derive a design criterion for constructing codes under this transmission scenario. We begin by establishing the notation and by reviewing the results from linear algebra that we will employ. This notation will also be used in the sequel to this paper [34]. Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$ and $\boldsymbol{y} = (y_1, y_2, \cdots, y_k)$ be complex vectors in the k-dimensional complex space $\mathbb{C}^k$. The inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{k} x_i \overline{y}_i$$

where $\overline{y}_i$ denotes the complex conjugate of y_i. For any matrix A, let A^* denote the Hermitian (transpose conjugate) of A. Recall from linear algebra that an $n \times n$ matrix A is *Hermitian* if $A = A^*$. The matrix A is *nonnegative definite* if $\boldsymbol{x}A\boldsymbol{x}^* \geq 0$ for any $1 \times n$ complex vector $\boldsymbol{x}$. An $n \times n$ matrix V is *unitary* if $VV^* = I$ where I is the identity matrix. An $n \times l$ matrix B is a *square root* of an $n \times n$ matrix A if $BB^* = A$. We shall make use of the following results from linear algebra [20].

- An eigenvector $\boldsymbol{v}$ of an $n \times n$ matrix A corresponding to eigenvalue λ is a $1 \times n$ vector of unit length such that $\boldsymbol{v}A = \lambda\boldsymbol{v}$ for some complex number λ. The vector space spanned by the eigenvectors of A corresponding to the eigenvalue zero has dimension $n - r$, where r is the rank of A.
- Any matrix A with a square root B is nonnegative definite.
- For any nonnegative-definite Hermitian matrix A, there exists a lower triangular square matrix B such that $BB^* = A$.
- Given a Hermitian matrix A, the eigenvectors of A span $\mathbb{C}^n$, the complex space of n dimensions and it is easy to construct an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors A. Furthermore, there exists a unitary matrix V and a real diagonal matrix D such that $VAV^* = D$. The rows of V are an orthonormal basis of $\mathbb{C}^n$ given by eigenvectors of A. The diagonal elements of D are the eigenvalues λ_i, $i = 1, 2, \cdots, n$ of A counting multiplicities.
- The eigenvalues of a Hermitian matrix are real.
- The eigenvalues of a nonnegative-definite Hermitian matrix are nonnegative.

Let us assume that each element of the signal constellation is contracted by a scale factor $\sqrt{E_s}$ chosen so that the average energy of the constellation elements is 1. Thus our design criterion is not constellation-dependent and applies equally well to 4-PSK, 8-PSK, and 16-QAM.

We consider the probability that a maximum-likelihood receiver decides erroneously in favor of a signal

$$\boldsymbol{e} = e_1^1 e_1^2 \cdots e_1^n e_2^1 e_2^2 \cdots e_2^n \cdots e_l^1 e_l^2 \cdots e_l^n$$

assuming that

$$\boldsymbol{c} = c_1^1 c_1^2 \cdots c_1^n c_2^1 c_2^2 \cdots c_2^n \cdots c_l^1 c_l^2 \cdots c_l^n$$

was transmitted.

Assuming ideal channel state information (CSI), the probability of transmitting $\boldsymbol{c}$ and deciding in favor of $\boldsymbol{e}$ at the decoder is well approximated by

$$P(\boldsymbol{c} \rightarrow \boldsymbol{e} | \alpha_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m) \leq \exp(-d^2(\boldsymbol{c}, \boldsymbol{e}) E_s / 4N_0) \quad (2)$$

where $N_0/2$ is the noise variance per dimension and

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{j=1}^{m} \sum_{t=1}^{l} \left| \sum_{i=1}^{n} \alpha_{i,j} (c_t^i - e_t^i) \right|^2. \quad (3)$$

This is just the standard approximation to the Gaussian tail function.

Setting $\Omega_j = (\alpha_{1,j}, \cdots, \alpha_{n,j})$, we rewrite (3) as

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{i'=1}^{n} \alpha_{i,j} \overline{\alpha_{i',j}} \sum_{t=1}^{l} (c_t^i - e_t^i) \overline{(c_t^{i'} - e_t^{i'})}.$$

After simple manipulations, we observe that

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{j=1}^{m} \Omega_j A \Omega_j^* \quad (4)$$

where $A_{pq} = \boldsymbol{x_p} \cdot \boldsymbol{x_q}$ and $\boldsymbol{x_p} = (c_1^p - e_1^p, c_2^p - e_2^p, \cdots, c_l^p - e_l^p)$ for $1 \leq p, q \leq n$. Thus

$$P(\boldsymbol{c} \rightarrow \boldsymbol{e} | \alpha_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m) \leq \prod_{j=1}^{m} \exp(-\Omega_j A(\boldsymbol{c}, \boldsymbol{e}) \Omega_j^* \, E_s / 4N_0) \quad (5)$$

where

$$A_{pq} = \sum_{t=1}^{l} (c_t^p - e_t^p) \overline{(c_t^q - e_t^q)}.$$

Since $A(\boldsymbol{c}, \boldsymbol{e})$ is Hermitian, there exists a unitary matrix V and a real diagonal matrix D such that $VA(\boldsymbol{c}, \boldsymbol{e})V^* = D$. The rows $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ of V are a complete orthonormal basis of $\mathbb{C}^n$ given by eigenvectors of A. Furthermore, the diagonal elements of D are the eigenvalues λ_i, $i = 1, 2, \cdots, n$ of A counting multiplicities. By construction, the matrix

$$B(\boldsymbol{c}, \boldsymbol{e}) = \begin{pmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \cdots & \cdots & e_l^1 - c_l^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \cdots & \cdots & e_l^2 - c_l^2 \\ e_1^3 - c_1^3 & e_2^3 - c_2^3 & \ddots & \vdots & e_l^3 - c_l^3 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ e_1^n - c_1^n & e_2^n - c_2^n & \cdots & \cdots & e_l^n - c_l^n \end{pmatrix} \quad (6)$$

is clearly a square root of $A(\boldsymbol{c}, \boldsymbol{e})$. Thus the eigenvalues of $A(\boldsymbol{c}, \boldsymbol{e})$ are nonnegative real numbers.

Next, we express $d^2(\boldsymbol{c}, \boldsymbol{e})$ in terms of the eigenvalues of the matrix $A(\boldsymbol{c}, \boldsymbol{e})$.

Let $(\beta_{1,j}, \cdots, \beta_{n,j}) = \Omega_j V^*$, then

$$\Omega_j A(\boldsymbol{c}, \boldsymbol{e}) \Omega_j^* = \sum_{i=1}^{n} \lambda_i |\beta_{i,j}|^2. \quad (7)$$

Next, recall that $\alpha_{i,j}$ are samples of a complex Gaussian random variable with mean $E\alpha_{i,j}$. Let

$$\boldsymbol{K}^j = (E\alpha_{1,j}, E\alpha_{2,j}, \cdots, E\alpha_{n,j}).$$

Since V is unitary, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is an orthonormal basis of $\mathbb{C}^n$ and $\beta_{i,j}$ are independent complex Gaussian random variables with variance 0.5 per dimension and mean $\boldsymbol{K}^j \cdot \boldsymbol{v}_i$. Let $K_{i,j} = |E\beta_{i,j}|^2 = |\boldsymbol{K}^j \cdot \boldsymbol{v}_i|^2$. Thus $|\beta_{i,j}|$ are independent Rician distributions with pdf

$$p(|\beta_{i,j}|) = 2|\beta_{i,j}| \exp(-|\beta_{i,j}|^2 - K_{i,j}) I_0(2|\beta_{i,j}|\sqrt{K_{i,j}})$$

for $|\beta_{i,j}| \geq 0$, where $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind.

Thus to compute an upper bound on the average probability of error, we simply average

$$\prod_{j=1}^{m} \exp(-(E_s/4N_0) \sum_{i=1}^{n} \lambda_i |\beta_{i,j}|^2)$$

with respect to independent Rician distributions of $|\beta_{i,j}|$ to arrive at

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \prod_{j=1}^{m} \left(\prod_{i=1}^{n} \frac{1}{1 + \frac{E_s}{4N_0}\lambda_i} \exp\left(- \frac{K_{i,j}\frac{E_s}{4N_0}\lambda_i}{1 + \frac{E_s}{4N_0}\lambda_i} \right) \right). \tag{8}$$

We next examine some special cases.

The Case of Rayleigh Fading: In this case, $E\alpha_{i,j} = 0$ and as *a fortiori* $K_{i,j} = 0$ for all i and j. Then the inequality (8) can be written as

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \left(\frac{1}{\prod_{i=1}^{n}(1 + \lambda_i E_s/4N_0)} \right)^m. \tag{9}$$

Let r denote the rank of matrix A, then the kernel of A has dimension $n - r$ and exactly $n - r$ eigenvalues of A are zero. Say the nonzero eigenvalues of A are $\lambda_1, \lambda_2, \cdots, \lambda_r$, then it follows from inequality (9) that

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \left(\prod_{i=1}^{r} \lambda_i \right)^{-m} (E_s/4N_0)^{-rm}. \tag{10}$$

Thus a diversity advantage of mr and a coding advantage of $(\lambda_1\lambda_2 \cdots \lambda_r)^{1/r}$ is achieved. Recall that $\lambda_1\lambda_2 \cdots \lambda_r$ is the absolute value of the sum of determinants of all the principal $r \times r$ cofactors of A. Moreover, it is easy to see that the ranks of $A(\boldsymbol{c}, \boldsymbol{e})$, and $B(\boldsymbol{c}, \boldsymbol{e})$ are equal.

Remark: We note that the diversity advantage is the power of SNR in the denominator of the expression for the pairwise error probability derived above. The coding advantage is an approximate measure of the gain over an uncoded system operating with the same diversity advantage.

Thus from the above analysis, we arrive at the following design criterion.

Design Criteria for Rayleigh Space–Time Codes:

- *The Rank Criterion:* In order to achieve the maximum diversity mn, the matrix $B(\boldsymbol{c}, \boldsymbol{e})$ has to be full rank for any codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. If $B(\boldsymbol{c}, \boldsymbol{e})$ has minimum rank r over the set of two tuples of distinct codewords, then a diversity of rm is achieved. This criterion was also derived in [15].
- *The Determinant Criterion:* Suppose that a diversity benefit of rm is our target. The minimum of rth roots of the sum of determinants of all $r \times r$ principal cofactors of $A(\boldsymbol{c}, \boldsymbol{e}) = B(\boldsymbol{c}, \boldsymbol{e})B^*(\boldsymbol{c}, \boldsymbol{e})$ taken over all pairs of distinct codewords $\boldsymbol{e}$ and $\boldsymbol{c}$ corresponds to the coding advantage, where r is the rank of $A(\boldsymbol{c}, \boldsymbol{e})$. Special attention in the design must be paid to this quantity for any codewords $\boldsymbol{e}$ and $\boldsymbol{c}$. The design target is making this sum as large as possible. If a diversity of nm is the design target, then the minimum of the determinant of $A(\boldsymbol{c}, \boldsymbol{e})$ taken over all pairs of distinct codewords $\boldsymbol{e}$ and $\boldsymbol{c}$ must be maximized.

We next study the behavior of the right-hand side of inequality (8) for large signal-to-noise ratios. At sufficiently high signal-to-noise ratios, one can approximate the right-hand side of inequality (8) by

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \left(\frac{E_s}{4N_0} \right)^{-rm} \left(\prod_{i=1}^{r} \lambda_i \right)^{-m} \left[\prod_{j=1}^{m} \prod_{i=1}^{r} \exp(-K_{i,j}) \right]. \tag{11}$$

Thus a diversity of rm and a coding advantage of

$$(\lambda_1\lambda_2 \cdots \lambda_r)^{1/r} \left[\prod_{j=1}^{m} \prod_{i=1}^{r} \exp(-K_{i,j}) \right]^{1/rm}$$

is achieved. Thus the following design criteria is valid for the Rician space–time codes for large signal-to-noise ratios.

Design Criteria for The Rician Space–Time Codes:

- *The Rank Criterion:* This criterion is the same as that given for the Rayleigh channel.
- *The Coding Advantage Criterion:* Let $\Lambda(\boldsymbol{c}, \boldsymbol{e})$ denote the sum of all the determinants of $r \times r$ principal cofactors of $A(\boldsymbol{c}, \boldsymbol{e})$, where r is the rank of $A(\boldsymbol{c}, \boldsymbol{e})$. The minimum of the products

  $$\Lambda(\boldsymbol{c}, \boldsymbol{e})^{1/r} \left[\prod_{j=1}^{m} \prod_{i=1}^{r} \exp(-K_{i,j}) \right]^{1/rm}$$

  taken over distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$ has to be maximized.

  Note that one could still use the coding advantage criterion, since the performance will be at least as good as the right-hand side of inequality (9).

C. The Case of Dependent Fade Coefficients

In this subsection, we assume that the coefficients $\alpha_{i,j}$ are samples of possibly dependent zero-mean complex Gaussian random variables having variance 0.5 per dimension. This is the Rayleigh fading, but the extension to the Rician case is straightforward.

To this end, we consider the $mn \times mn$ matrix

$$Y(\boldsymbol{c}, \boldsymbol{e}) = \begin{pmatrix} A(\boldsymbol{c}, \boldsymbol{e}) & 0 & \cdots & \cdots & 0 & 0 \\ 0 & A(\boldsymbol{c}, \boldsymbol{e}) & \cdots & \cdots & 0 & 0 \\ 0 & 0 & A(\boldsymbol{c}, \boldsymbol{e}) & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & A(\boldsymbol{c}, \boldsymbol{e}) \end{pmatrix}$$

where 0 denotes the all-zero $n \times n$ matrix. If

$$\Omega = (\Omega_1, \cdots, \Omega_m)$$

,then (5) can be written as

$$P(\boldsymbol{c} \rightarrow \boldsymbol{e} | \alpha_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m) \leq \exp(-\Omega Y(\boldsymbol{c}, \boldsymbol{e})\Omega^* E_s/4N_0). \quad (12)$$

Let $\Theta = E\Omega^*\Omega$ denote the correlation matrix of Ω. We assume that Θ is full rank. The matrix Θ, being a nonnegative-definite square Hermitian matrix, has a square root C which is an $nm \times nm$ lower triangular matrix. The diagonal elements of Θ are unity, so that the rows of C are of length one. Let $\nu = \Omega(C^*)^{-1}$, then it is easy to see that the components of ν are uncorrelated complex Gaussian random variables with variance 0.5 per dimension. The mean of the components of ν can be easily computed from the mean of $\alpha_{i,j}$ and the matrix C. In particular, if the $\alpha_{i,j}$ have mean zero, so do the components of ν.

By (12), we arrive at the conclusion that

$$P(\boldsymbol{c} \rightarrow \boldsymbol{e} | \alpha_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m) \leq \exp(-\nu C^* Y(\boldsymbol{c}, \boldsymbol{e}) C \nu^* E_s/4N_0). \quad (13)$$

We can now follow the same argument as in the case of independent fades with $A(\boldsymbol{c}, \boldsymbol{e})$ replaced by $C^*Y(\boldsymbol{c}, \boldsymbol{e})C$. It follows that the rank of $C^*Y(\boldsymbol{c}, \boldsymbol{e})C$ has to be maximized. Since C is full rank, this amounts to maximizing

$$\text{rank}\,[Y(\boldsymbol{c}, \boldsymbol{e})] = m\,\text{rank}\,[A(\boldsymbol{c}, \boldsymbol{e})].$$

Thus the rank criterion given for the independent fade coefficients holds in this case as well.

Since $\alpha_{i,j}$ have zero mean, so do the components of ν. Thus by a similar argument to that of the case of independent fade coefficients, we arrive at the conclusion that the determinant of $C^*Y(\boldsymbol{c}, \boldsymbol{e})C$ must be maximized. This equals to

$$\det(\Theta)\det(Y(\boldsymbol{c}, \boldsymbol{e})) = \det(\Theta)[\det(A(\boldsymbol{c}, \boldsymbol{e}))]^m.$$

In this light the determinant criterion given in the case of independent fade coefficients holds as well. Furthermore, by comparing this case to the case of independent fade coefficients, it is observed that a penalty of $(10/nm)\log_{10}(\det(\Theta))$ decibels in the coding advantage occurs. This approximately quantifies the loss due to dependence.

It follows from a similar argument that the rank criterion is also valid for the Rician case and that any code designed for the Rayleigh channel performs well for the Rician channel even if the fade coefficients are dependent. To obtain the coding advantage criterion, one has to compute the mean of the components of ν and apply the coding advantage criterion given in the case of independent Rician fade coefficients. This is a straightforward but tedious computation.

D. The Case of Rapid Fading

When the fading is rapid, we model the channel by the mathematical equation

$$d_t^j = \sum_{i=1}^{n} \alpha_{i,j}(t) c_t^i \sqrt{E_s} + \eta_t^j. \quad (14)$$

The coefficients $\alpha_{i,j}(t)$ for $t = 1, 2, \cdots, l$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$ are modeled as independent samples of a complex Gaussian random variable with mean zero and variance 0.5 per dimension. This assumption corresponds to very fast Rayleigh fading but the generalization to Rician fading is straightforward. Also, η_t^j are samples of independent zero-mean complex Gaussian random variables with variance $N_0/2$ per dimension.

As in previous subsections, we assume that the coefficients $\alpha_{i,j}(t)$ for $t = 1, 2, \cdots, l$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$ are known to the decoder. The probability of transmitting

$$\boldsymbol{c} = c_1^1 c_1^2 \cdots c_1^n c_2^1 c_2^2 \cdots c_2^n \cdots c_l^1 c_l^2 \cdots c_l^n$$

and deciding in favor of

$$\boldsymbol{e} = e_1^1 e_1^2 \cdots e_1^n e_2^1 e_2^2 \cdots e_2^n \cdots e_l^1 e_l^2 \cdots e_l^n$$

at the maximum-likelihood decoder is well approximated by

$$P(\boldsymbol{c} \rightarrow \boldsymbol{e} | \alpha_{i,j}(t), i, j, t) \leq \exp(-d^2(\boldsymbol{c}, \boldsymbol{e}) E_s/4N_0)$$

where

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{j=1}^{m} \sum_{t=1}^{l} \left| \sum_{i=1}^{n} \alpha_{i,j}(t)(c_t^i - e_t^i) \right|^2.$$

This is just the standard approximation to the Gaussian tail function. Let

$$\Omega_j(t) = (\alpha_{1,j}(t), \alpha_{2,j}(t), \cdots, \alpha_{n,j}(t))$$

and $C(t)$ denote the $n \times n$ matrix with the element at pth row and qth column equal to $(c_t^p - e_t^p)\overline{(c_t^q - e_t^q)}$. Then it is easy to see that

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{j=1}^{m} \sum_{t=1}^{l} \Omega_j(t) C(t) \Omega_j^*(t).$$

The matrix $C(t)$ is Hermitian, thus there exist a unitary matrix $V(t)$ and a diagonal matrix $D(t)$ such that $C(t) = V(t)D(t)V^*(t)$ [20]. The diagonal elements of $D(t)$, denoted here by $D_{ii}(t)$, $1 \leq i \leq n$, are the eigenvalues of $C(t)$ counting multiplicities. Since $C(t)$ is Hermitian, these eigenvalues are real numbers. Let

$$(\beta_{1,j}(t), \cdots, \beta_{n,j}(t)) = \Omega_j(t) V(t)$$

then $\beta_{i,j}(t)$ for $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$, $t = 1, 2, \cdots, l$ are independent complex Gaussian variables with mean zero and variance 0.5 per dimension and

$$\Omega_j(t) C(t) \Omega_j^*(t) = \sum_{i=1}^{n} |\beta_{i,j}(t)|^2 D_{ii}(t).$$

By combining this with (3) and (15) and averaging with respect to the Rayleigh distribution of $|\beta_{i,j}(t)|$, we arrive at

$$P(\mathbf{c} \to \mathbf{e}) \le \prod_{i,t} \left(1 + D_{ii}(t)\frac{E_s}{4N_0}\right)^{-m}. \tag{15}$$

We next examine the matrix $C(t)$. The columns of $C(t)$ are all different multiples of

$$\mathbf{c}_t - \mathbf{e}_t = (c_t^1 - e_t^1, c_t^2 - e_t^2, \cdots, c_t^n - e_t^n).$$

Thus $C(t)$ has rank 1 if $c_t^1 c_t^2 \cdots c_t^n \neq e_t^1 e_t^2 \cdots e_t^n$ and rank zero otherwise. It follows that $n-1$ elements in the list

$$D_{11}(t), D_{22}(t), \cdots, D_{nn}(t)$$

are zeros and the only possibly nonzero element in this list is $|\mathbf{c}_t - \mathbf{e}_t|^2$. By (15), we can now conclude that

$$P(\mathbf{c} \to \mathbf{e}|\alpha_{i,j}(t), i, j, t) \le \prod_{t=1}^{l} \left(1 + |\mathbf{c}_t - \mathbf{e}_t|^2 \frac{E_s}{4N_0}\right)^{-m}. \tag{16}$$

Let $\mathcal{V}(\mathbf{c}, \mathbf{e})$ denote the set of time instances $1 \le t \le l$ such that $|\mathbf{c}_t - \mathbf{e}_t| \neq 0$ and let $|\mathcal{V}(\mathbf{c}, \mathbf{e})|$ denote the number of elements of $\mathcal{V}(\mathbf{c}, \mathbf{e})$. Then it follows from (16) that

$$P(\mathbf{c} \to \mathbf{e}) \le \prod_{t \in \mathcal{V}(\mathbf{c}, \mathbf{e})} \left(|\mathbf{c}_t - \mathbf{e}_t|^2 \frac{E_s}{4N_0}\right)^{-m}. \tag{17}$$

It follows that a diversity of $m|\mathcal{V}(\mathbf{c}, \mathbf{e})|$ is achieved. Examining the coefficient of $(E_s/4N_0)^{-m\mathcal{V}(\mathbf{c}, \mathbf{e})}$ leads to a design criterion.

Design Criteria for Rapid fading Rayleigh Channels:

- *The Distance Criterion:* In order to achieve the diversity vm in a rapid fading environment, for any two codewords $\mathbf{c}$ and $\mathbf{e}$ the strings $c_t^1 c_t^2 \cdots c_t^n$ and $e_t^1 e_t^2 \cdots e_t^n$ must be different at least for v values of $1 \le t \le l$.
- *The Product Criterion:* Let $\mathcal{V}(\mathbf{c}, \mathbf{e})$ denote the set of time instances $1 \le t \le l$ such that $c_t^1 c_t^2 \cdots c_t^n \neq e_t^1 e_t^2 \cdots e_t^n$ and let
$$|\mathbf{c}_t - \mathbf{e}_t|^2 = \sum_{i=1}^{n} |c_t^i - e_t^i|^2.$$
Then to achieve the most coding advantage in a rapid fading environment, the minimum of the products
$$\prod_{t \in \mathcal{V}(\mathbf{c}, \mathbf{e})} |\mathbf{c}_t - \mathbf{e}_t|^2$$
taken over distinct codewords $\mathbf{e}$ and $\mathbf{c}$ must be maximized.

III. Code Construction

A. Fundamental Limits on Outage Capacity

Let us consider a communication system employing n transmit and one receive antennas where the fading is quasistatic and flat. Intuition suggests that, there must come a point where adding more transmit antennas will not make much of a difference and this can be seen in the mathematics of outage capacity. Foschini and Gans [14] prove that the capacity of the aforementioned system is a random variable of the form $\log_2(1 + (\chi_{2n}^2/2n)\text{SNR})$, where χ_{2n}^2 is a random variable formed by summing the squares of $2n$ independent Gaussian random variables with mean zero and variance one. This means that by the strong law of large number $\chi_{2n}^2/2n \to 1$ in distribution. Practically speaking, for $n = 4$, $\chi_{2n}^2/2n \simeq 1$ and the capacity is the familiar Gaussian capacity $\log_2(1 + \text{SNR})$ per complex dimension. Thus in the presence of one receive antenna, little can be gained in terms of outage capacity by using more than four transmit antennas. A similar argument shows that if there are two receive antennas, almost all the capacity increase can be obtained using $n = 6$ transmit antennas. These observations also follow from the capacity plots given by Telatar [35]. This paper considers communication systems with at most two receive antennas, so we focus on the case that the number of transmit antennas is less than six. If more transmit and receive antennas are used, we can use the coding methods given in [33], where array processing and space–time coding are combined.

Our focus is mostly on low-delay applications. We thus only allow coding inside a frame of data as coding across different frames introduces delay. This emphasis on the method of coding motivated the choice of outage capacity (rather than Shannon's capacity) as the measure of achievable performance.

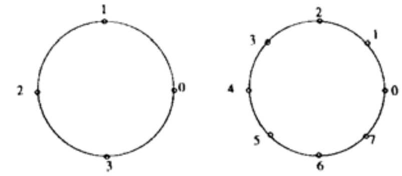

Fig. 3. 4-PSK and 8-PSK constellations.

B. Code Construction for Quasi-Static Flat Fading Channels

We proceed to use the criteria derived in the previous section to design trellis codes for a wireless communication system that employs n transmit antennas and (optional) receive antenna diversity where the channel is quasistatic flat fading channel. The encoding for these trellis codes is obvious, with the exception that *at the beginning and the end of each frame, the encoder is required to be in the zero state*. At each time t, depending on the state of the encoder and the input bits a transition branch is chosen. If the label of this branch is $q_t^1 q_t^2 \cdots q_t^n$, then transmit antenna i is used to send constellation symbols q_t^i, $i = 1, 2, \cdots, n$ and all these transmissions are simultaneous.

Let us consider the 4-PSK and 8-PSK constellations as given in Fig. 3. In Figs. 4–6, we provide 4-PSK codes for transmission of 2 b/s/Hz using two transmit antennas. Assuming, one receive antenna, these codes provide a diversity advantage of two. Similarly, in Figs. 7–9, we provide 8-PSK codes for transmission of 3 b/s/Hz using two transmit antennas. Assuming, one receive antenna, these codes provide a diversity advantage of two. We did not include the 64-state 4-PSK and 8-PSK codes for brevity of presentation.

We next consider decoding of these codes. Assuming ideal channel state information, the path gains $\alpha_{i,j}$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$ are known to the decoder.

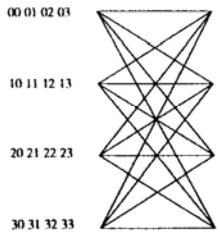

Fig. 4. 2-space–time code, 4-PSK, 4 states, 2 b/s/Hz.

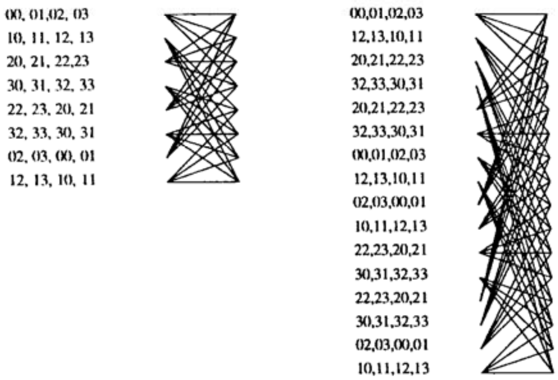

Fig. 5. 2-space–time codes, 4-PSK, 8 and 16 states, 2 b/s/Hz.

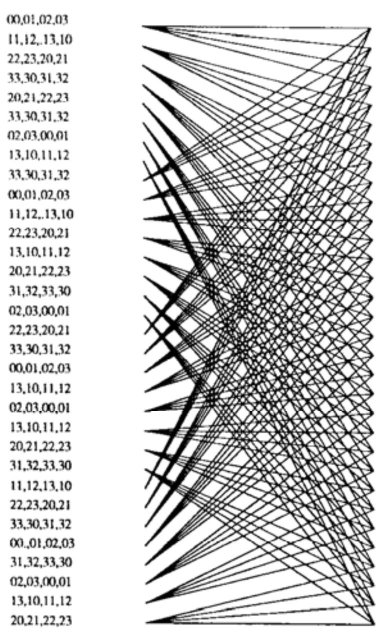

Fig. 6. 2-space–time code, 4-PSK, 32 states, 2 b/s/Hz.

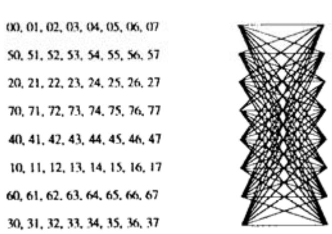

Fig. 7. 2-space–time code, 8-PSK, 8 states, 3 b/s/Hz.

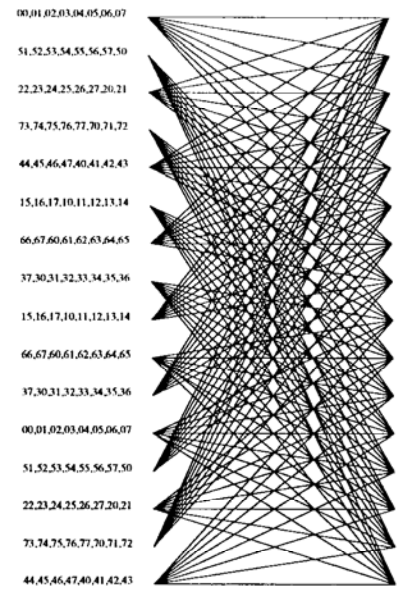

Fig. 8. 2-space–time code, 8-PSK, 16 states, 3 b/s/Hz.

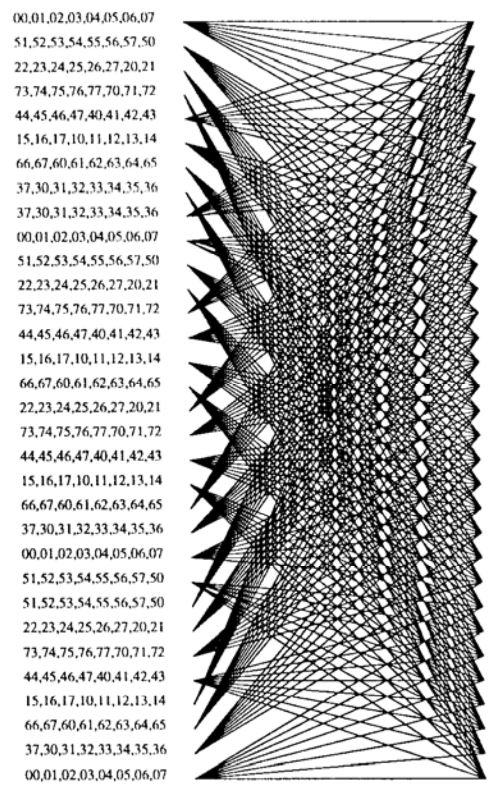

Fig. 9. 2-space–time code, 8-PSK, 32 states, 3 b/s/Hz.

Assuming that r_t^j is the received signal at receive antenna j at time t, the branch metric for a transition labeled $q_t^1 q_t^2 \cdots q_t^n$ is given by

$$\sum_{j=1}^{m} \left| r_t^j - \sum_{i=1}^{n} \alpha_{i,j} q_t^i \right|^2.$$

The Viterbi algorithm is then used to compute the path with the lowest accumulated metric. In the absence of ideal

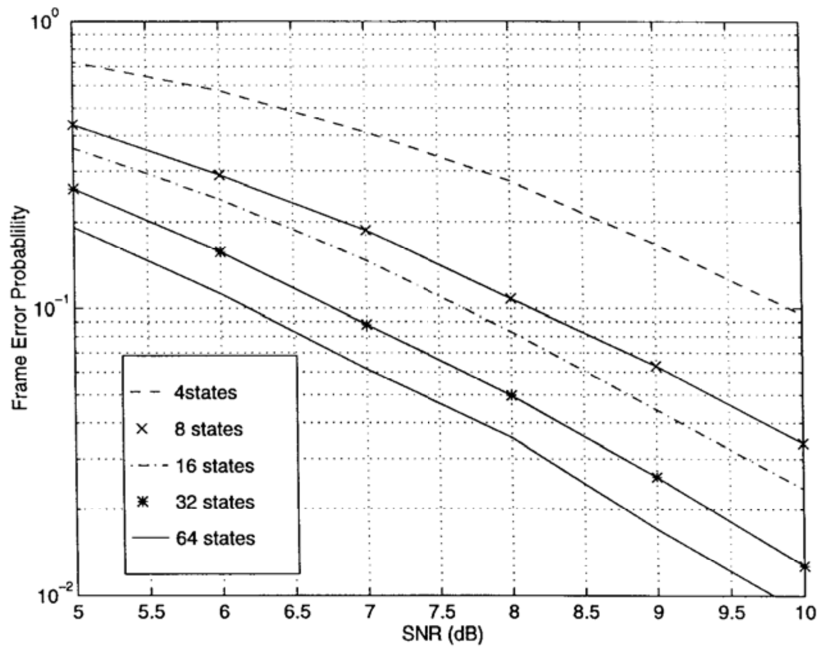

Fig. 10. Codes for 4-PSK with rate 2 b/s/Hz that achieve diversity 4 with two receive and two transmit antennas.

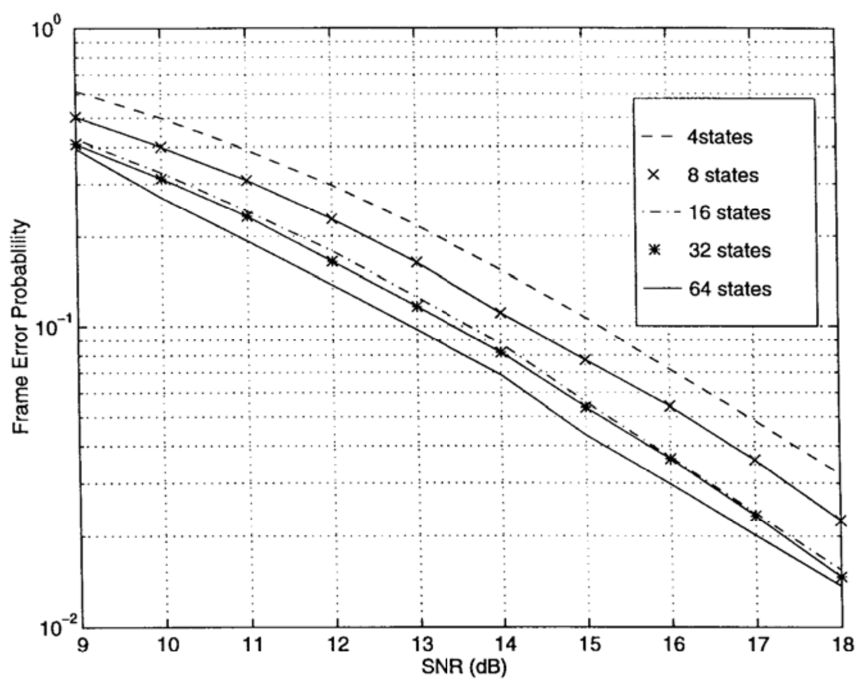

Fig. 11. Codes for 4-PSK with rate 2 b/s/Hz that achieve diversity 2 with one receive and two transmit antennas.

channel state information, an analysis carried in [34] gives the appropriate branch metrics. Channel estimation algorithm for this case is also considered in [34].

The aforementioned trellis codes are *space–time* trellis codes, as they combine spatial and temporal diversity techniques. Furthermore, if a space–time trellis code guarantees a diversity advantage of r for the quasistatic flat fading channel model described above (given one receive antenna), we say that it is an r-*space–time* trellis code. Thus the codes of Figs. 4–9 are 2-space-time codes.

In Figs. 10–13, we provide simulation results for the performance of these codes with two transmit and with one and two receive antennas. For comparison, the outage capacity given in [14] is included in Figs. 14 and 15. We observe that, at the frame error rate of 0.10 (In these simulations, each frame consists of 130 transmissions out of each transmit antenna.), the codes perform within 2.5 dB of the outage capacity. It appears from the simulation results that the coding advantage obtained by increasing the number of states increases as the number of receive antennas is increased. We also observe that

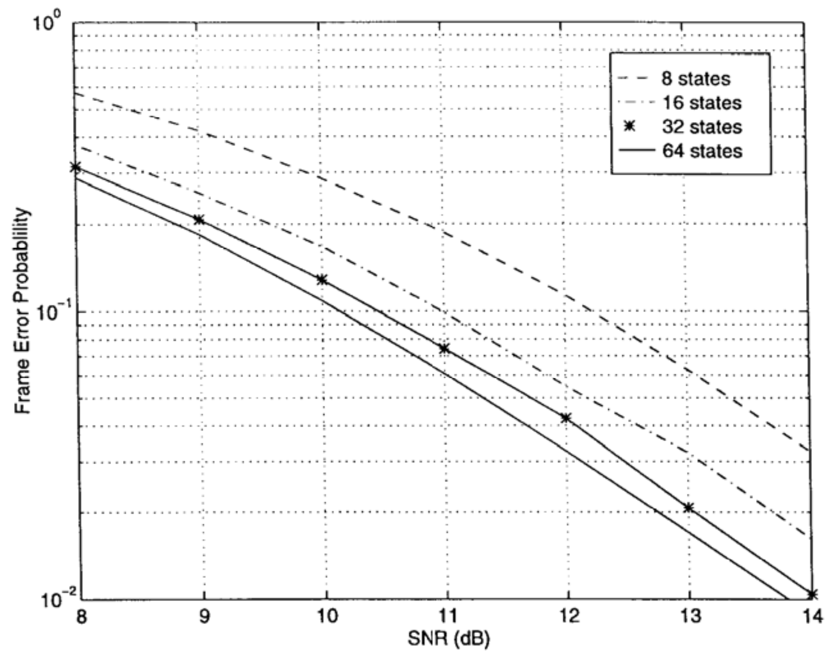

Fig. 12. Codes for 8-PSK with rate 3 b/s/Hz that achieve diversity 4 with two receive and two transmit antennas.

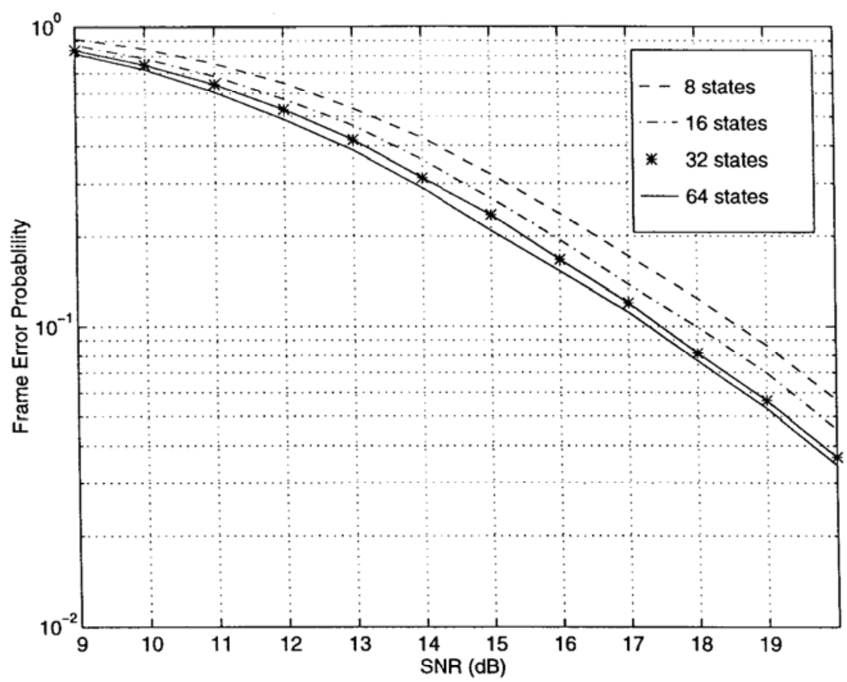

Fig. 13. Codes for 8-PSK with rate 3 b/s/Hz that achieve diversity 2 with one receive and two transmit antennas.

the coding advantage over the 4-state code is not as large as that forecasted by the determinant criterion. This is not unexpected, since the determinant criterion is approximate. For instance, it takes no account of path multiplicity. Furthermore, in the derivation of the design criteria, only the probability of confusing two distinct codewords was considered. In any case, simulation results demonstrate that the codes we constructed perform very well.

The above codes are designed by hand and for fixed rate, diversity advantage, constellation size, and trellis decoding complexity the designer sought to maximize the coding advantage given by the determinant criterion. A natural question is whether higher transmission rates are possible for 4-PSK and 8-PSK constellation rates using 2-space–time codes? A second question is whether simpler coding schemes exist? Fundamental questions of this type are the focus of the next section.

C. Tradeoff Between Rate, Diversity, Constellation Size, and Trellis Complexity

We shall derive fundamental tradeoff between transmission rate, diversity advantage, constellation size, and trellis

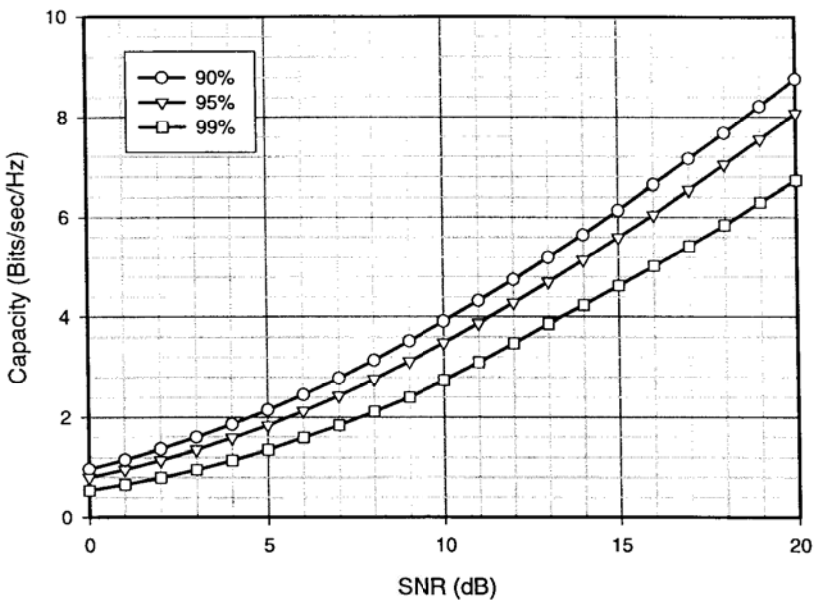

Fig. 14. Outage capacity for two receive and two transmit antennas.

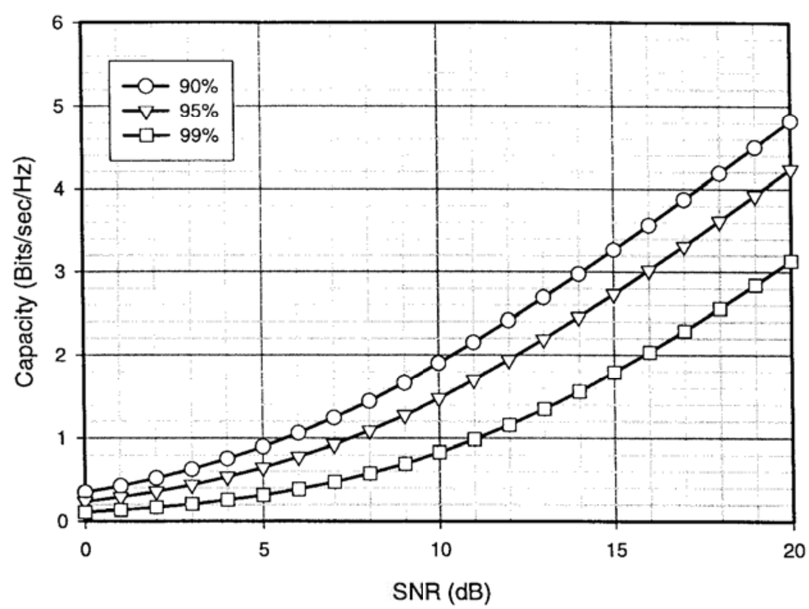

Fig. 15. Outage capacity for one receive and two transmit antennas.

decoding complexity. For fixed rate, diversity advantage, constellation size, and trellis decoding complexity, we seek to maximize the coding advantage given by the determinant criterion.

Consider a wireless system with n transmit and m receive antennas. It is known, from the result of previous sections that a maximum diversity of mn can be achieved. Our objective of code design must be achieving the maximum possible rate at a diversity advantage of rm. The following theorem addresses this issue.

Theorem 3.3.1: Consider an n transmit, m receive antenna mobile communication system with a Rician transmission model as given in the previous section. Let rm be the diversity advantage of the system. Assuming that the signal constellation Q has 2^b elements, the rate of transmission R satisfies

$$R \leq \frac{\log[A_{2^{bl}}(n, r)]}{l} \tag{18}$$

in bits per second per Hertz, where $A_{2^{bl}}(n, r)$ is the maximum size of a code length n and minimum Hamming distance r defined over an alphabet of size 2^{bl}.

Proof: Let l denote the frame length. We consider the superalphabet $Q^l = Q \times Q \times \cdots \times Q$ given by the l-folded Cartesian product of Q with itself. The mapping $f\colon Q^{ln} \to [Q^l]^n$ taking the codeword

$$e_1^1 e_1^2 \cdots e_1^n e_2^1 e_2^2 \cdots e_2^n \cdots e_l^1 e_l^2 \cdots e_l^n$$

in Q^{nl} to

$$[(e_1^1, e_2^1, \cdots e_l^1), (e_1^2, e_2^2, \cdots e_l^2), \cdots (e_1^n, e_2^n, \cdots, e_l^n)]$$

in $[Q^l]^n$ is one-to-one. By the rank criterion, the matrix $B(\boldsymbol{c}, \boldsymbol{e})$ given in (6) is of rank at least r for any two distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. Thus at least r rows of $B(\boldsymbol{c}, \boldsymbol{e})$ are nonzero. It follows that $f(\boldsymbol{c})$ and $f(\boldsymbol{e})$ have Hamming distance at least r as codewords defined over Q^l. The alphabet Q^l has size 2^{bl}, thus the number of codewords is bounded above by $A_{2^{bl}}(n, r)$. It follows that the rate of transmission is bounded above by (18). □

Corollary 3.3.1: Consider the Rician transmission model with n transmit and m receive antennas. If the diversity advantage is nm, then the transmission rate is at most b bits per second per hertz.

Proof: It is known that $A_{2^{bl}}(n, n) = 2^{bl}$ and this is achieved by a repetition code [24]. □

Remark: For 4-PSK, 8-PSK, or 16-QAM constellations, respectively, a diversity advantage of nm places an upper bound on transmission rate of 2, 3, and 4 b/s/Hz.

It follows also from the above that there is a fundamental tradeoff between constellation size, diversity, and the transmission rate. We relate this tradeoff to the trellis complexity of the code.

Lemma 3.3.1: The constraint length of an r-space–time trellis code is at least $r - 1$.

Proof: Consider two parallel transitions corresponding to the constraint length ν in the trellis diagram. Without loss of generality, we may assume that one of these transitions corresponds to all-zero path $00000\cdots 0$ and the other corresponds to

$$e_1^1 e_1^2 \cdots e_1^n e_2^1 e_2^2 \cdots e_2^n \cdots e_{\nu+1}^1 e_{\nu+1}^2 \cdots e_{\nu+1}^n 00\cdots 0000.$$

If $\nu < r - 1$, the rank criterion is easily seen to be violated. □

Lemma 3.3.2: Let b denote the transmission rate of a multiple-antenna system employed in conjunction with an r-space–time trellis code. The trellis complexity of the space–time code is at least $2^{b(r-1)}$.

Proof: Since the transmission rate is b bits per second per hertz, the number of branches leaving each state of the trellis diagram is 2^b. Thus at time instance $r - 1$, there are $2^{b(r-1)}$ paths that have diverged from the zero state of the trellis at time zero. By Lemma 3.3.1, none of these paths can merge at the same state. Thus there are at least $2^{b(r-1)}$ states in the trellis. □

The codes constructed in Figs. 4 and 7 and some of those to be constructed later, achieve this upper bound. Thus the bound of Theorem 3.3.1 is tight. This also means that these codes produce the optimal tradeoff between the transmission rate, diversity, trellis complexity, and constellation size.

D. Geometrical Uniformity and Its Applications

For the Gaussian channel, the method of constructing trellis codes based on lattices and cosets allowed coding theorists to work with larger constellations and more complicated set partitioning schemes [7]. Here, we examine the algebraic structure of the codes presented in Section III-B. We begin with the code of Fig. 4. This is an example of delay diversity codes to be discussed later.

Example 3.4.1: Here the signal constellation is 4-PSK, where the signal points are labeled by the elements of $\mathbb{Z}_4$, the ring of integers modulo 4 as shown in Fig. 3. We consider the 4-state trellis code shown in Fig. 4. The edge label $x_1 x_2$ indicates that signal x_1 is transmitted over the first antenna and that signal x_2 is transmitted over the second antenna. This code has a very simple description in terms of a sequence (b_k, a_k) of binary inputs. The output signal pair $x_1^k x_2^k$ at time k is given by

$$(x_1^k, x_2^k) = b_{k-1}(2, 0) + a_{k-1}(1, 0) + b_k(0, 2) + a_k(0, 1) \tag{19}$$

where the addition takes place in $\mathbb{Z}_4$ (cf. Calderbank and Sloane [7]).

Following Forney [12], we shall say that a code is geometrically uniform if given any two codewords x, y, there is an isometry $\phi_{x,y}$ permuting the set of codewords such that $\phi_{x,y}(x) = y$. For Rician transmission as above, the isometries are unitary transformations of the underlying Complex space. If a space–time code is geometrically uniform, then it is easy to see that the performance is independent of the transmitted codeword [12]. We claim that the code of Fig. 4 is geometrically uniform.

To this end, let $R_1 = (1, 0)(2, 3)$ and $R_2 = (1, 2)(0, 3)$ be permutations of the elements of 4-PSK constellation. The permutations R_1 and R_2 are realized by reflection in the bisectors of the first and second quadrants of the complex plane, respectively. In this light, they are isometries of the complex space.

Given a codeword $\boldsymbol{x}$ of the code of Fig. 4, we consider the corresponding sequence (b_k, a_k) of binary inputs. Let $\phi_k : (\mathbb{C} \times \mathbb{C}) \to (\mathbb{C} \times \mathbb{C})$ be the isometry given by

$$(R_1^{a_{k-1}}(R_1 R_2)^{b_{k-1}}, R_1^{a_k}(R_1 R_2)^{b_k})).$$

Then

$$\cdots \times \phi_0 \times \phi_1 \times \phi_2 \times \cdots$$

maps the all zero codeword to $\boldsymbol{x}$ while preserving the code. This proves the claim.

For a diversity advantage of 2, it is required that for any pair of distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$ the matrix

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

must have rank 2. This is evident from Fig. 4 or from the algebraic description (19), for if the paths corresponding to codewords $\boldsymbol{c}$ and $\boldsymbol{e}$ diverge at time t_1 and remerge at time t_2, then the vectors $(e_1^{t_1} - c_1^{t_1}, e_2^{t_1} - c_2^{t_1})$ and $(e_1^{t_2} - c_1^{t_2}, e_2^{t_2} - c_2^{t_2})$ are linearly independent. In fact, $e_1^{t_1} - c_1^{t_1} = e_2^{t_2} - c_2^{t_2} = 0$, $e_2^{t_1} - c_2^{t_1} \neq 0$, and $e_1^{t_2} - c_1^{t_2} \neq 0$.

To compute the coding advantage, we need to find codewords $\boldsymbol{c}$ and $\boldsymbol{e}$ such that the determinant

$$\det\left(\sum_{k=1}^{l} (e_k^1 - c_k^1, e_k^2 - c_k^2)^*(e_k^1 - c_k^1, e_k^2 - c_k^2)\right) \tag{20}$$

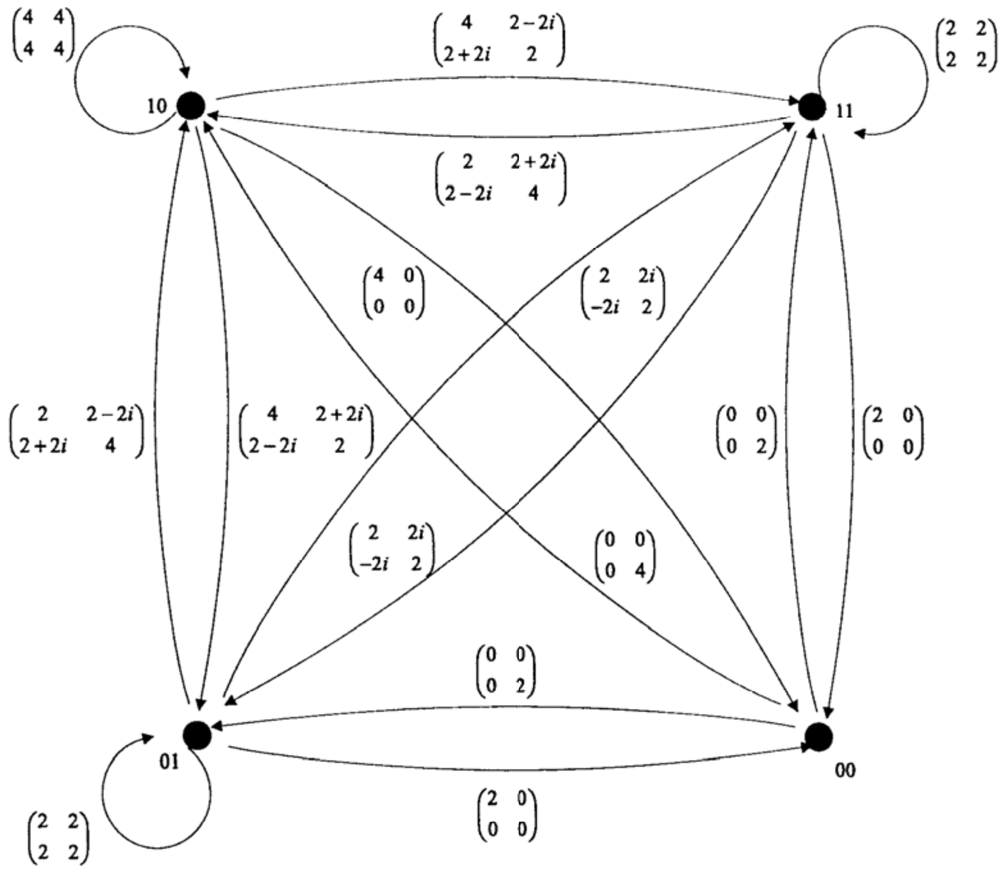

Fig. 16. State diagram of Example 3.4.1.

is minimized. As the code of this example is geometrically uniform, we could assume without loss of generality that $\boldsymbol{c}$ is the all zero codeword. We can attack (20) by replacing the edge label (x_1, x_2) by the complex matrix

$$\begin{pmatrix} (\boldsymbol{j}^{-x_1}-1)(\boldsymbol{j}^{x_1}-1) & (\boldsymbol{j}^{x_1}-1)(\boldsymbol{j}^{-x_2}-1) \\ (\boldsymbol{j}^{-x_1}-1)(\boldsymbol{j}^{x_2}-1) & (\boldsymbol{j}^{x_2}-1)(\boldsymbol{j}^{-x_2}-1) \end{pmatrix}.$$

This labeling is shown in Fig. 16.

Diverging from the zero state contributes a matrix of the form

$$\begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}$$

and remerging to the zero state contributes a matrix of the form

$$\begin{pmatrix} s & 0 \\ 0 & 0 \end{pmatrix}$$

where $s, t \geq 2$. Thus (20) can be written as

$$\det\left[\begin{pmatrix} s & 0 \\ 0 & t \end{pmatrix} + \begin{pmatrix} a & b \\ \bar{b} & d \end{pmatrix}\right] \tag{21}$$

with $a, d \geq 0$ and $|b|^2 \leq ad$. Hence the minimum determinant is 4. □

Remark: It is straightforward to prove that the codes of the previous section are geometrically uniform. Indeed, we examine the 4-PSK trellis codes with 8, 16, and 32 states in Figs. 5 and 6. These codes can be, respectively, expressed by equations

$$\begin{aligned}(x_1^k, x_2^k) &= a_{k-2}(2, 2) + b_{k-1}(2, 0) + a_{k-1}(1, 0) \\ &\quad + b_k(0, 2) + a_k(0, 1) \\ (x_1^k, x_2^k) &= b_{k-2}(0, 2) + a_{k-2}(2, 0) + b_{k-1}(2, 0) \\ &\quad + a_{k-1}(1, 2) + b_k(0, 2) + a_k(0, 1) \\ (x_1^k, x_2^k) &= a_{k-3}(2, 2) + b_{k-2}(3, 3) + a_{k-2}(2, 0) \\ &\quad + b_{k-1}(2, 2) + a_{k-1}(1, 1) \\ &\quad + b_k(0, 2) + a_k(0, 1)\end{aligned}$$

in $\mathbb{Z}_4$, using the same notation as the one employed in Example 3.4.1. These codes are geometrically uniform. The minimum determinants are, respectively, 12, 20, and 28.

The *design rules* that guarantee the diversity in Figs. 4 and 7 are as follows.

- *Design Rule 1:* Transitions departing from the same state differ in the second symbol.
- *Design Rule 2:* Transitions arriving at the same state differ in the first symbol.

The rest of the codes are a bit trickier to analyze but it can be confirmed using geometrical uniformity that the diversity advantage is actually achieved.

E. Optimal Codes

Here, we construct some other codes that are optimal with respect to the fundamental tradeoffs between rate, diversity, constellation size, and trellis complexity. First, we consider the case when $n = 2$ and design 2-space–time trellis codes. Suppose that the constellation has $Q = 2^b$ elements. By Corollary 3.3.1, the maximum transmission rate is b bits per second per hertz. On the other hand, Lemma 3.3.2 implies that the number of states of any 2-space–time trellis code is at least 2^b. The following lemma proves that all these bounds can be attained together.

Lemma 3.5.1: There exists 2-space–time trellis codes defined over a constellation of size 2^b having trellis complexity 2^b and transmission rate b bits per second per hertz.

Proof: Every block of b bits naturally corresponds to an element of $\mathbb{Z}_{2^b}$, the ring of integers modulo 2^b. The constellation alphabet can also be labeled with elements of $\mathbb{Z}_{2^b}$ in a one-to-one and onto manner. Thus without loss of generality, we identify both the input blocks and the states of the encoder with the elements of $\mathbb{Z}_{2^b}$. We consider the trellis code having 2^b states corresponding to elements of $\mathbb{Z}_{2^b}$ as defined next. Given that a block of length b of bits corresponding to i is the input to the encoder and the encoder is at state $s \in \mathbb{Z}_{2^b}$, the label of the transmission branch is (i, s). The new state of the encoder is i.

Given two distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$, the associated paths in the trellis emerge from a state at time t_1 and remerge in another state at a later time $t_2 \le l$. It is easy to see that the t_1th and t_2th columns of the matrix $B(\boldsymbol{c}, \boldsymbol{e})$ are independent. □

Remark: The construction given above is just delay diversity expressed in algebraic terms.

For the 4-PSK constellation, the code given by the above Lemma appears in Fig. 4.

For the 8-PSK constellations, the code given by the above lemma appears in Fig. 17. One can also consider the code of Fig. 7. Assuming that the input to the encoder at time k is the 3 input bits (d_k, b_k, a_k), the output of the encoder at time k is

$$\begin{aligned}(x_1^k, x_2^k) &= d_{k-1}(4, 0) + b_{k-1}(2, 0) + a_{k-1}(5, 0)\\ &\quad + d_k(0, 4) + b_k(0, 2) + a_k(0, 1)\end{aligned}$$

where the computation is performed in $\mathbb{Z}_8$, the ring of integers modulo 8, and the elements of the 8-PSK constellation have the labeling given in Fig. 3. Design Rules 1 and 2 guarantee diversity advantage 2 for this code. We believe that the above code optimizes the coding advantage (determinant criterion), but unfortunately have not been able to prove this conjecture. The minimum determinant of this code is 2.

As in the 4-PSK case, one can improve the coding advantage of the above codes by constructing encoders with more states. In fact, using the design criterion established in this paper, we have designed 2-space–time trellis codes with number of states up to 64 for 8-PSK and 16-QAM constellations. We include the 16-state 16-QAM code as well (Figs. 18 and 19), but for brevity, we avoided including the rest of these codes. Design rules 1 and 2 (or simple extensions thereof) guarantee diversity 2 in all cases.

We conjecture that most of the codes presented above are the best in terms of the determinant criterion, but we do not have a proof to this effect.

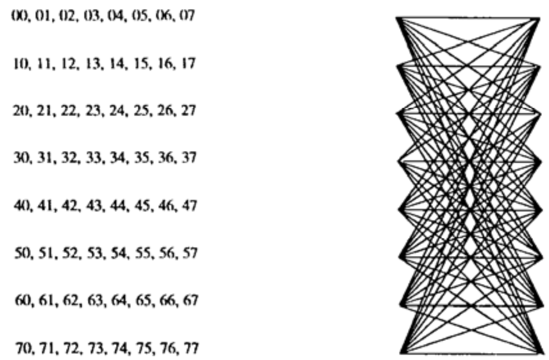

Fig. 17. Space–time realization of a delay diversity 8-PSK code constructed from a repetition code.

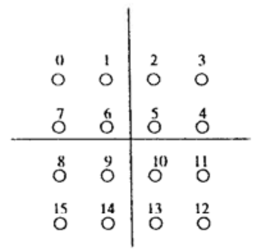

Fig. 18. The QAM constellation.

F. An r-Space–Time Trellis Code for $r > 2$

Here, we design r-space–time codes for $r > 2$. We construct a 4-space–time code for a 4 transmit antenna mobile communication system. The limit on transmission rate is 2 b/s/Hz. Thus the trellis complexity of the code is bounded below by 64. The input to the encoder is a block of length 2 of bits a_1, b_1 corresponding to an integer $i = 2a_1 + b_1 \in \mathbb{Z}_4$. The 64 states of the trellis correspond to the set of all three tuples (s_1, s_2, s_3) with $s_i \in \mathbb{Z}_4$ for $1 \le i \le 3$. At state (s_1, s_2, s_3) upon input data i, the encoder outputs (i, s_1, s_2, s_3) elements of 4-PSK constellation (see Fig. 3) and moves to state (i, s_1, s_2). Given two codewords $\boldsymbol{c}$ and $\boldsymbol{e}$, the associated paths in the trellis diverge at time t_1 from a state and remerge in another state at a later time $t_2 \le l$. It is easy to see that the t_1th, (t_1+1)th, (t_2-1)th, and t_2th columns of the matrix $B(\boldsymbol{c}, \boldsymbol{e})$ are independent. Thus the above design gives a 4-space–time code.

G. Coding with Delay Diversity

Here we observe that the delay diversity scheme of [32] and [44] can be viewed as space–time coding, and that our methods for analyzing performance apply to these codes. Indeed, consider the delay diversity scheme of Fig. 1, where the channel encoder is a rate $1/2$ block repetition code defined over some signal constellation alphabet. This can be viewed as a space–time code by defining

$$\begin{aligned}c_t^1 &= \tilde{c}_{t-1}^1\\ c_t^2 &= \tilde{c}_t^2\end{aligned}$$

where c_t^1 and c_t^2 are the symbols of the equivalent space–time code at time t and $\tilde{c}_t^1 \tilde{c}_t^2$ is the output of the encoder at time t.

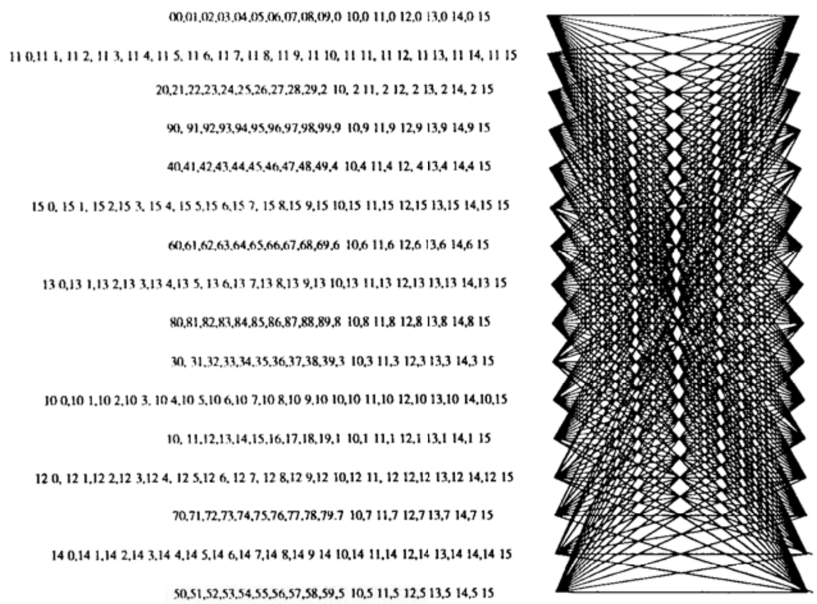

Fig. 19. 2-space–time 16-QAM code, 16 states, 4 b/s/Hz.

Next consider the 8-PSK signal constellation, where the encoder maps a sequence of three bits $a_k b_k c_k$ at time k to ii with $i = 4a_k + 2b_k + a_k$. It is easy to show that the equivalent space–time code for this delay diversity code has the trellis representation given in Fig. 17. The minimum determinant of this code is $(2 - \sqrt{2})^2$.

Next, we consider the block code

$$\mathcal{C} = \{00,\ 15,\ 22,\ 37,\ 44,\ 51,\ 66,\ 73\}$$

of length 2 defined over the alphabet 8-PSK instead of the repetition code. This block code is the best in the sense of product distance [32] among all the codes of cardinality 8 and of length 2 defined over the alphabet 8-PSK. This means that the minimum of the product distance $|c_1 - e_1||c_2 - e_2|$ between pairs of distinct codewords $\boldsymbol{c} = c_1c_2 \in \mathcal{C}$ and $\boldsymbol{e} = e_1e_2 \in \mathcal{C}$ is maximum among all such codes. The delay diversity code constructed from this block code is identical to the space–time code given by trellis diagram of Fig. 7. The minimum determinant of this delay diversity code is thus 2.

The 16-state code for the 16-QAM constellation given in Fig. 19, is obtained from the block code

$$\{0\,0,\ 1\,11,\ 2\,2,\ 3\,9,\ 4\,4,\ 5\,15,\ 6\,6,\ 7\,13,\ 8\,8,\ 9\,3,\ 10\,10,\ 11\,1,\ 12\,12,\ 13\,7,\ 14\,14,\ 15\,5\}$$

using the same delay diversity construction. Again, this block code is optimal in the sense of product distance.

The delay diversity code construction can also be generalized to systems having more than two transmit antennas. For instance, the 4-PSK 4-space–time code given before is a delay diversity code. The corresponding block code is the repetition code. By applying the delay diversity construction to the 4-PSK block code

$$\{0\,0\,0\,0,\ 1\,2\,3\,1,\ 2\,1\,2\,3\,,\ 3\,3\,1\,2\}$$

one can obtain a more powerful 4-PSK 4-space–time code having the same trellis complexity.

It is an interesting open problem whether it is possible to construct good space–time codes of a given complexity using coding in conjunction with delay diversity. Note that coding is an integral part of the delay diversity arrangement and is not to be confused with outer coding.

H. Multilevel Space–Time Coding

Imai and Hirakawa [21] described a multilevel method for constructing codes where the transmitted symbols are obtained by combining codeword symbols from the component codes. They also introduced a suboptimal multistage decoding algorithm. Multilevel coding has been extended to Gaussian channels (see [6] and the references therein).

Space–time codes may be designed with multilevel structure, and multistage decoding can be useful in some practical communication systems, particularly when the number of transmit antennas is high. This has the significant advantage of reducing the decoding complexity.

Without loss of generality, we assume a signal constellation Q_0 consisting of 2^{b_0} signal points and a set partitioning of Q_0 based on subsets

$$Q_{f-1} \subset Q_{f-2} \subset \cdots \subset Q_1 \subset Q_0$$

where the number of elements of Q_k is equal to 2^{b_k} for all $0 \le k \le f-1$. By such a set partitioning, we mean that Q_0 is the union of $2^{b_0-b_1}$ disjoint sets called *cosets of* Q_1 in Q_0, each having 2^{b_1} elements. The collection of $2^{b_0-b_1}$ cosets of Q_1 in Q_0 must include Q_1 as an element. Having the cosets of Q_1 in Q_0 at hand, each coset is then divided into $2^{b_1-b_2}$ disjoint sets each having 2^{b_2} elements. The $2^{b_1-b_2}$ subsets of Q_1 are called cosets of Q_2 in Q_1. The collection of cosets of Q_2 in Q_1 must include Q_2. Thus there are $2^{b_0-b_2}$ subsets of Q_0 with 2^{b_2} elements called the cosets of Q_2 in Q_0. Trivially, the collection of cosets of Q_2 in Q_0 includes Q_2. This procedure is repeated until we arrive at cosets of Q_k in Q_w for all $0 \le w < k \le f-1$. Let $r_{f-1} = b_{f-1}$ and $r_k = b_{k+1} - b_k$ for

$k = 0, 1, \cdots, f-2$. Then Q_k contains 2^{r_k} cosets of Q_{k+1} for all $k = 0, 1, \cdots, f-2$. Set partitioning of QAM and PSK constellations were first introduced by Ungerboeck [36].

Corresponding to the aforementioned set partitioning, there exist f space–time encoders, namely, encoders $E_0, E_2, \cdots, E_{f-1}$. It is required that all these encoders have a trellis representation. Every $K = K_0 + \cdots + K_{f-1}$ bits of input data is encoded using encoders $E_0, \cdots, E_{f-1}$ corresponding to the f levels.

At each time t depending on the state of the jth encoder and the input data, a branch of the trellis of the kth encoder is chosen which is labeled with n blocks of r_k bits denoted by $B_t^1(k), B_t^2(k), \cdots, B_t^n(k)$. For each $1 \le i \le n$, the blocks $B_t^i(0), \cdots, B_t^i(f-1)$ then choose a point of the signal constellation in the following way: the block $B_t^i(0)$ chooses a coset Q_1' of Q_1 in Q_0. The block $B_t^i(1)$ chooses a coset Q_2' of Q_2 in Q_0 which is also a subset of Q_1', and so forth. Finally, the block $B_t^i(f-1)$ chooses a point of Q_{f-1}'. The chosen point is then transmitted using the ith antenna at time t. Multilevel decoding is described in [21].

Let us suppose that the encoder of the kth level has 2^{S_k} states at time t. One can view the multilevel code described above as a space–time code C with $2^{(S_0+\cdots+S_{f-1})}$ states at time t. The states of C at time t correspond to f-tuples $(s_t^0, s_t^1, \cdots, s_t^{f-1})$ of states of encoders $0, 1, \cdots, f-1$ at that time. There is a branch between states $(s_t^0, s_t^1, \cdots, s_t^{f-1})$ and $(s_{t+1}^0, s_{t+1}^1, \cdots, s_{t+1}^{f-1})$ when E_k goes from state s_t^k to s_{t+1}^k for all $0 \le k \le f-1$. In this case, the branch labels between these states is the set of symbols that are sent via antennas $1, 2, \cdots, n$ when encoders E_k, $k = 0, \cdots, f-1$ move from state s_t^k to the state s_{t+1}^k for all $0 \le k \le f-1$. In this way, one can view a multilevel space–time code as a regular space–time code with a multilevel structure that allows simplified decoding. The penalty for this simplified decoding is a loss in performance due in part to magnification of the effective error coefficient. Also, in this way the design criterion derived previously could be applied to the space–time code C. Alternatively, the criteria can be applied to the trellis of each encoder providing different diversities at each level with the levels decoded first given the higher diversities.

We provide an example of multilevel coding.

Consider a scheme using $n = 3$ transmit antennas and an 8-PSK constellation. Suppose that a data rate of 5 b/s/Hz is desired. We construct a multilevel scheme that has this data rate and provides diversity advantage 2. If trellis space–time coding is employed, at least $2^5 = 32$ states are required with 32 transitions leaving each state of the trellis. Instead, we employ a multilevel code with multistage decoding [30].

At each time t the input to the encoder is five bits of information $b_t^1 b_t^2 b_t^3 b_t^4 b_t^5$. The input sequence b_t^5 is encoded using a repetition code of rate 1/3 giving the output sequence $b_t^5 b_t^5 b_t^5$. The pair of bits $b_t^1 b_t^2$ and $b_t^3 b_t^4$ are encoded using a parity-check code of rate 2/3 yielding sequences $b_t^1 b_t^2 b_t^6$ and $b_t^3 b_t^4 b_t^7$. Let

$$\begin{aligned} c_t^1 &= 4b_t^1 + 2b_t^3 + b_t^5 \\ c_t^2 &= 4b_{t-1}^2 + 2b_{t-1}^4 + b_{t-1}^5 \\ c_t^3 &= 4b_{t-2}^6 + 2b_{t-2}^7 + b_{t-2}^5 \end{aligned}$$

be elements of the 8-PSK constellation, where the labeling is given in Fig. 3. The transmitted signal from antenna $1 \le i \le 3$ at time t is c_t^i.

At the decoder multistage decoding is performed. At first, a decision on b_t^5 is made. A trellis diagram for b_t^5 has only four states where the states depend on b_{t-1}^5 and b_{t-2}^5. In such a trellis diagram each branch has 15 parallel branches. There are 32 branches leaving each state. It is easy to use the criterion developed in this paper and observe that a diversity advantage of 3 on deciding the bits $b_1^5, b_2^5, \cdots, b_l^5$ is guaranteed.

Assuming that $b_1^5, b_2^5, \cdots, b_l^5$ are determined, the multistage decoder performs decoding to determine $b_t^3 b_t^4$. Here, the states at time t are given by the triplet $(b_{t-1}^3, b_{t-1}^4, b_{t-2}^7)$, so there are eight states in the trellis diagram. There are four parallel transitions between any two connected states. The criteria for diversity can be used to observe that assuming correct decisions in the first stage of decoding, a diversity advantage of two is achieved in the second stage.

In the third stage, the multilevel decoder determines $b_t^1 b_t^2$ using a trellis. The states at time t are given by the triplet $(b_{t-1}^1, b_{t-1}^2, b_{t-2}^6)$, so there are eight states in the trellis diagram. There are no parallel transitions between any two connected states. Assuming correct decisions in the first and second stage of decoding, a diversity advantage of two is achieved in the third stage.

The total number of branches visited in decoding this multilevel scheme is almost half as much as the one given by the trellis space–time code having 32 states. Thus it is natural to expect that multilevel coding is a good way to produce powerful space–time codes for various high-bit-rate applications if the number of antennas at the base-station is high.

I. Space–Time Codes That Exploit Temporal Variations: Smart Greedy Space–Time Codes

This subsection addresses the important problem of constructing codes for data transmission, not at rates greater than today's wireless systems, but operating at significantly lower signal-to-noise ratios. This provides a better frequency reuse factor. The second key issue addressed here is designing codes that can take advantage of possible temporal variations in a wireless channel to provide additional diversity. This has use in providing quality service to both low- and high-mobility users.

We provide examples of codes that address both these key issues and refer to them as *low-rate multidimensional space–time codes for both slow and rapid fading channels* or *smart–greedy space–time* codes. At the very highest level, these are concatenated codes. As the function of the outer code is fixing a small number of symbol errors, we focus on the design of the inner code. The code is called *smart and greedy* because the encoder does not know the channel but can exploit the benefits provided both by the transmit and receive antennas as well as by possible rapid changes in the channel. It is assumed that the transmitter does not know the channel but seeks to choose a codebook that guarantees a diversity gain of r_1 when there is no mobility and a diversity gain of $r_2 \ge r_1$ when the channel is fast fading.

Fig. 20. The BPSK constellation.

When the fading is slow, it will be modeled as in Section II as quasistatic. When fading is rapid, it will be modeled as in Section II-D. In reality, we know that the situation is something between these two extremes. It is thus expected that a code designed using a hybrid criteria given by these two extremes will perform well in a variety of mobility conditions.

We thus combine the criteria obtained in those subsections to arrive at a hybrid design criteria.

A Hybrid Design Criteria for Smart–Greedy Space–Time Codes:

- *The Distance/Rank Criterion:* In order to achieve the diversity vm in a rapid fading environment, for any two codewords $\boldsymbol{c}$ and $\boldsymbol{e}$ the strings $c_t^1 c_t^2 \cdots c_t^n$ and $e_t^1 e_t^2 \cdots e_t^n$ must be different at least for v values of $1 \le t \le l$. Furthermore, let

$$B(\boldsymbol{c}, \boldsymbol{e}) = \begin{pmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \cdots & \cdots & e_l^1 - c_l^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \cdots & \cdots & e_l^2 - c_l^2 \\ e_1^3 - c_1^3 & e_2^3 - c_2^3 & \ddots & \vdots & e_l^3 - c_l^3 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ e_1^n - c_1^n & e_2^n - c_2^n & \cdots & \cdots & e_l^n - c_l^n \end{pmatrix}. \quad (22)$$

  If $B(\boldsymbol{c}, \boldsymbol{e})$ has minimum rank r over the set of pairs of distinct codewords, then a diversity of rm is achieved in static flat fading environments.

- *The Product/Determinant Criterion:* Let $\mathcal{V}(\boldsymbol{c}, \boldsymbol{e})$ denote the set of time instances $1 \le t \le l$ such that $c_t^1 c_t^2 \cdots c_t^n \ne e_t^1 e_t^2 \cdots e_t^n$ and let

$$|\boldsymbol{c}_t - \boldsymbol{e}_t|^2 = \sum_{i=1}^{n} |c_t^i - e_t^i|^2.$$

  Then to achieve the most coding advantage in a rapid fading environment, the minimum of the products

$$\prod_{t \in \mathcal{V}(\boldsymbol{c}, \boldsymbol{e})} |\boldsymbol{c}_t - \boldsymbol{e}_t|^2$$

  taken over distinct codewords $\boldsymbol{e}$ and $\boldsymbol{c}$ must be maximized. For the case of a static fading channel, the minimum of rth roots of the sum of determinants of all $r \times r$ principal cofactors of $A(\boldsymbol{c}, \boldsymbol{e}) = B(\boldsymbol{c}, \boldsymbol{e}) B^*(\boldsymbol{c}, \boldsymbol{e})$ taken over all pairs of distinct codewords $\boldsymbol{e}$ and $\boldsymbol{c}$ corresponds to the coding advantage, where r is the rank of $A(\boldsymbol{c}, \boldsymbol{e})$.

Using the above design criteria, we constructed smart–greedy codes for both slow and fast fading channels. We illustrate the construction of these codes by some examples. In all these examples, it is again assumed that at the beginning and the end of each frame, the encoder of the code is at zero state.

Example 3.9.1: Suppose that a transmission rate of 0.5 b/s/Hz is required. In this example, we will use the BPSK constellation. The constellation points are given in Fig. 20.

Our objective is to guarantee a diversity advantage of 2 and 4, respectively, in slow and rapid flat fading environments.

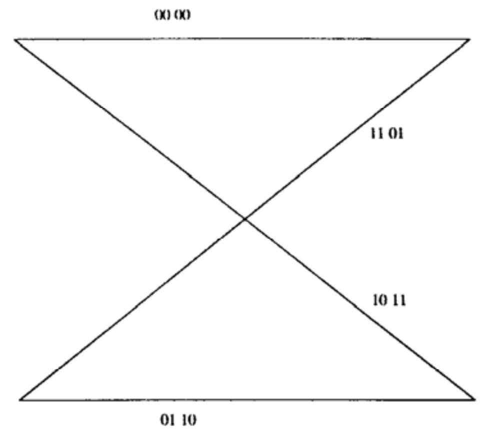

Fig. 21. BPSK smart–greedy code.

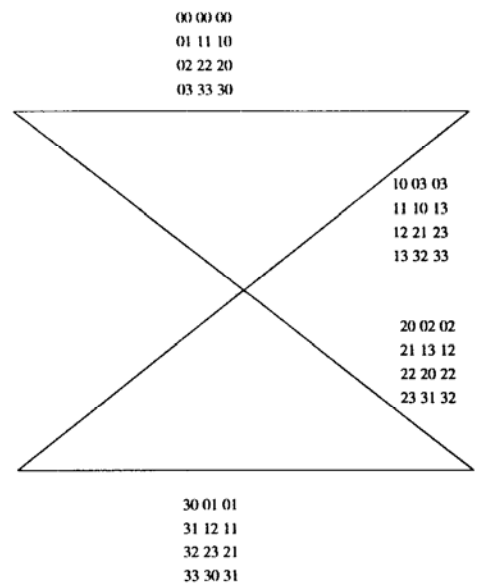

Fig. 22. 4-PSK smart–greedy code.

The following code (see Fig. 21) using M-TCM construction guarantees these diversity gains.

At time $2k+1$, $k = 0, 1, 2, \cdots$, depending on the state of the encoder and the input bit, a branch is chosen by the encoder and the first coordinate and second coordinates of the labels are sent simultaneously from the transmit antennas at times $2k+1$ and $2k+2$. For instance, at time 1, if the branch label 10 11 is chosen, symbols 1, 0 and 1, 1 are sent, respectively, from transmit antennas one and two at times one and two. From the design criteria established, it is easy to see that this code guarantees the desired diversities in static and rapid fading environments.

Example 3.9.2: Here a transmission rate of 1 b/s/Hz and diversity gains of 2 and 3, respectively, in static and rapid flat fading environments are desired. From the criteria, we know that a diversity gain of 2 is possible in a static flat fading environment and this transmission rate can be accomplished using a BPSK constellation. In this example, we will use the 4-PSK constellation instead (see Fig. 3). Our objective is to guarantee a diversity gain of 2 and 3, respectively, in slow

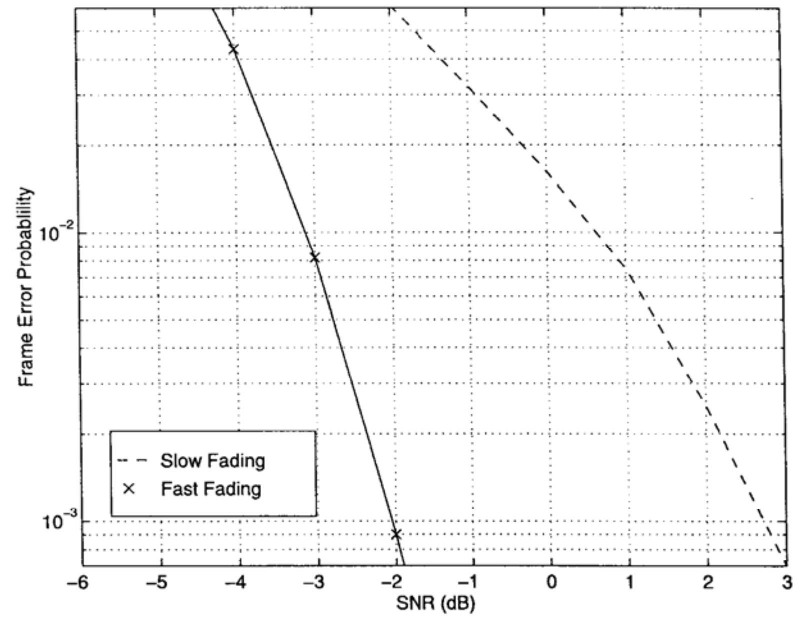

Fig. 23. Performance of the code of Example 3.9.3 with two receive and two transmit antennas.

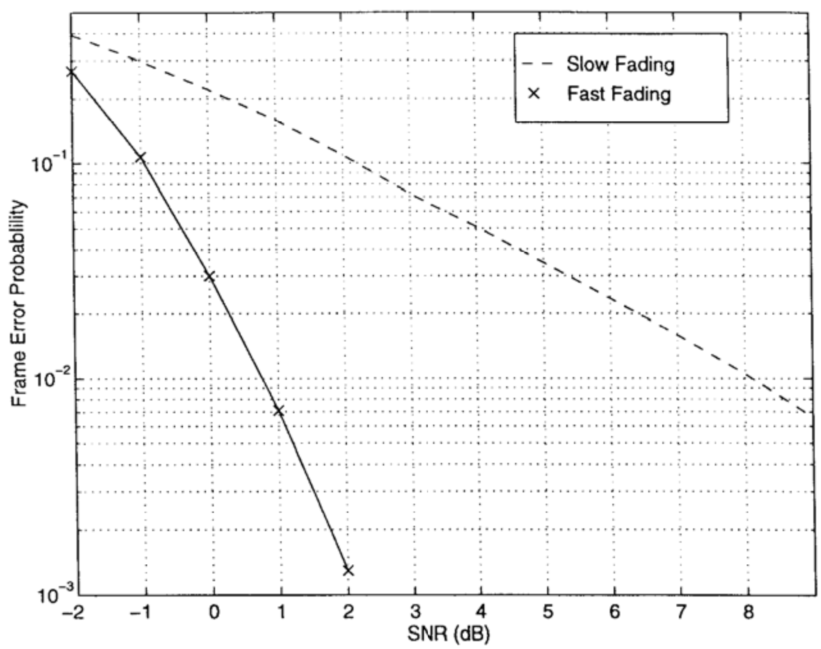

Fig. 24. Performance of the code of Example 3.9.3 with one receive and two transmit antennas.

and rapid flat fading environments. The code of Fig. 22 using M-TCM construction guarantees these diversity gains. From the design criteria established above, it is easy to see that this code guarantees the desired diversities in static and rapid fading environments.

In both these examples the design of smart–greedy codes of the same rate and better performance having higher number of states is also possible. Another possibility is concatenation with appropriate RS codes. We demonstrate the performance of these codes by the following examples.

Example 3.9.3: Consider the code of Example 3.9.1 as the inner code and a $[16, 12, 5]$ extended RS code over GF(16) as an outer code. For 48 coded input bits (12 symbols of GF(16)) and one terminating bit set equal to zero, the output of the outer code corresponds to 65 bits which is used as the input to the smart–greedy space–time encoder. The output of the smart–greedy space–time encoder is two frames of length 130 symbols of BPSK symbols corresponding to two transmit antennas. The uncoded zero bit guarantees that the encoder of the inner code is at zero state at the end of each frame. The rate of this smart–greedy code is almost 0.37 b/s/Hz. The performance of this concatenated code is given in Figs. 23 and 24 for, respectively, two and one receive antennas.

Example 3.9.4: Consider the code of Example 3.9.2 as the inner code and a $[25, 21, 5]$ shortened RS code over GF(32) as an outer code. For 105 coded input bits, four uncoded input

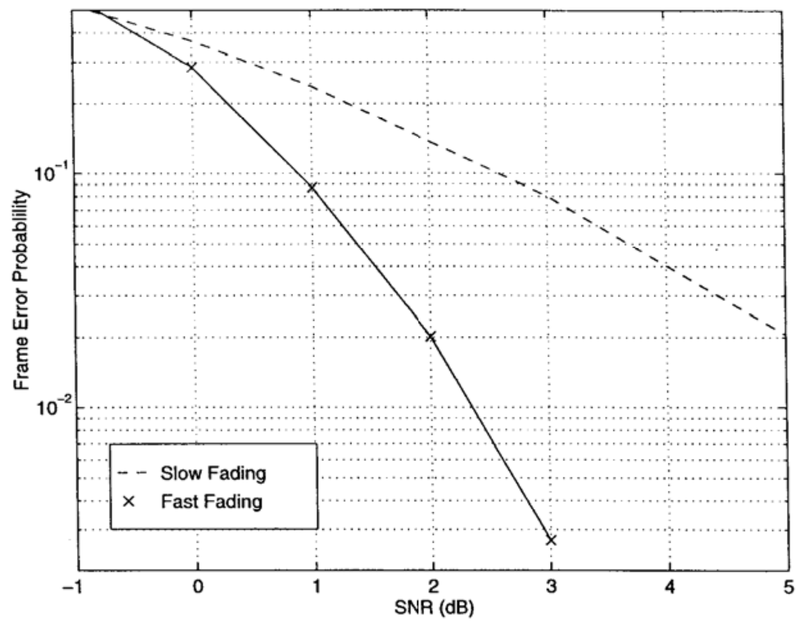

Fig. 25. Performance of the code of Example 3.9.4 with two receive and two transmit antennas.

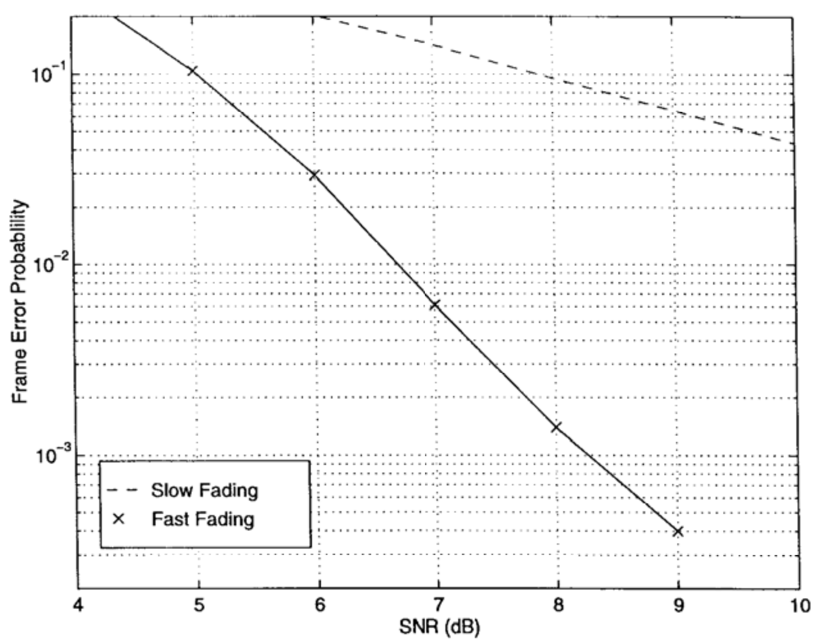

Fig. 26. Performance of the code of Example 3.9.4 with one receive and two transmit antennas.

bits, and one terminating bit set equal to zero, the output of the outer code corresponds to 130 bits which is used as the input to the smart–greedy space–time encoder. The output of the smart–greedy space–time encoder is two frames of length 130 symbols of 4-PSK symbols corresponding to two transmit antennas. The uncoded zero bit guarantees that the encoder of the inner code is at zero state at the end of each frame. The rate of this smart–greedy code is almost 0.83 b/s/Hz. The performance of this concatenated code is given in Figs. 25 and 26 for, respectively, two and one receive antennas.

The greediness and smartness of the codes can be observed from the above performance curves. These codes are also ideal for improving the frequency reuse factor.

J. Trellis Versus Block-Coded Modulation

There has been recently an explosion of interest in the trellis complexity of codes. In this light, one may ask if a block-coded modulation scheme can outperform the space–time trellis codes in terms of the tradeoff between complexity of implementation rate and diversity advantage.

It is well known that a block-code trellis is time-varying and harder to implement than that of a space–time trellis code. Space–time trellis codes have significant advantage over the block codes that only one time section of the trellis must be stored in memory. Moreover, for a block-coded modulation scheme the number of ACS (Add–Compare–Select) elements required at each time instance is different, making both DSP and VLSI implementation less attractive. The problem of designing a block-coded modulation scheme that satisfy the rank criterion is also an open problem. These deficiencies aside, we further discourage the possibility of potential application of block codes by proving that they cannot outperform the space–time trellis codes in terms of the tradeoff between the diversity advantage, rate, and trellis complexity.

It will be assumed that the reader is familiar with the theory of trellis complexity of block codes (for details, we refer the reader to [13]).

Let Q denote the constellation of $q = 2^b$ (where b is not necessarily an integer) signals to be used for information transmission. Consider a block code $\mathcal{C}$ and a trellis for $\mathcal{C}$. Suppose that the code is used to transmit q-ary symbols via n antennas using frames of length l. Let us fix the transmission model of the previous sections. The codewords are then blocks of length nl of q-ary symbols. At any time instance n symbols of branches of a path in the trellis are assigned to n points of the constellation in some manner. The n signal points are then simultaneously sent via the n transmit antennas. We have the following theorem.

Theorem 3.10.1: Consider a block code $\mathcal{C}$ defined over a q-ary alphabet and a trellis for $\mathcal{C}$. Suppose that $\mathcal{C}$ is employed as above for transmission of information using n transmit and m receive antennas. If the achieved diversity gain is mr and the transmission rate is R bits per second per hertz, then $s_{\max}$, the number of maximum states in the trellis of $\mathcal{C}$, satisfies

$$s_{\max} \geq 2^{R(r-1)}. \tag{23}$$

Furthermore, the above bound is still valid even if the trellis is sectionalized into segments of length n.

Proof: It suffices to prove the last statement of the theorem. To this end, suppose that a sectionalized trellis of $\mathcal{C}$ is given with each branch labeled by n constellation symbols. Then, it follows from the rank criterion, that no two paths of this sectionalized trellis diverging from some state can remerge at another state in a time interval of length less than r.

Then a straightforward variant of [22] proves that $s_{\max}$, the maximum number of states in the sectionalized trellis, satisfies the inequality $s_{\max} \geq q^{(r-1)\log_q(|\mathcal{C}|)/l}$. Given that $q = 2^b$ and observing that $R = \log_2(|\mathcal{C}|)/l$, we arrive at the inequality $s_{\max} \geq 2^{R(r-1)}$. □

Corollary 3.10.1: No block code (that admits a trellis representation) can outperform the designs of this paper in terms of the tradeoff between diversity gain, rate, and trellis complexity.

Proof: The above bound is similar to the bound established for the space–time trellis codes which can be attained for our designs. □

IV. Conclusions

We unveiled a new family of codes called the Space–Time codes for transmission using multiple transmit antennas over Rayleigh or Rician wireless channels. Many subfamilies of space–time codes were also introduced. The performance of these codes was shown to be excellent, and the decoding complexity comparable to codes used in practice on Gaussian channels. Space–time codes have simple systolic architecture and can be readily implemented in DSP and VLSI.

Various fundamental theoretical limits on rate, trellis complexity, diversity, constellation size, and their tradeoffs were established. Examples were provided confirming that the limits we established are attainable in practice.

We believe that the studies we initiated here, only scratch the tip of the iceberg and many important questions remain to be answered. Research on the interactions and combinations of the space–time coding technology with other techniques such as orthogonal frequency division multiplexing [3], array processing [33], and numerous other topics is now being pursued.

References

[1] Special Issue on the European Path Toward UMTS, *IEEE Personal Commun. Mag.*, vol. 2, Feb. 1995.

[2] Ericsson, "Mobile telephony—Market overview." [On Line]. WWW: http://www.ericsson.com/BR/market/.

[3] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-Time coded OFDM for high data rate wireless communication over wideband channels," *IEEE VTC'98,* submitted.

[4] N. Balaban and J. Salz, "Dual diversity combining and equalization in digital cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 342–354, May 1991.

[5] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis Coded Modulation With Applications.* New York: Macmillan, 1991.

[6] A. R. Calderbank, "Multilevel codes and multistage decoding," *IEEE Trans. Commun.*, vol. 37, pp. 222–229, 1989.

[7] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. 33, pp. 177–195, 1987.

[8] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis coded modulations in Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, Jan 1992.

[9] L. J. Cimini, Jr. and N. R. Sollenberger, "OFDM with diversity and coding for high bit-rate mobile data applications," in *Proc. 3rd Int. Workshop on Mobile Multimedia Communications*, Sept. 1996, paper A3.1.1.

[10] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channel: Performance criteria," *IEEE Trans. Commun.*, vol. 36, pp. 1004–1012, Sept. 1988.

[11] ——, "The design of trellis coded MPSK for fading channel: Set partitioning for optimum code design," *IEEE Trans. Commun.*, vol. 36, pp. 1013–1021, Sept. 1988.

[12] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, 1991.

[13] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: state spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, 1993.

[14] G. J. Foschini, Jr. and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, submitted for publication.

[15] 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," in *Proc. IEEE VTC'96*, pp. 136–140.

[16] P. S. Henry and B. S. Glance, "A new approach to high capacity digital mobile radio," *Bell Syst. Tech. J.*, vol. 51, pp. 1611–1630, Sept. 1972.

[17] R. D. Gitlin, J. Salz, and J. H. Winters, "The capacity of wireless communication systems can be substantially increased by the use of antenna diversity," *IEEE J. Select. Areas Commun.*, submitted for publication.

[18] T. Hattori and K. Hirade, "Multitransmitter simulcast digital signal transmission by using frequency offset strategy in land mobile radio-telephone system," *IEEE Trans. Veh. Technol.*, vol. VT-27, pp. 231–238, 1978.

[19] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 170–176, May 1992.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis.* New York: Cambridge Univ. Press, 1988.

[21] A. Imai and S. Hirakawa, "A new multilevel coding method using error control codes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 264–268, 1978.
[22] A. Lafourcade and A. Vardy, "Asymptotically good codes have infinite trellis complexity," *IEEE Trans. Inform. Theory*, vol. 41, pp. 555–559, Mar. 1995.
[23] A. D. Little, *Wireless Business and Finance,* July 1996.
[24] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes.* New York: North-Holland, 1977.
[25] A. Narula, M. Trott, and G. Wornell, "Information theoretic analysis of multiple-antenna transmission diversity," *IEEE Trans. Inform. Theory*, submitted for publication.
[26] G. J. Pottie, "System design issues in personal communications," *IEEE Personal Commun. Mag.*, vol. 2, pp. 50–67, Oct. 1995.
[27] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," in *Proc. IEEE GLOBECOM'96*, pp. 1809–1814.
[28] C.-E. W. Sundberg and N. Seshadri, "Digital cellular systems for North America," in *IEEE GLOBECOM'90*, Dec. 1990, pp. 533–537.
[29] ——, "Coded modulation for fading channels: An overview Invited Paper," *European Trans. Telecommun. and Related Technol.* (Special Issue on Applications of Coded Modulation Techniques), pp. 309–324, May 1993.
[30] N. Seshadri and C.-E. W. Sundberg, "Multi-level trellis coded modulation for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 41, pp. 1300–1310, Sept. 1993.
[31] W. C. Jakes, *Microwave Mobile Communications.* Piscataway, NJ: IEEE Press, 1993.
[32] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, no. 1, 1994.
[33] V. Tarokh, A. Naguib, N. Seshadri and A. R. Calderbank, "Array processing and space-time coding for very high data rate wireless communication," preprint.
[34] ——, "Space-time codes for wireless communications: Practical considerations," *IEEE Trans. Commun.*, submitted for publication.
[35] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T-Bell Labs Internal Tech. Memo., June 1995.
[36] G. Ungerboeck, "Channel coding for multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
[37] V. Weerackody, "Diversity for direct-Sequence spread spectrum system using multiple transmit antennas," in *Proc. IEEE ICC'93*, May 1993, pp. 1775–1779.
[38] L.-F. Wei, "Coded M-DPSK with built-in time diversity for fading channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1820–1839, Nov. 1993.
[39] S. G. Wilson and Y. S. Leung, "Trellis coded phase modulation on Rayleigh fading channels," in *Proc. IEEE ICC'97*, June 1997.
[40] J. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42. pp. 1740–1751, Feb./Mar./Apr. 1994.
[41] J. H. Winters, "Switched diversity with feedback for DPSK mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. VT-32, pp. 134–150, Feb. 1983.
[42] ——, "Diversity gain of transmit diversity in wireless systems with Rayleigh fading," in *Proc. IEEE ICC*, 1994, vol. 2, pp. 1121–1125.
[43] A. Wittneben, "Base station modulation diversity for digital SIMULCAST," in *Proc. IEEE'VTC*, May 1993, pp. 505–511.
[44] ——, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE'ICC*, 1993, pp. 1630–1634.